Universitat Pompeu Fabra Barcelona

Tonality Estimation in Electronic Dance Music

A Computational and Musically Informed Examination

Ángel Faraldo Pérez

TESI DOCTORAL UPF / 2017

Thesis Director:

Dr. Sergi Jordà Puig

Music Technology Group

Department of Information and Communication Technologies

Universitat Pompeu Fabra, Barcelona, Spain

Dissertation submitted to the Department of Information and Communication Technologies of Universitat Pompeu Fabra in partial fulfilment of the requirements for the degree of

DOCTOR PER LA UNIVERSITAT POMPEU FABRA

Music Technology Group (http://mtg.upf.edu), Department of Information and Communication Technologies (http://www.upf.edu/dtic), Universitat Pompeu Fabra (http://www.upf.edu), Barcelona, Spain.

The doctoral defence was held on .... at the Universitat Pompeu Fabra and scored as ....

Dr. Sergi Jordà Puig
(Thesis Supervisor)
Universitat Pompeu Fabra (UPF), Barcelona, Spain

Dr. Anja Volk
(Thesis Committee Member)
Universiteit Utrecht, The Netherlands

Dr. Gianni Ginesi
(Thesis Committee Member)
Escola Superior de Musica de Catalunya (ESMUC), Barcelona, Spain

Dr. Agustín Martorell Domínguez
(Thesis Committee Member)
Universitat Pompeu Fabra (UPF), Barcelona, Spain

This thesis has been carried out at the Music Technology Group (MTG) of Universitat Pompeu Fabra in Barcelona, Spain, from January 2014 to December 2017, supervised by Dr. Sergi Jordà Puig and Prof. Perfecto Herrera Boyer.

My work has been supported by the Department of Information and Communication Technologies (DTIC) PhD fellowship (2015-17), Universitat Pompeu Fabra, and the European Research Council under the European Union's 7$^{\text{th}}$ Framework Program, as part of the GiantSteps project (FP7-ICT-2013-10, grant agreement 610591).

Efforts in several parts of this thesis were closely related to the GiantSteps project and have been carried out in collaboration with the partner institutions (Johannes Kepler University, Linz; Native Instruments, Berlin; Reactable Systems, Barcelona; Studio for Electro-Instrumental Music, Amsterdam), and with the Music and Multimodal Interaction Lab at the MTG, led by Dr. Sergi Jordà. My closest collaborators include, in alphabetical order, Pere Calopa, Daniel Gómez-Marín, Peter Knees, Martin Hermant and Cárthach Ó Nuanáin.

Abstract

This dissertation revolves around the task of computational key estimation in electronic dance music, upon which three interrelated operations are performed. First, I attempt to detect possible misconceptions within the task, which is typically accomplished with a tonal vocabulary overly centred in Western classical tonality, reduced to a binary major/minor model which might not accomodate popular music styles. Second, I present a study of tonal practises in electronic dance music, developed hand in hand with the curation of a corpus of over 2,000 audio excerpts, including various subgenres and degrees of complexity. Based on this corpus, I propose the creation of more open-ended key labels, accounting for other modal practises and ambivalent tonal configurations. Last, I describe my own key finding methods, adapting existing models to the musical idiosyncrasies and tonal distributions of electronic dance music, with new statistical key profiles derived from the newly created corpus.

Resum

Aquesta tesi doctoral versa sobre anàlisi computacional de tonalitat en música electrònica de ball. El meu estudi es concentra en tres operacions fonamentals. Primer, intento assenyalar possibles equívocs dins de la pròpia tasca, que normalment es desenvolupa sobre un vocabulari tonal extremadament centrat en el llenguatge de la música clàssica europea, reduït a un model binari major/menor que podria no acomodar estils de música popular. Seguidament, presento un estudi de pràctiques tonals en música electrònica de ball, efectuat en paral·lel a la recol·lecció i anàlisi d'un corpus de més de 2.000 fragments de música electrònica, incloent diversos subgèneres i graus de complexitat tonal. Basat en aquest corpus, suggereixo la creació d'etiquetes tonals més obertes, que incloguin altres pràctiques modals així com configuracions tonals ambigües. Finalment, descric el meu sistema d'extracció automàtica de tonalitat, adaptant models existents a les particularitats de la música electrònica de ball, amb la creació de distribucions tonals específiques a partir d'anàlisis estadístiques del recentment creat corpus.

Resumen

Esta tesis doctoral versa sobre análisis computacional de tonalidad en música electrónica de baile. Mi estudio se concentra en tres operaciones fundamentales. Primero, intento señalar posibles equívocos dentro de la propia tarea, que normalmente se desarrolla sobre un vocabulario tonal extremadamente centrado en el lenguaje de la música clásica europea, reducido a un modelo binario mayor/menor que podría no acomodar estilos de música popular. Seguidamente, presento un estudio de prácticas tonales en música electrónica de baile, efectuado en paralelo a la recolección y análisis de un corpus de más de 2.000 fragmentos de música electrónica, incluyendo varios subgéneros y grados de complejidad tonal. Basado en dicho corpus, sugiero la creación de etiquetas tonales más abiertas, que incluyan otras prácticas modales así como configuraciones tonales ambiguas. Por último, describo mi sistema de extracción automática de tonalidad, adaptando modelos existentes a las particularidades de la música electrónica de baile, con la creación de distribuciones tonales específicas a partir de análisis estadísticos del recién creado corpus.

la vida es la memoria
y el hombre es el azar
Fernando Arrabal

Acknowledgements

Let me begin my acknowledgements with my loved ones, since without them, I would not have even considered embarking on this journey. Padre and Madre, thanks for supporting me so instinctively, and for making me embrace the doubt. Marta, the most intense supporter, the greatest lover, the worst cook, partner of so many new things. I would have died of starvation and sadness if you weren't around me during these last months. Thank you for being, simply, as you are. Yolanda, we moved from the Dark places into the Bright, together. That was the very beginning of my whole PhD adventure. Then, our souls decided to find out about the world separately. Today, I arrive at the end of what brought us to Barcelona, and I want to thank you for all your support and understanding. Silvia, Emilio, you are the true friends everyone deserves. I am so lucky to have you around, and it is partly because of you that I have not gone completely mad during my writing. You introduced me to so many new souls, to Xiana, the people of Can Masdeu... *Moitas grazas por tanto agarimo*.

Cárthach, Martinuqui, Dani, you were the best imaginable companions for such a journey. We shared everything. Drugs, doubts, study. We shaped our projects together so they could complement each other. We tried. Most of the times we failed. But the few things we achieved, we did greatly together. You were the most gorgeous occupants of the MTG, the place that was our home for around four years, and where we met other amazing fellows. Juanjo, Marius, the GiantSteps project and the people involved in it, Peter Knees, the two Mickael's, Kristina, Bucci. The Great Bucci. It was so good to shape my project through the various cities, beers and conversations. Thank you for that.

Parts of my research would have not even been imagined without three years of quite intensive teaching, thanks to a number of students testing prototypes enthusiastically, providing many helpful insights. I have learnt a lot from that process. Eduard Mas, Sergio Latre, your help annotating and correcting almost 1,500 audio tracks of electronic dance music is priceless, as much as it is the exceptional work of Daniel G. Camhi revising in careful detail a corpus of 500 additional fragments. Your contributions have been so important, that a whole chapter of this dissertation is written in your honour.

I am almost about to put an end to a process that has taken four years of my life. And I still have not mentioned the two beings thanks to whom this journey has come true. Perfe, you have been the greatest scientific supervisor I have ever met. Your patience deserves a monument; your knowledge, a library. Your ability to release tension and give support with small actions, is simply inestimable. Sergi, the genius in disguise. We briefly met on Skype and that was enough to make up my mind and move to Barcelona, leaving eight Dutch years behind. You have been an excellent boss and advisor, an extremely balanced Yin complementing Perfe's Yang. A trip companion, and after these four years, a friend. Thank you so very much for trusting.

Barcelona, 21st December, 2017
Ángel Faraldo

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

"You should look for a completely different idea,
elsewhere, in another area, so that something passes
between the two which is neither in one nor in the other."
Gilles Deleuze, *Dialogues* (1977)

Electronic dance music, or its acronym EDM, is a meta-label that refers to a number of musical practises originating in the 1980's and extending into the present, made almost solely with electronic equipment and a strong presence of percussion imposing a steady beat, and mostly intended for dancing at nightclubs and raves.

In my opinion, these two broad descriptors —*dance* and *electronic*— are at the very origin of what is arguably the most drastic shift in the development of popular music in the Twentieth Century, exerting an influence in music production and consumption habits comparable to the arousal of musical notation, the standardisation of equal temperament or the arrival of recording technology in previous historical moments. From the music industry to music education, EDM has revolutionised the ways of composing and performing music, the acts of collective music consumption, and the very notions of authorship and musicianship.

This highly technological turn, brings in a number of opportunities for those working in areas related to computer engineering, artificial intelligence, information retrieval, and music technology, including a myriad of real-world applications, such as music recommendation systems, educational resources, and creative tools for the electronic music producer or DJ.

1.1 Motivation

As a music professional, the principal domains in which I develop my work are music technology and music composition, magnetised by the appeals of musical formalisation and the expressive powers of new musical instruments alike. In particular, I have been long interested in the tension between digital technology and musical expression, and in how this friction influences the development of musical language. Therefore, it felt natural to embark on a project that could couple certain aspects of computational data extraction with more orthodox musical analysis, materialising into tools that could offer the electronic —dance— music maker analytical insights to guide or support her creative flow in non-intrusive ways.

There seems to be an ample interest in understanding the creative processes behind electronic dance music, as suggested by the increasing number of online magazines and user fora, offering discussions on new technology and production techniques, interviews with artists and producers, and release reviews.[1] The relevance of electronic dance music has also been acknowledged by music scholars (e.g. Tagg, 1994; Middleton & Manuel, 2015), and is reflected in the proliferation of publications addressing EDM from several interdisciplinary perspectives, including social and cultural studies (e.g. Thornton, 1995; Rietveld, 1998), music journalism (e.g. Reynolds, 1998; Brewster & Broughton, 2000), ethnomusicology (Fikentscher, 2000), and, to a lesser extent, music theory, where most efforts have gone into elucidating aspects of rhythm and structure, as they are the most salient aspects of EDM (Butler, 2006). The study of its tonal practises, on the other hand, has remained somehow unattended, as pitch and harmony normally play secondary roles in these genres. Nevertheless, there is some evidence that novel tonal techniques —detached from previous conceptions of tonality— are developed by EDM practitioners as part of their musical language (Wooller & Brown, 2008), creating new sound aggregates (e.g. by combining different musical sources together) and temporal structures, such as large-scale DJ sets. Furthermore, EDM producers are in demand of tools providing tonal descriptions of tracks, in order to facilitate the classification and mixing of sound files.

For these reasons, the task of automatic key estimation seemed an interesting starting point for my research, connecting scientific domains —mostly music information retrieval— to music-theoretical interests, such as elucidating how EDM musical practises might have caused novel tonal configurations, just as much as they have produced new rhythmic and formal structures.

[1] http://www.synthzone.com/mags.htm, for example, lists over 20 "electronic music magazines, publications and journals". Moreover, a simple query in any online search engine should provide quick access to numerous online resources.

1.2 Contexts of Action

However, before I continue clarifying my research goals, this section provides a brief description of what I consider the most important aspects of electronic dance music, contributing to the consolidation of a unique practice, differentiated from other popular music styles. Although this initial report might appear a bit lengthy, I regard it an important asset for a better understanding of the nature of my research objectives and thesis contributions, discussed in the next two sections, and for this reason is presented at this point. Similarly, I dedicate a few paragraphs to introduce the scientific domain of music information retrieval, within which most of my research should be framed.

1.2.1 Electronic, Dance, Music

In the introductory chapter to his compilation of essays on EDM, Butler (2012, pp. xi–xii) outlines four pervading aspects that characterise what is an otherwise heterogeneous group of practises, styles, contexts and geographical locations. The first of these four aspects, he claims, is the central position of the *recording*, not as mode of distribution —as in most popular music nowadays— but as the primary element of performance itself. Secondly, *dancing* is taken as the principal producer of meaning, dancers being the "performing audience", in contrast to the more passive consumer of other types of music, as Butler put it. The third common trace is related to the site-specificness of "collective dancing to recorded music", be this in the *club*, a unique space designed specifically for this purpose, or in the *rave*, normally one-time massive events happening at picturesque locations. The last element tying these musics together is, according to Butler, their common roots in 1970's *disco*, in which essential practices of what later would be recognised as DJ culture originated, such as the constant musical flow throughout the session, or techniques like beat- and tempo-matching (Brewster & Broughton, 2000).

Therefore, although other authors have expressed their discontent with the term 'electronic dance music' as an umbrella for such a diverse account of practices (e.g. McLeod, 2001; Doehring, 2015), the fact that it condenses essential elements of these manifestations in a open-ended way —*electronic, dance, music*— together with the absence of a better denomination, has made it consolidate as an appropriate metagenre label, and in that sense it is used throughout this dissertation.[2]

[2]However, it is perhaps worth noting that the label 'electronic dance music' has been appropriated by North American music industry to refer to a specific subgenre of US post-dubstep arising around 2010.

One of the most salient aspects of EDM is, without a doubt, its 'all-electronic' quality (Butler, 2006, p. 33), establishing a clear boundary with other popular music styles, mostly vocal and guitar-centred. Electronic dance music originates from the record itself —even if the record contains vocals and acoustic instruments— bringing into play a new type of musician (the DJ), a new instrument (the turntable), and a new notion of musical skill, consisting in playing records instead of notes, in combining existing musics to arrive at a new sound, rather that composing with a previously defined palette of notes and chords. This technological orientation soon embraced all sorts of electronic appliances, including synthesisers, drum-machines, sequencers, samplers and —later on— computers, with which EDM makers add additional layers to their mixes, eventually creating music from scratch, giving raise to the figure known as the *producer*. These diametrically different and complementary approaches to EDM —mixing vs. producing— pervade the whole development of EDM, and are already present at the very origin of the metagenre.

Another distinctive characteristic of EDM is its purely 'sonic' nature, contrasting with the enormous importance of vocals in all other popular music styles. Contrary to this tendency, EDM is predominantly instrumental, and the use of voices —sung or spoken— is at best restricted to a repeating sentence or a few scattered words. The exception to this norm is clearly represented by hip hop, which, although undoubtedly grounded in DJ's mixing culture, it inherits the strophic nature of lyrics.

The *loop* represents the quintessential structural unit of EDM. The origin of loop-based composition can be arguably traced back to *musique concrète* (Schaeffer, 2017, first published in 1966) and 1960's rock (Spicer, 2004), as a natural consequence of the developments in multi-track recording and production technology. However, in EDM, the loop stands as the main appropriative matter —consisting mainly of drum-kit breaks and bass snippets— upon which the musical structure unfolds.

Perhaps symptomatically, EDM compositions are typically presented as 'tracks', denoting the characteristic reductionism of EDM compared to other musical styles. In contrast with the song format (multi-layered, based on strophic alternation, with pitch and semantic implications), EDM tracks appeal to the driving role of the percussion track as the principal organiser of the musical flow, upon which additional layers might become as little as ornamental (Doehring, 2015, p. 133). In Butler's words, "in EDM, *drums* are the music, to the extent that the few melodic elements that are present [...] frequently assume a percussive role as well" (Butler, 2006, p. 93).

Figure 1.1 presents an imaginary 'map' of three relatively nearby cities, where the first genres of EDM originated almost simultaneously.[3] My intention in locating these cities in such an arrangement is to illustrate some formative contrasts that, in my humble opinion, pervade the whole history of EDM. The right side of Figure 1.1

[3] Although geographically speaking Chicago is West of Detroit, Figure 1.1 is arranged according to musical similarities rather than to their geographical location.

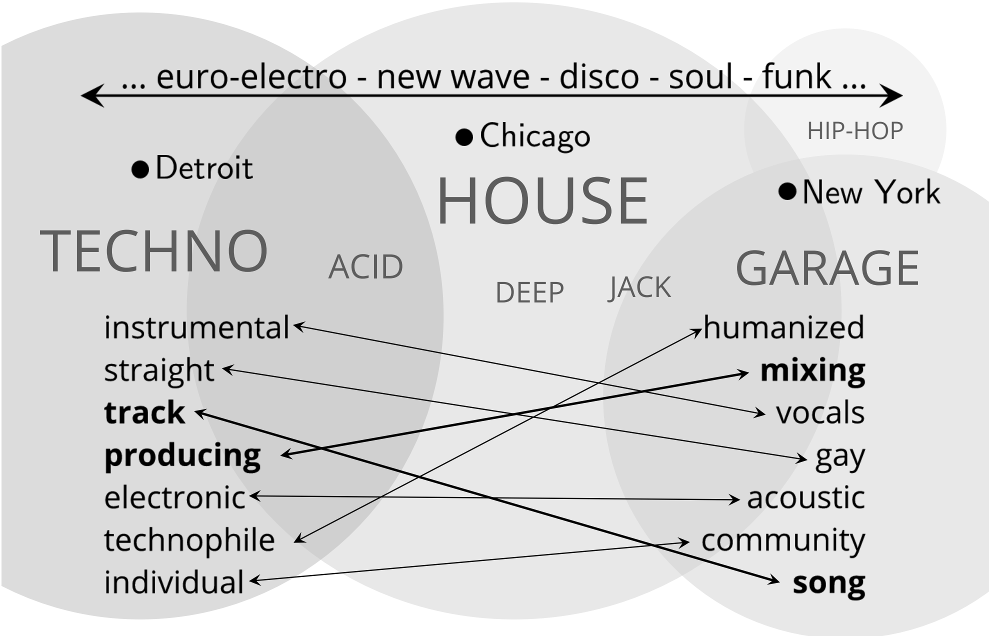


FIGURE 1.1: Musical and social space in early EDM subgenres. This imaginary map intends to show some of the main tendencies and contrasts that characterise different EDM styles, already present in these early manifestations.

represents EDM styles bearing a clearer influence of disco music, as typified by New York's garage and other early variants of house. These subgenres are closer to the song form characteristic of other popular styles, presenting acoustic instrumentation (although from recordings) and vocal parts, and exerting an influence in most forms of mainstream pop music nowadays. Early house styles were originally integrated within the black and gay social network, organised around emblematic collective dancing spaces (New York's *Paradise* or Chicago's *Warehouse*), where an extremely refined practise of mixing originated and developed. In contrast, in the left side of Figure 1.1 (bear in mind that arrows represent a continuum between both extremes), Detroit's techno music exemplifies a more introverted tendency, originating mostly in the studio, and exploiting the expressive powers of drum machines, sequencers and synthesisers, stimulated by a certain dystopian and technophile imaginary, with influences of Kraftwerk and new wave. Current sequels of early techno can be traced in styles such as minimal techno, progressive, or tech house, for example. It is worth noting that Figure 1.1 locates hip hop in a different orbit. With this representation, I intend to illustrate that although hip hop shares certain practises of mixing and appropriation with other EDM styles, its preeminently vocal and lyrical aspect, leaves it out of the sphere of EDM, at least in what regards the remainder of this dissertation.

Musical Characteristics of EDM

According to Moore (2012), popular music styles can be differentiated and characterised by observing four basic textural/functional layers, namely the *explicit-beat* layer, the *functional-bass*, the *harmonic-filler* and the *melodic* layer. Within this descriptive framework, many EDM genres often prescind of the melodic and harmonic layers, keeping a tight interaction of the bass and beat layers.

As I have already advanced, EDM is mostly about the beat, and, regarding this aspect, Butler establishes a useful differentiation between two broad tendencies, dividing the EDM ecosystem into 'breakbeat-driven' and 'four-on-the-floor' styles (2006, p. 78). The first ones originated in Britain after house music was exported from the US, and typically comprise subgenres such as jungle, hardcore, UK garage or drum'n'bass. The essence of breakbeat styles is the 'dereconstruction' of classic soul and funk drum breaks into all sorts of temporal rearrangements. These styles tend to deemphasise strong beats, placing considerable stress on metrically weak parts (Butler, 2006, p. 78). On the other hand, four-on-the-floor genres originate in disco's steady bass-drum pattern, evolving —via house music— into a large variety of genres progressively further from the initial disco reference, including techno and trance, for example.

In general, EDM tracks are relatively fast, with tempos ranging between 120 and 150 BPM, although there are styles defined exactly by laying out of these boundaries, such as trip hop and downtempo electronica (with tempos as slow as 80 BPM), or gabber techno, reaching extremely high speeds of over 200 BPM (Butler, 2006, p. 34). As a matter of fact, tempo characterisation seems to be a reliable indicator of certain subgenres, for example, dubstep (140 BPM), drum'n'bass (160–180 BPM) or house (120–130 BPM).[4,5,6]

Besides rhythm and tempo, EDM subgenres are defined by the musical activity —if any— present in other textural layers, as much as by their 'instrumentation'. For example, in house and trance, is common to find a clear harmonic-filler layer, with chord progressions borrowed from either soul or jazz music (in deep house), or harmonic sequences from Western classical music, in some trance derivates. Similarly, trance is typically melodic, with lines achieving epic resonances in subgenres such as psy trance, whereas the house melodic layers remind of soul and rhythm'n'blues, with some variants like balearic house, mixing with and extending into mainstream pop music. On the contrary, techno and the subgenres under its sphere of influence, are much sparser regarding both the harmonic and melodic layers. Some styles such

[4] http://techno.org/electronic-music-guide

[5] http://www.complex.com/music/an-idiots-guide-to-edm-genres/grime

[6] Snoman (2009) provides a practical account of musical features of EDM subgenres in a recipe-like presentation, suggesting 'tricks' to face the production of different styles.

as minimal techno almost completely prescind of pitched materials, whereas hybrid forms such as progressive or tech house tend to integrate chordal or melodic units within the more intricate rhythmic layer. Sparsity in mid to high pitch registers is also characteristic of breakbeat-driven styles, which focus on heavy audio sample manipulation —breaks— counterbalanced with prominent basslines.

Regarding the instrumentation of the various subgenres, the general 'dichotomy' suggested in Figure 1.1 seems to easily accommodate newly created genres all the way to the present. On the one hand, genres owing to the disco/soul/funk traditions naturally bear resemblance with the instrumentation found in these popular music styles. This is true for much house music, which incorporates samples from acoustic instruments and vocal lines. Other subgenres such as minimal, techno, tech house, progressive house and trance, tend to favour fully synthesised textures, whilst hardcore and dance variants abuse sampled material of incidental sounds such as sirens, horns, spoken voices and other types of field recordings. Jungle, drum'n'bass and other breakbeat-driven genres seem to be especially inclined towards extreme sample manipulation, originating, as I have already noted, in acoustic drum breaks from soul records, and presenting quite an unique scenario within the EDM ecosystem, given the more noticeable influence of styles such as dub, reggae and hip hop, from which it borrows the presence of vocals.[7]

What all EDM subgenres have in common —disregarding their instrumentation and basic rhythmic layout— is their structural organisation based in repetition. The essential structural unit of EDM is the loop, a short excerpt of music that is rhythmically aligned, layered and repeated, alone or in combination with other loops. A typical EDM track is composed by aggregation and juxtaposition of smaller units of different lengths, coexisting at different musical layers. It is common to find one-bar units for most rhythmic-percussive patterns, two- or four-bar loops for harmonic-melodic content, and eight-bar loops for complete textural-structural sequences.

This "modular" structural organisation, as coined by Butler (2006), promotes what Spicer (2004) has denominated "accumulative form", by which a musical composition unfolds as an accumulation of thematic fragments —loops— creating a thickening texture, thus replacing "the climactic presentation of the main theme with the climactic accumulation of riffs into a texturally thick groove" (Garcia, 2005, par. 4.2). Furthermore, according to Garcia, accumulative forms in EDM are often populated with "aural signposts", guiding the listener/dancer throughout the musical structure (2005, par. 4.4). Accumulation in EDM is typically resolved with an expressive formula culminating in the so-called *drop*, which is divided into three subsequent

[7] As I stated before, for the sake of convenience, I do not regard hip hop within the umbrella of EDM, mostly appealing to the prominent role of prose and vocals in this music.

steps (Solberg, 2014). First, the *breakdown* introduces a sudden and momentary release of the rhythmic-percussive activity. Its main dramatic effect is the removal of the bass drum —what Butler has called "withholding the beat" (2006, p. 92). The breakdown is typically followed by a *build-up* stage, in which various types of "uplifters" (Solberg, 2014, p. 70) might be used, such as ascending arpeggios, glissandi, or other pitched-up components, normally precipitated by an acceleration in the quantisation of rhythmic elements, and necessarily resolving into the actual drop, the moment at which the foundational bass drum is reintroduced at maximum power, supplying the emotional peak of the track.

The Histories of EDM

To conclude this brief excursion into the contexts and characteristics of EDM, it is perhaps worth pointing at the various few historical accounts in circulation. The most ambitious of these is possibly the monograph by Reynolds (1998), a music journalist and rave addict himself, who reports on the origins and development of EDM through its various time-spaces, from an arguably personal vision. This is complemented by a compilation by Shapiro (2000), with chapters dedicated to individual subgenres. The role —and history— of the DJ has been studied by Brewster & Broughton (2000) and Fikentscher (2000), the first author to introduce an ethnomusicological perspective into EDM studies. Regarding specific genres, Rietveld (1998) has dedicated an individual study to house music, whereas techno has also been object of a number of publications (Sicko, 1999; Barr, 2000). All of this monographs were published just before the turn of the century, and address 'histories' from the first twenty years of EDM. Similarly, Butler (2006) only reports on the historical origins of EDM (New York garage, Chicago house and Detroit techno), admitting the much lengthier implications of a proper history of EDM.

1.2.2 Music, Information, Retrieval

The relatively young area of music information retrieval[8] (MIR) —a discipline consolidating towards the year 2000— attempts to extract, analyse and otherwise study aspects of music with computational approaches. MIR emerges as an interdisciplinary field combining music-related studies, mostly music theory, musicology and music cognition, with engineering domains such as signal processing, machine learning, statistics and data science (Downie, 2003; Schedl et al., 2014).

[8] Also referred to as music information *research*.

Broadly speaking, MIR tends to break down research problems into music's constitutive parameters, such as pitch, rhythm or structure (e.g. chord detection, meter recognition, structural segmentation), in connection with perceptual questions (e.g. tonality inference, downbeat detection), cultural aspects (e.g. genre recognition, cover song identification) and other domain-specific problems (e.g. audio source separation, audio-to-musical-score alignment). In its origins, MIR was mainly devoted to extracting information from symbolic formats, most notably MIDI, although research on symbolic realms has expanded to incorporate refined score following systems that are successfully used in live performances (Cont, 2008) and image recognition endeavours, hand in hand with musicological study and library science (Rebelo et al., 2012). Nowadays, most research efforts lay in extracting information from audio signals, although the overwhelming presence of the internet and digitisation of knowledge, is drawing an increasing attention towards the study of music-related textual and semantic data (Oramas & Sordo, 2016).

Research outcomes from the MIR community, are presented in music-and-technology conferences, such as the International Computer Music Conference[9] or the Sound and Music Computing Conference,[10] as well as in engineering pools including IEEE[11] or the Audio Engineering Society.[12] Since the year 2000, the International Society for Music Information Retrieval organises its own annual conference, ISMIR,[13] with notable impact across academia and industry. Very recently, the same organisation launched an open-access journal initiative.[14]

Regarding electronic dance music, MIR research has mostly focused in the domains of meter and rhythm (Heittola & Klapuri, 2002; Hockman et al., 2012; Leimeister et al., 2014; Panteli et al., 2014; Hörschlager et al., 2015; Gómez-Marín et al., 2016), given the central position that these elements bear in EDM. However, there has been research looking at other musical aspects, such as timbre characterisation (Rocha et al., 2013; Honingh et al., 2015), structure detection (Aljanaki et al., 2014; Glazyrin, 2014; Yadati et al., 2014; Scarfe et al., 2014; López-Serrano et al., 2016), genre identification (Kirss, 2007; Jacobson et al., 2007; Collins, 2012), and to a lesser extent, key estimation (Sha'ath, 2011), an area to which I have contributed two publications in the course of my research (Faraldo et al., 2016a, 2017).

[9] http://computermusic.org/page/23
[10] http://www.smc-conference.org
[11] https://www.ieee.org
[12] http://www.aes.org
[13] http://www.ismir.net
[14] https://transactions.ismir.net

One of the reasons for the good health of the discipline is the applicability of MIR research to real-world scenarios. These include classification and study of music libraries for scholar research, recommender systems for music streaming services, such as *Spotify*,[15] and the development of creative tools for electronic musicians and amateurs alike, facilitating otherwise tedious tasks, such as organising sample collections by tempo or key, providing musical descriptions and intuitive visualisations of musical knowledge, or offering alternatives for creative variation and continuation. The *GiantSteps* project, an European initiative gathering together partners from academia, industry and music education, was an important effort to bridge current advancements in music computing with the needs of EDM creatives. It is in this context that almost the totality of my doctoral research has been carried.[16]

1.3 Research Objectives

Although a popular area in the music information retrieval community, the task of automatic key extraction from audio has been slightly overseen in recent years, perhaps considered somewhat of a solved problem. Academic algorithms and commercially available applications provide relatively solid key estimation solutions, although their performance changes drastically when addressing different musical styles. This suggests that differences in the musical function of pitch and harmony call for different engineering approaches, taking into account stylistic particularities rather than aiming for all-purpose solutions, something that has been already noted by Gómez (2006a).

There are a number of available methods tailored specifically to EDM. Most of them arise as aiding tools for *harmonic mixing*, a technique largely used by DJs to sequence music tracks according to their tonal similarity (Vorobyev & Coomes, 2012). However, these solutions tend to present similar limitations: (a) they are restricted to a binary classification into major and minor keys, and (b) they normally produce one single label per track, given their orientation towards large-scale DJ sets.

In my own listening experience, however, these restrictions do not correspond with the apparent complexity of EDM, where I frequently find myself surprised with rare pitch combinations, modal configurations other than typical major or minor scales, tonally ambiguous passages, and other atonal or 'atonical' excerpts.

Therefore, in the particular context of EDM, the task of automatic key detection from audio seemed an interesting objective, more relevant than other tonality-related MIR problems, such as automatic chord extraction or melody identification, since

[15] https://www.spotify.com
[16] http://www.giantsteps-project.eu

musical actors such as chords, melodies and tonal directionality do not seem to be all that characteristic of EDM. On the other hand, the interplay of different pitch-class sets from various simultaneous musical sources at various textural levels, conveying different degrees of tonal strength and modal ambiguity, seemed a passionating and promising area of study.

As a consequence of this, the main goal of my research has been to diagnose the performance of key estimation algorithms in EDM, proposing musically informed alternatives to existing methods. In turn, this endeavour has been the principal motivation behind studying idiosyncratic tonal practises in EDM, what has become a second objective of this research on its own.

Since I started my doctoral program, I got increasingly convinced that electronic music production techniques —first revolving around record players, sequencers and samplers, nowadays mostly around digital audio workstations— have had a noticeable impact in the development of tonal language in EDM, for such production practises seem closer to cinematographic montage —based on splicing, layering and processing sound files— than to musical operations based on traditional compositional operations on symbolic notation. While this has been demonstrated for other musical parameters (e.g. Butler, 2006, regarding rhythm and structure) and genres (e.g. Spicer, 2004, in pop-rock), the influence of layering and looping in the materialisation of unique tonal layouts —and the ways listeners integrate them together— has been paid insufficient attention. Furthermore, since the structural and functional 'obligations' of EDM —centred around dancing and intense emotional exposure— are mostly achieved through rhythmic and timbral means, harmonic, tonal and pitched materials in general could assume a rather open-ended and experimental function in EDM, freed from the musical structuring role typically assigned to pitch, especially in Western classical music.

In order to clarify these potential effects, I decided to embark on a study of tonality in EDM that could inform my research in automatic key estimation. I wanted to attain a descriptive —not prescriptive, or critical— study of tonality in EDM, in line with what Meyer had described as "style analysis" (1973, pp. 6–9), by identifying a group of idiomatic tonal configurations, distinct from other genres, and statistically observable.

I also wanted that my observations could directly benefit the makers —at least beginners and amateurs— in the form of computational methods that could eventually be offered as compositional aids or classification tools in digital creative environments, consummating a feedback loop between my interest in musical analysis and my desire to promote musically driven MIR research.

1.4 Actions in Context

In line with the research goals defined in the previous section, my report on the original outcomes contained in this manuscript is divided into two different chapters, corresponding to musical analysis and music information retrieval, respectively. I have tried to establish a nutrient dialogue between both areas throughout my research: musical analyses have informed the key detection methods proposed, and various MIR techniques have been intended to support analytical enquiries. However, for the sake of clarity, findings in either domain are discussed separately.

1.4.1 A Study of Tonal Practices in EDM

As a first set of contributions, I discuss a series of novel tonal configurations, that —I believe— are a consequence of several interrelated factors. These include:

- The use idiosyncratic production techniques and technologies, revolving around playing and mixing records in the first place, and directed towards the digital audio workstation (DAW) afterwards.
- A generalised lack of directional dynamics and other tonal artefacts, such as chord sequences, cadential points, and key changes, in connection with the cyclical and repetitive structure of the music.
- A shift in importance of pitch structures, from playing a primary role in other musical genres, to occupying a secondary —and sometimes decorative— position in EDM.

Moreover, I suggest that such tonal manipulations are conscious compositional elements in EDM, materialising both in the simultaneity of sounds and in their temporal arrangement. In addition, perhaps as a methodological side-effect, this study also contributes to the research community with the following evidence:

- Two datasets of two-minute audio excerpts of a variety of EDM subgenres, with global-key annotations, adding up to more than 2,000 labels.
- A collection of 500 musical analyses, including detailed pitch-class set annotations, global tonal labelling, modal changes, characteristic musical features, and verbose descriptions of salient or unfrequent attributes.

I am aware that the study of tonal practises in EDM presented in subsequent chapters is necessarily partial an incomplete. First, I admittedly decided to study tonality within the constraints of what Tagg has called the "extended present" (2012, pp. 272–273), normally assimilated to single musical phrases or sequences. This notion is particularly useful in EDM, and I have equated Tagg's notion to the span of a cycling loop, typically comprising between two and eight measures, and representing a complete musical unit in EDM. This excludes from my study large-scale tonal structures such as DJ sets, focusing on short-term tonal relationships.

Another complication arises from the fact that EDM encompasses a large variety of subgenres, with clearly differentiated attributes. I could have limited my enquiry to a given historical or socio-geographical arena, for example, to studying the role of synthesised basslines in early acid house, or chord sequences in Chicago's soulful house. However, my intention was to identify certain constants across practises and subgenres, even at the risk of providing necessarily vaguer descriptions. Nevertheless, insights on specific genres will appear at several points in the thesis.

Similarly, I could have investigated potentially transversal practises, such as the tonal complexity and transformations between original tracks and remixes, a sort of natural ground for studying tonal variation in tracks with a common origin. Another thrilling perspective would have been to study harmonic complexity from a psychoacoustical stand, based on an intuition that relates intervallic simplicity to the high degree of timbral complexity found in EDM. Fortunately, research addressing harmonic mixing from the perspective of psychoacoustics has been recently initiated by Gebhardt et al. (2015, 2016) and Bernardes et al. (2017a).

Nonetheless —and although modest— I regard my preliminary study of tonal practises in EDM as significant on its own, and wishfully capable of motivating further computational analysis research, creative tools and applications, as well as supplementary musicological study.

1.4.2 Algorithms for Key Estimation in EDM

However, it is important to bear in mind that my study of tonal practices in EDM is mainly oriented towards the development of better informed algorithms for tonal recognition tasks in EDM. As I have already noted, it is frequent to use computational tools to estimate the musical key of audio tracks in EDM production environments, with a direct application in the sequential mixing of musics that are tonally related. Unfortunately, the available solutions only provide a binary major/minor classification, without offering any finer detail regarding modality or tonal ambiguity, what, I think, could be of great utility in the simultaneous mixing of sounds.

In this dissertation, I contribute two methods for automatic key determination, which are evaluated with existing and newly created datasets, and compared to existing state of the art approaches. In varying degrees, the algorithms described in this text are capable of:

- Characterising pieces globally, with a global-key label, improving the performance of most available algorithms in EDM.
- Providing a finer detail, regarding modal ambiguity and tonical ambivalence, enriching the binary major/minor classification with labels accounting for atonal fragments and other modal practises falling out of such binary model.

One of the algorithms developed during my research has already been integrated into a commercial application,[17] obtaining good critical reception.[18,19] The method providing additional verbose, on the other hand, has proved valuable in observing and elaborating on some of the tonal practises that I claim as characteristic of EDM, establishing and interesting dialogue between my two areas of interest.

1.5 Structure of this Dissertation

Following this Introduction, the next two chapters lay down the music-theoretical foundations and scientific background for the remainder of this thesis. In Chapter 2, I introduce the basic musical terminology used throughout my explanation, the concepts of key and tonality, and their particular uses across various musical practises, including EDM. Chapter 3, complementarily, contextualises the domain of music information retrieval, discussing tonality-related research and providing a detailed review of the literature on automatic key estimation.

Chapter 4 represents a turning point in the dissertation, a hinge between the contextual chapters and the original contributions, as illustrated in Figure 1.2. As such, it gives account of the main methodological aspects of my research, including a discussion on available datasets for computational tonal analysis, and a description of common evaluation procedures for key-finding algorithms. Additionally, the chapter advances the description of a newly created data collection, the *GiantSteps key dataset*, which is used in conjunction with already available data to present a preliminary evaluation of existing systems, in order to make an argument supporting the contributions described in subsequent chapters.

[17] http://reactable.com/rotor

[18] https://www.soundonsound.com/reviews/reactable-systems-rotor

[19] http://ipadloops.com/reactable-rotor-tangible-modular-music-synth-for-ipad

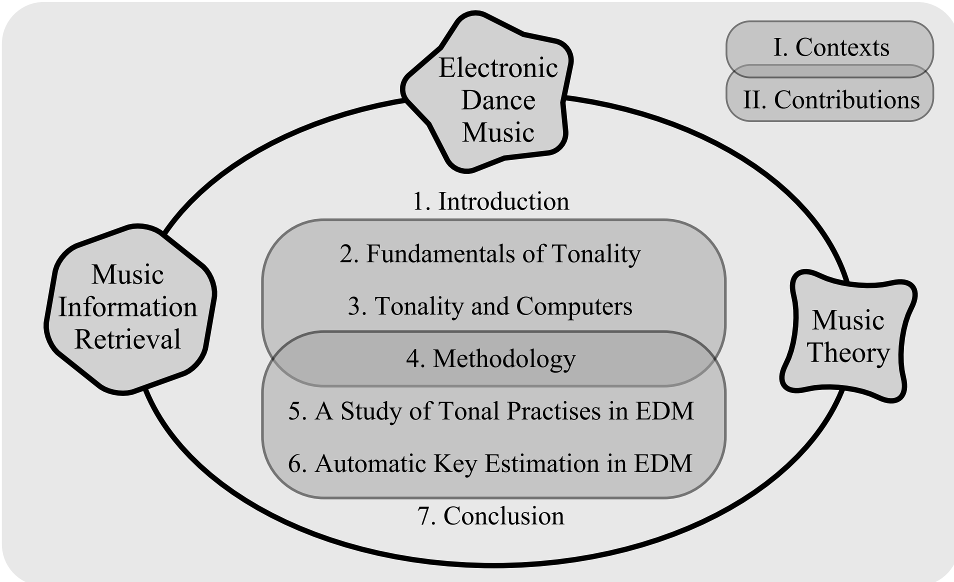


FIGURE 1.2: Overall structure of this dissertation.

In Chapter 5, I share my findings regarding tonal practises in EDM. This report is grounded in tonal analyses of EDM tracks, adding up to over 2,000 audio excerpts with new key annotations, and a detailed analysis of 500 audio excerpts, providing evidence of novel tonal configurations, as well as tracing distinctive tonal behaviours across various subgenres. Chapter 6 builds upon some of the aspects referred in the previous chapter, and describes the contributed methods for key finding in electronic dance music. The discussion unfolds in a bottom-up fashion, from an explanation of low-level signal-processing decisions, through the description of tonality profiles, derived from various corpora of EDM, to a discussion of the scope and degree of descriptive detail of the proposed solutions. I conclude the chapter with a final evaluation, comparing the results of the proposed methods with state of the art key-finding algorithms.

The main body of this work ends with a concluding chapter, where I summarise the contributions discussed herein, and share some of the limitations and difficulties I found during my research, suggesting potentially interesting ways of continuing this work.

For the sake of completion, I have prepared three appendices with complementary information and resources originated in the course of my doctorate. Appendix A presents a lists of peer-reviewed publications to which I have contributed, related to the contents of this dissertation. In Appendix B, I condense the typewriting conventions used throughout this dissertation, and is meant to serve as a quick reference guide. Last, Appendix C points to the materials generated throughout my research, including datasets and computer programs to reproduce the experiments described in this manuscript.

***

In the following chapters, I shall change my voice to the first person plural, for I would not have been able to accomplish this work alone. In that plural voice, resonate the echoes of my supervisors and fellow doctorandi, as much as those of all the people that helped me in unaccountable ways, by pointing at flaws in my discourse, suggesting roads of enquiry, and giving unconditional support. However, this Introduction —as well as the concluding chapter— is narrated in singular person, assuming the complete responsibility for all the opinions contained herein, the organisation and readability of the whole manuscript, and especially, regarding any possible misunderstanding that it could motivate.

Chapter 2

Fundamentals of Tonality

"If you have built castles in the air
your work need not be lost.
That is where they should be.
Now put the foundations under them."
Henry David Thoreau, *Walden* (1854)

This chapter presents the foundations upon which the music-theoretical elaborations and contributions of this thesis are supported. We begin our narration by defining the basic musical terminology that is used throughout this work, before addressing the fundamentals of Western classical tonality in Section 2.2. Section 2.3 examines particular practises across popular music styles, presumably closer to our object of study, which is considered in Section 2.4, with a review of the scarce literature on tonality in EDM.

We have intended to adjust our explanation to the requirements of our research, providing significant music-theoretical background to the extent that it will prove useful when considering tonal characterisation and automatic key estimation in subsequent chapters. For this reason, our report has been intendedly simplified, in order to remain accessible to the reader less familiar with music-theoretical literature.

2.1 Basic Tonal Terminology

Throughout this dissertation, we try to use musical terminology that is both all-embracing and precise regarding the denotation of musical objects and concepts. Philip Tagg (2012, 2013, 2014) has made a significant effort to normalise musical terminology based on notions of cultural equity —across Western and Non-Western musics, popular or with enduring classical traditions— as well as on lexicological and etymological consistence. It is for these reasons, that we incorporate some of his acceptations and neologisms in the lexicon of this thesis, especially those designating tonal aspects, where conflicting terminology mostly appears. This is probably because a specific type of tonality is the most characteristic artefact of Western classical music, functioning as the yardstick upon which any other possible interplay of musical tones is normally considered. Different periods and musical styles have developed different practices of tonality. However, the variety of tonal practises has been often neglected in scholar work —or expressed in terms of cultural inferiority— although this situation is perceptibly changing due to the consolidation of popular music studies. As Tagg puts it,

> "The concepts of tonality circulating in Western academies of music, whatever their canonic repertoire, are still all too often inadequate, illogical and ethnocentric. They simply don't do much to help music students living in a multicultural, internet linked, 'global' world to get to grips with the tonal nuts and bolts of all those musics that don't fit the conceptual grid of categories developed to explain certain aspects of the euroclassical or classical and jazz traditions. [...] The difficulty is that the vast majority of those other musics is under-theorised, in the sense that existing music theory often seems to have either misleading terms or no terms at all to designate their specific tonal dynamics." (Tagg 2014, p. 14)

2.1.1 Frequency, Pitch, Octave

In a broad sense, sound can be thought of as an oscillating pattern of movement within a given medium —be this air, water or other material— that can be perceived by our auditory system. Sound patterns can be divided into periodic and aperiodic signals. Periodic sounds present a repeating or quasi-repeating oscillation through time, whereas aperiodic signals tend to be more difficult to predict. The most important property of a periodic oscillation is its *frequency*, defined as the number of equal-length cycles that a signal completes over a period of time. A convenient measure of frequency is in cycles-per-second (cps), more commonly referred to as Hertz (Hz). Periodic sounds in the range between 20 and 20,000 Hz are experienced by humans as *tones*, with perceived *heights* that change in correlation to their frequency. Figure 2.1a depicts a short fragment of a sine tone at 100 Hz, in its time domain (above) and spectral (below) representations. Sine tones are the simplest periodic oscillations, consisting of only one frequency component. However, musical tones are usually complex oscillations, made up of aggregate *harmonics*, 'children' oscillations at proportional ratios, such as, for example, the sung vowel 'e' shown in Figure 2.1b. Even when these complex oscillations occur —which is most of the times

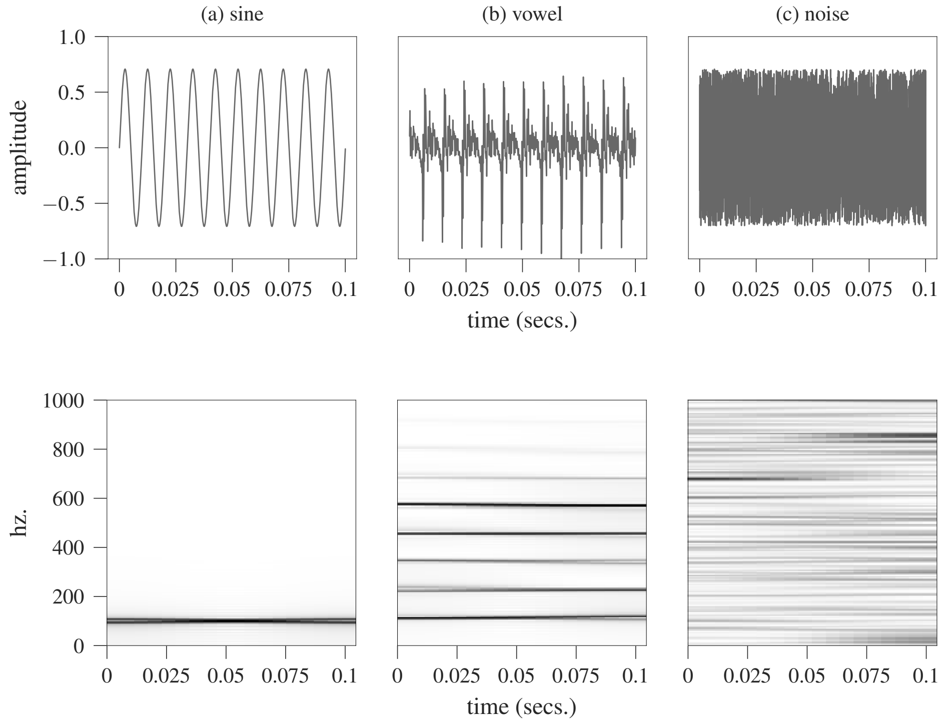


FIGURE 2.1: Time- (above) and frequency-domain (below) representations of periodic and aperiodic signals: the graphs represent 0.1 seconds of (a) a sine tone, (b) a sung vowel 'e' and (c) white noise. The spectrograms have been truncated at 1000 Hz. for visualisation purposes.

in the physical realm— we are still able to perceive one prominent, fundamental tone, typically corresponding to the largest audible period, or inferred psychoacoustically from the signal's components. We refer to the perceptual height of periodic sounds as *pitch*. Since human's cognitive apparatus tends to perceive physical magnitudes with different logarithmic or non-linear curves, pitch is generally not reported in Hertz, but according to conventions accounting for the variety of tones in a given musical milieu. On the contrary, completely aperiodic signals, typically known as *noise*, do not evoke a sense of height, distributing their energy uniformly across the whole spectrum, as Figure 2.1c illustrates.

An important connection between the physical and cultural realms seems to be the fact that tones doubling or halving their frequency are perceived as highly similar, to the extent that they are considered equivalent in most musical systems across the world (Honingh & Bod, 2011). This is especially the case in musical cultures where "men, women and children sing together in unison" (Trehub et al., 2015). In Western musical culture, pitches with a frequency ratio of 1:2 are said to be one *octave* apart.

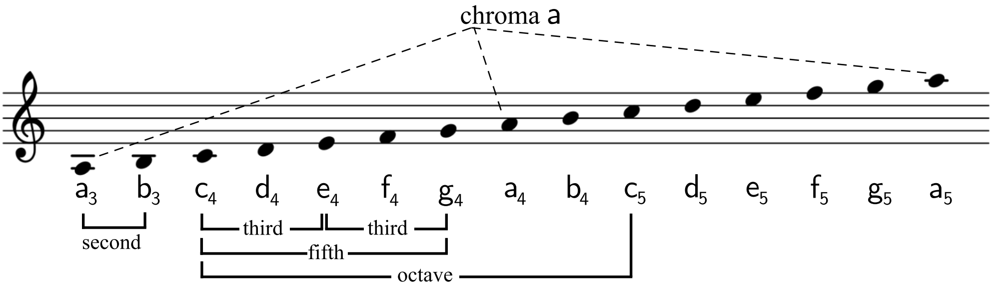


FIGURE 2.2: Two octaves written in Western musical notation, illustrating the relationship between pitch names, octaves and simple intervals.

This denomination comes from the fact that this ratio has been typically divided into seven musically related tones, the subsequent eighth named just like the first, given its perceptual similarity. However, it is worth noting that different traditions divide the octave in a different number of tones.

2.1.2 Pitch, Chroma, Diatonic Interval

In the anglophone world, the musical divisions of the octave are named with the first seven letters of the alphabet (a–g), as illustrated in Figure 2.2.[20] To specify pitches from a particular octave, an index can be added to the pitch name. According to the standard pitch notation, the lowest note of a piano (the reference instrument for Western music theory) is an a_0. However, contrary to what one would intuitively deduce, octave cycles start in c and not in a. The reason for this is grounded in the centrality of c in western musical theory, for reasons that will become apparent in the following paragraphs. Pitch names without octave specification usually designate octave-equivalent families of tones, called *chromas* or *pitch classes* (pc's). For example, the chroma a is made up of a's from across all octaves. Conventionally, the distance or *interval* between two different pitches is given by counting the total ordinal of letters from the first —typically the lowest— to the second. For example, the interval between a and b is a *second*, between c and e, or e and g, a *third*, and between c and g, a *fifth*.

In Western music theory, however, the intervallic distance between consecutive natural pitches is not constant. Consolidating in the Eighteenth Century, Western music widely adopted what is known as the *equal temperament* system, by which octaves are divided into twelve perceptually equal intervals, called *semitones*. The semitone

[20]For the sake of clarity and to minimise confusion between various musical objects, throughout this dissertation we write single pitch names in lower-case sans-serif typeface. This and other typesetting conventions are summarised in Appendix B.

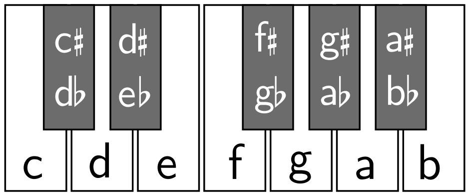


FIGURE 2.3: Distribution of the twelve chromas in a piano keyboard.

is the smallest musical interval to be found in most Western music, corresponding to adjacent keys on a piano, or subsequent frets on a guitar. Intervals between consecutive pitch names are either one or two semitones (i.e. one *tone*) apart[21], and musical orthography must differentiate between a total of twelve different pitch classes with only seven note names. This is achieved by indicating a raise or a decline from the so-called *natural* tones, by appending or prepending the sharp (♯) or flat (♭) symbol, respectively.[22] Figure 2.3 depicts the layout of an octave in a piano keyboard. Natural tones are represented by the white keys, whereas altered pitches correspond to the black keys. In the figure, the leftmost black key can be indistinctly referred to as c♯ and d♭, which are said to be *enharmonically* equivalent tones. The preference for one or other label is normally determined by the specific musical context. For disambiguation purposes, a natural symbol (♮) can be used to cancel the effect of any accidental when accompanying a pitch or chroma.

Figure 2.4, presents, in musical notation, the *chromatic* division of the octave into twelve chromas. Besides pitch names, it is common to refer to pitch classes by using numerical indexes, written above the staff in the figure.[23] This numerical equivalence facilitates and generalises typical musical transformations —especially when using computers— and it has been widely adopted in circles embracing set theory (Forte, 1973; Straus, 2005), as we will shortly see. Throughout this thesis we use duodecimal notation for the representation of pitch classes, substituting 10 and 11 with ↊ and ↋, respectively.[24]

[21]Here we introduce the first of a series of polysemic and potentially confusing terms. Up to this point, 'tone' denoted any sound with a pitched quality. In other common acceptation, the word designates the musical interval comprising two semitones.

[22]Typically, when referring to altered pitches —in written or spoken language— the alteration or *accidental* is reported after the pitch name (a♯). In musical notation, however, the alteration precedes the notehead.

[23]Although semantically equivalent, we normally use 'chroma' when dealing with alphabetical labels, and 'pitch class' to refer to numerical notation.

[24]It is an extended practice to write pc's in duodecimal notation, particularly in computational analysis environments. The most common alphabet assimilates 10 to A and 11 to B; however, since these two letters can be mistaken with pitch names, we follow the convention proposed by The Dozenal Society of Great Britain, included in the Unicode standard (http://www.dozenalsociety.org.uk).

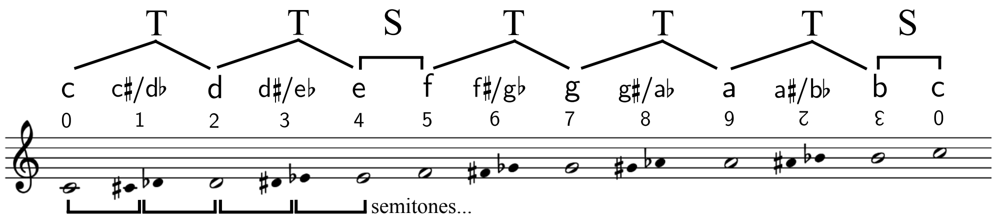


FIGURE 2.4: Musical notation of an octave divided into 12 semitones, with labels showing the chroma names and pc-integers in duodecimal notation (↊ = 10, ↋ = 11). The intervallic distance between the natural pitches (tone [T] or semitone [S]) is also shown.

2.1.3 Scale, Interval Quality

In previous paragraphs, we discussed pitches as single abstract units, without reference to any particular musical context. In most pitched-centric musics, the basic vocabulary to establishing such a context is given by the *scale*. A musical scale can be thought of as a palette of 'available' chromas, arranged alphabetically, typically in ascending order. The starting pitch of such stepwise ordering holds the most important position in the scale, providing it with a referential name, and normally occupying the gravitational centre around which all other chromas are arranged. This notion of musical organisation around a central note, the *tonic*, has been referred to as *tonicality* (Reti, 1958; Tagg, 2014), and is the foundation over which the broader concepts of key and tonality are grounded, as we will see in Section 2.2.

Besides its tonal centre, the most distinctive feature of a scale is its particular pattern of intervals. If we consider the sequence of all the notes in Figure 2.4, we observe a pattern of eleven successive semitones, forming what is known as a *chromatic* scale, since it contains all the available 'colours' in equally tempered music. Nevertheless, scales typically consist of sequences of intervals of variable length, presenting asymmetric patterns. Figure 2.4 also highlights the interval pattern between the white noteheads, presenting a sequence of two consecutive tones, plus one semitone, plus three subsequent tones, plus one last semitone (T–T–S–T–T–T–S), dividing the octave into seven pitches before returning to the initial chroma c. This particular pattern is called the *major* or *ionian* scale, being the most common pitch structure in Western music.[25] Any particular scale is identified by its tonic and its intervallic structure (therefore, the white-notehead scale in Figure 2.4 is named C ionian).[26] Interestingly,

[25] It is far more common to refer to this scale as the major scale. However, we reserve the term 'major' to characterise larger musical contexts —as we will shortly explain— preferring the label 'ionian' to refer to this scalar pattern specifically.

[26] In order to differentiate musical contexts (scales, chords, or keys) from single pitches, we use capitalised pitch names to refer to the former, and lower-case letters to refer to the latter (see Appendix B).

pitch	st.	interval label	relative degree
c	**0**	**perfect unison**	$\hat{\mathbf{1}}$
c♯	1	augmented unison	♯$\hat{1}$
d♭	1	minor second	♭$\hat{2}$
d	**2**	**major second**	$\hat{\mathbf{2}}$
d♯	3	augmented second	♯$\hat{2}$
e♭	3	minor third	♭$\hat{3}$
e	**4**	**major third**	$\hat{\mathbf{3}}$
f	**5**	**perfect fourth**	$\hat{\mathbf{4}}$
f♯	6	augmented fourth	♯$\hat{4}$
g♭	6	diminished fifth	♭$\hat{5}$
g	**7**	**perfect fifth**	$\hat{\mathbf{5}}$
g♯	8	augmented fifth	♯$\hat{5}$
a♭	8	minor sixth	♭$\hat{6}$
a	**9**	**major sixth**	$\hat{\mathbf{6}}$
a♯	10	augmented sixth	♯$\hat{6}$
b♭	10	minor seventh	♭$\hat{7}$
b	**11**	**major seventh**	$\hat{\mathbf{7}}$
c	**12**	**perfect octave**	$\hat{\mathbf{8}}$

TABLE 2.1: Typical musical intervals and/or relative scale degrees from c. In bold font, we emphasise the degrees of the ionian scale, which are either major or perfect. For other intervals, note that the same distance in semitones has different denotations depending on the start and end alphabetic pitch names.

this scale in particular is formed by all and only the natural chromas, what may partially explain why c has consolidated as the reference pitch for music theoretical explanations instead of a, for example.

In Figure 2.2, musical intervals were labelled as ordinal numbers counting the simple distance between two pitches. Accordingly, the interval between c and d is a second, just as much as the distance between e and f. However, the scale in Figure 2.4 shows that these two distances comprise two and one semitones, respectively. Similarly, the thirds c–e and g–e take in four and three semitones, whereas thirds c–e♭ and e–g♯ present three and four semitones. To solve this ambiguity (same alphabetic distance, different semitonal interval), a clearer labelling of intervals can be achieved by adding the *major*, *minor* and *perfect* interval types to the basic distances, differentiating major seconds (c–d) from minor seconds (e–f), minor thirds (c–e♭) from major thirds (c–e). In order to measure these compound distances, the ionian scale is usually taken as a reference, given that all the intervals between the tonic and the other notes in the scale are either major or perfect. All other intervals falling out of the ionian scale are

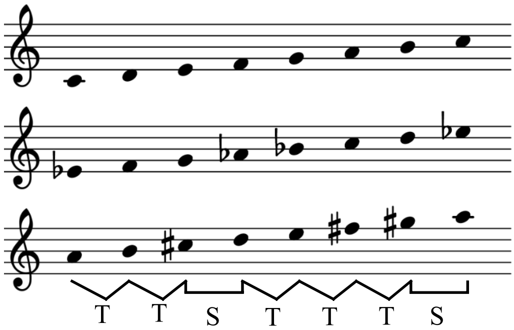


FIGURE 2.5: Ionian scales from three different tonics: C ionian (top), E♭ ionian (middle), A ionian (bottom). In order to keep the same intervallic pattern, the use of accidentals becomes necessary.

either minor intervals (when a major interval is lowered one semitone), diminished or augmented (i.e. when perfect intervals, typically fourths and fifths, are lowered or raised one semitone, respectively). Table 2.1 lists the most common intervals counted from c, expressing their distance in semitones, their labels and their scale degree. The latter notation is extremely useful to describe pitch patterns in relative terms, highlighting their musical quality without a reference specific pitches. Relative scale degrees and intervals are indicated with circumflex accents on Arabic numerals, following a widespread convention. A flat (♭) preceding the number indicates a minor or diminished interval, whereas a sharp symbol (♯) defines an augmented step (e.g. $\hat{1}$, ♭$\hat{3}$, ♯$\hat{4}$, ♮$\hat{7}$).

2.1.4 Transposition, Rotation, Mode

Figure 2.5 shows the results of *transposing* the C ionian scale to other tonics, obtaining the scales of E♭ ionian and A ionian. Note that in order to preserve the ionian intervallic pattern, different alterations are used. These scales, although providing different tonal contexts (C, E♭, A) still convey the same ionian mood or quality. The operation called transposition merely consists in adding a constant interval to a collection of pitches. In general, we can think of transposition as an operation that preserves the same musical character, only that with a different collection of pitches.

Another typical operation is scalar *rotation* (called ‘diatonic transposition’ in traditional music theory). Rotation implies that pitches in a scale are shifted circularly, maintaining the same ordered collection starting at different points. A scale has as many rotational variants as chromas. These variants are typically called *modes*, although in practice, there is absolutely no difference between a scale and a mode. Both are alphabetically ordered sequences of pitches, dividing the octave in a number of intervals. Figure 2.6 illustrates the effects of rotation (left) and transposition (right),

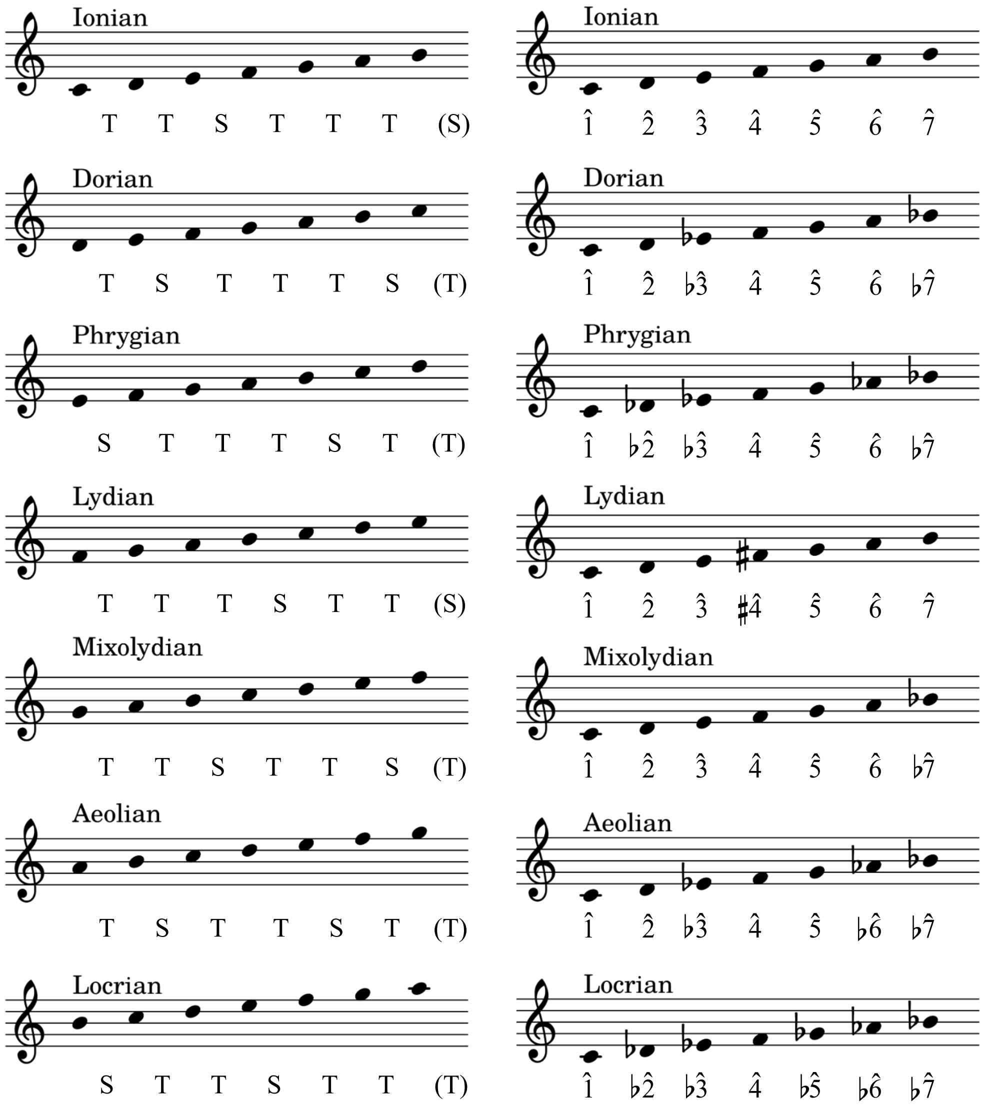


FIGURE 2.6: The seven diatonic modes, written as subsequent rotations of the C major (ionian) scale (left) and transposed to c (right). Each pattern provides a distinctive musical quality due to the intervallic relations with the tonic.

introducing the seven modes of the *diatonic* collection —yet another acceptation to refer to the ionian pattern. These diatonic modes are also known as the 'greek' modes, due to a misreading of Hellenic music theory by Mediaeval scholars (Bower, 1984), and each iteration in the circular shift receives a demonym from a former greek region (ionian, dorian, phrygian, lydian, mixolydian, aeolian and locrian). What is most important is that each rotation of the scale produces a different intervallic pattern, leading to a new scale type. On the contrary, the effect of transposition keeps the same intervallic structure across different tonic notes. Left and right sides of Figure 2.6 show how rotation of the natural pitches (in the left) alter significantly the scale. This becomes visible in the right column, where all modes have been transposed to c. The accidentals accompanying the notes show the deviations from the ionian pattern. Below each pattern, we annotate the different interval sequences (left) and their scale degrees (right). Analogously, Table 2.2 presents the seven diatonic modes arranged by pattern similarity, so that each mode differs with the preceding and succeeding scale in only one pitch (in grey), where the distinctive character of each mode lays. Ionian and aeolian scales are highlighted, for they have served as models to explain Western tonal practise, assimilated as the major and minor modes, respectively. However, the layout of Table 2.2 suggests that at least three different modes could provide a sense of 'majorness' or 'minorness', based on the third scale degree, as we will shortly see.

2.1.5 Modality, Key, Chord

Western tonal theory has been typically explained as divided into two different *modalities*, known as *major* and *minor*. The notion of modality implies a context beyond a specific scale, comprising a number of procedures by which a tonal centre is established and articulated, including particular phrasings and specific sequences of musical objects.[27] A related concept, the idea of *key* refers to the materialisation of a specific modality from a particular tonal centre. Therefore, just as a scale is defined by a pattern and a tonic, a key implies a tonic and a given modality. Consequently, Western tonal theory recognises twenty-four possible keys, half major, half minor.

Although any particular key can be suggested by a simple sequence of single tones, Western music has been characterised by its polyphonic nature, that is, by the use of simultaneous pitches conveying various degrees of *consonance*. Consonance is mainly a perceptual magnitude (sensory consonance), although with elements of cultural construction (musical consonance), indicating the degree by which musical intervals appear as 'pleasant' to the ear (Terhardt, 1974). Typically, the key of a piece of music will be suggested by its melody as much as by its polyphonic units. A good synthesis of the workings of polyphony can be found in the musical discipline of *harmony*, which summarises the unification of tonal practise in *euroclassical* music, a term coined by Tagg (2014) to designate European music from the so-called 'Common Practise Era', roughly spanning from 1600 to 1900, and comprising the consolidation, development and crisis of European tonal language, as traced in the oeuvre of composers such as Haydn, Mozart, Beethoven or Brahms. Numerous

[27] We deliberately use the term 'modality' to establish a semantic difference with 'mode', which simply refers to an ordered collection of chromas.

overall quality	*mode name*	*scale degrees*						
	Lydian	$\hat{1}$	$\hat{2}$	$\hat{3}$	$\sharp\hat{4}$	$\hat{5}$	$\hat{6}$	$\hat{7}$
major	**Ionian**	$\hat{\mathbf{1}}$	$\hat{\mathbf{2}}$	$\hat{\mathbf{3}}$	$\hat{\mathbf{4}}$	$\hat{\mathbf{5}}$	$\hat{\mathbf{6}}$	$\hat{\mathbf{7}}$
	Mixolydian	$\hat{1}$	$\hat{2}$	$\hat{3}$	$\hat{4}$	$\hat{5}$	$\hat{6}$	$\flat\hat{7}$
	Dorian	$\hat{1}$	$\hat{2}$	$\flat\hat{3}$	$\hat{4}$	$\hat{5}$	$\hat{6}$	$\flat\hat{7}$
minor	**Aeolian**	$\hat{\mathbf{1}}$	$\hat{\mathbf{2}}$	$\flat\hat{\mathbf{3}}$	$\hat{\mathbf{4}}$	$\hat{\mathbf{5}}$	$\flat\hat{6}$	$\flat\hat{\mathbf{7}}$
	Phrygian	$\hat{1}$	$\flat\hat{2}$	$\flat\hat{3}$	$\hat{4}$	$\hat{5}$	$\flat\hat{6}$	$\flat\hat{7}$
diminished	Locrian	$\hat{1}$	$\flat\hat{2}$	$\flat\hat{3}$	$\hat{4}$	$\flat\hat{5}$	$\flat\hat{6}$	$\flat\hat{7}$

TABLE 2.2: Characteristic scale degrees and relationships in the diatonic modes. The scales are ordered to maximise their similarity, presenting only one different degree across subsequent rows (in grey). The modes are further divided into major, minor and diminished, according to their overall quality. Scale degrees in grey provide the distinctive musical character of each mode.

handbooks on harmony have formalised the construction of pitch-aggregates and their timely succession, since the publication of Rameau's treatise in 1722 (e.g. Rameau, 1971; Schoenberg, 1974; Piston, 1991).

Abstract polyphonic units are normally referred to as *chords*. Although in its broader acceptation a chord is virtually any aggregate of two or more pitches at any intervallic distance, most musical practises have favoured chordal systems based in similar interval types, either fourths or fifths, but most notably, in the aggregation of thirds.

Figure 2.7 shows the seven diatonic *triads* of C major, obtained by stacking thirds over the tones of the ionian scale. Triads (chords made of three pitch classes) are the basic units of Western harmony, although there is no theoretical limitation to pile up as many thirds as desired, other than exhausting the chromas in the scale. A closer look at the various triads in the figure, reveals that their internal structure differs slightly, based on the differences between intervals of the same basic distance mentioned above. For instance, the first chord in Figure 2.7 {ceg} presents a structure of 4 + 3 semitones, constituting a major triad ($\hat{1}$, $\hat{3}$, $\hat{5}$). Alternatively, the second aggregate {dfa} represents a minor chord, with a pattern of 3 + 4 semitones ($\hat{1}$, $\flat\hat{3}$, $\hat{5}$). There is yet another unique chord type in Figure 2.7, forming on the seventh degree, resulting from two stacked minor thirds. This chord type is known as a diminished chord, because of the $\flat\hat{5}$ interval between the extreme notes of the triad. Chords are labelled after their *root* note, that is, the lowest pitch in the stack of thirds, and their chord type (e.g. Cmaj, Dmin, Bdim).[28] They can also be referred in relative notation using Roman numerals, taking the ionian scale as a reference, just like with regular

[28]We use chord labels in abbreviated form in order to differentiate them from key names, e.g. C major (key) vs. Cmaj (chord). Appendix B lists the abbreviations used, and summarises this and other writing conventions.

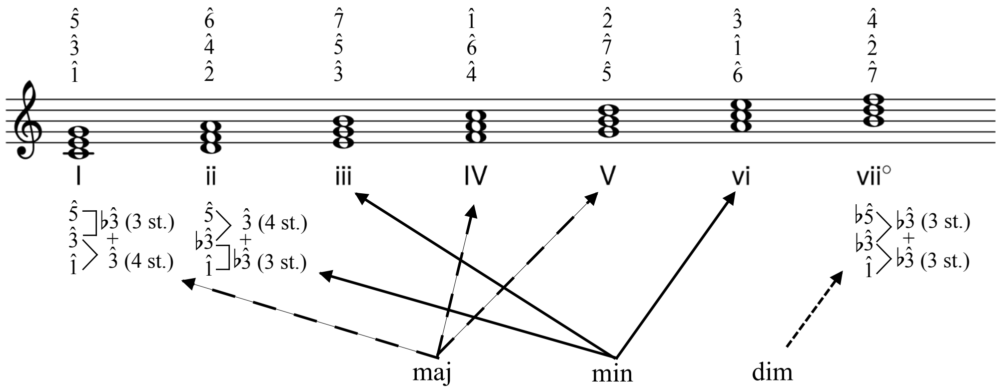


FIGURE 2.7: The seven diatonic triads of C major. The scale degrees that form each chord are written above the staff. Below, Roman numerals express the type of each chord, which is detailed under the first occurrence of each type.

intervals. In such cases, letter capitalisation differentiates major and minor chords, and a numero sign ('°') indicates a diminished chord. The chord-type distribution in Figure 2.7 applies to any major context, which natively presents one diminished chord (vii°), three minor chords (ii, iii and vi) and three major chords (I, IV and V). The latter ones are also called the 'tonal chords' for being arguably the most important elements in establishing a musical key, and are normally referred to as the tonic, subdominant and dominant chords, respectively.

A reason of the predominance of ionian tertian harmony in Western music theory can be given in the light of three facts: First, euroclassical and pop music have been predominantly written in major modality —as much as 60 to 75% of the repertoire, depending on the source (Krumhansl, 1990, pp. 62–75). Second, major modality is quite straightforward regarding the theoretical formation of its elements and their musical materialisation; on the contrary, minor modality is normally taken as subsidiary of the major, and needs multiple scales and 'exceptions' to adjust the theory to the practise, as we will see in Section 2.2.2. A third reason —contributing to the previous two— is to be found in the inner structure of musical sounds. As stated at the beginning of this primer, most pitched sounds are the result of complex oscillations composed by a number of harmonics. Figure 2.8 shows the harmonic series of a hypothetical g2. More exactly, it presents its first eight harmonics, together with their scale degree (taking g as the tonic) and their approximate frequency. The progression of frequencies illustrates the 'harmonic' quality of the signal, showing that frequency components are integer multiples of the fundamental frequency. Furthermore, the example illustrates that each component of this hypothetical g2 could be perceived as a definite pitch, in which case, the first six harmonics correspond to the components of

harmonic num.	1	2	3	4	5	6	7	8
pitch	g_2	g_3	d_4	g_4	b_4	d_5	f_5	g_5
degree	$\hat{1}$	$\hat{1}$	$\hat{5}$	$\hat{1}$	$\hat{3}$	$\hat{5}$	$\flat\hat{7}$	$\hat{1}$
frequency (Hz)	$\approx$100	$\approx$200	$\approx$300	$\approx$400	$\approx$500	$\approx$600	$\approx$700	$\approx$800

FIGURE 2.8: Harmonic series of a theoretical g_2, indicating the approximated equal-tempered pitches, the musical interval with the first harmonic, their approximate frequency and the harmonic number.

the tonic major triad (forming intervals related by octave, perfect fifth and major third with the fundamental frequency). This probably has had an influence in consolidating the major triad as a the most stable tonal aggregate since Rameau.

2.1.6 Pitch-Class Set Operations

We conclude this first section on basic musical terminology by proposing a slightly different approach to looking into pitch collections, based in the so-called pitch-class set theory. Drawing from twelve-tone composition theory by composer Milton Babbitt (1955), pitch-class set theory has been thoroughly formalised by Forte (1973), gaining widespread acceptation across analytical circles worldwide. Although initially conceived as an analytical device for early Twentieth Century music —typically referred to as 'atonal' music— it has proved a powerful tool to study 'post-tonal' musical expressions, such as the works of the Repetitive Minimalists and other tonal practises that do not conform with Western tonal standards (Straus, 2005). Accordingly, it shall prove useful when approaching the study of reduced pitch collections in electronic dance music.

A pitch-class set (pc-set) is simply a collection of unique pitch classes, numerical indexes representing the twelve chromas.[29] Just like chords or scales, pc-sets can be ordered and manipulated in various ways by rotation, transposition or inversion. Ordering a pc-set involves arranging it in ascending order. Typically, an ordered set is expressed in its most condensed form, with the smallest interval between the first and last notes, representing the pc-set's *normal order*. If there are various possibilities meeting this condition, the arrangement with smallest intervals at the beginning is chosen as the normal order. For example, the scale of C ionian, is represented the pitch class set {024579↋}, which is expressed in its normal order as {↋024579}, since this is the expression that minimises the distance between the edges (10 semitones) with the smallest intervals towards the beginning (S–T–T–S–T–T–T). The main power of pc-set expression lays in its capacity of abstraction, allowing comparisons between

[29] Recall that pitch classes 10 (b♭) and 11 (b) are respectively written as ↊ and ↋, as explained in Section 2.1.2 and Appendix B.

interval	0	1	2	3	4	5	6
inverse	0	Ɛ	Ƨ	9	8	7	6

TABLE 2.3: Pitch-class inversion equivalence.

different collections of pitches. The most common way of classifying pc-sets is by reducing them to their *prime form*. According to Forte (1973), a prime form is a pc-set in normal order, transposed so that its first pitch-class is 0. Transposing a pc-set is just a matter of adding a constant number of semitones to all the elements in the set, and calculating its modulo 12, since pc's only comprise the octave range. Therefore, we obtain the prime form of the diatonic set by adding one semitone to its normal order: $(\{1\} + \{\text{Ɛ}024579\})\ mod\ 12 = \{013568\text{Ƨ}\}$.

It is important to notice that this prime form (corresponding to the locrian pattern) represents all the modal variants of the diatonic scale, neutralising the effect of intervallic rotation and therefore of modal differentiation. Pitch-class set operations normally de-emphasise the tonicality of pitch collections —is not for nothing that the theory originated to describe *atonical* music. Another common operation is *inversion*, which literally consists in calculating an interval in its opposite direction. However, given that pitch-classes are always positives in range 0–Ɛ, the inversion of a pitch-class can be seen as a substitution with its inverse or complementary interval, obtained by subtracting the interval to twelve $(12 - i)$, as shown in Table 2.3.[30]

2.2 From Key to Tonality

Up to this point, we have presented basic terminology that hopefully will prove useful when discussing aspects of tonality throughout this dissertation. For a matter of focus, we have decided to concentrate mostly on pitch aspects —the prime matter of tonality— excluding from the discussion essential musical parameters such as rhythm or form, that will be addressed only when they become necessary to our explanation. Similarly, the descriptions provided in this section must be taken as an overview of an otherwise enormously arborescent topic, with unaccountable publications and speculative perspectives. In any case, we have tried to present theoretical notions that will recur in subsequent chapters, either to formulate our hypotheses or to ground our criticism.

[30] Strictly speaking, pitch-class set theory considers inversionally equivalent sets as identical. However, for practical reasons, throughout this text we only consider transposition and rotation as identity operations.

2.2.1 'Tonality' Under Suspicion

The notion of tonality is definitely one of the most prominent concepts across Western musics —and in a great deal of non-Western cultures too, although perhaps under different denominations. In its broadest sense, it defines the systematic arrangement of pitch phenomena and the relations between them, especially in reference to a main pitch class called the tonic (Hyer, 2012). However, the influence of scholar literature, articulated around the currency of tonality, may have hindered the study of other musics with different structuring paradigms. Musics, for example, without a predefined system of pitch relationships and motifs, like post-minimalist drone music as introduced by La Monte Young, or non-idiomatic free improvisation as defended by Bailey (1993). Furthermore, other types of music such as Japanese *taiko* drumming, countless *musique concrète* compositions, or some subgenres of electronic dance music, like minimal techno, do not present definite pitch at all.

We think that in its most basic sense, the term 'tonal' should be used to denote any music made with tones, that is, with perceivable pitch units, in opposition to *atonal* music, made with various kinds of un-pitched elements. This differentiation is important for our purposes, since we will be dealing with a type of music —EDM— that is both tonal and atonal —in these acceptations— presenting sections with just special effects or spoken voices, and sometimes even whole tracks composed only with percussive elements, with sparse or none pitch content at all. Therefore, 'music made with tones' establishes a clear baseline to think of 'tonal music', constituting the basic requirement for our study of tonality in EDM and one of the features that will help us identify some EDM subgenres. Consequently, we prefer the term *tonical* (Reti, 1958) to denote music where pitched elements suggest the presence of one or more pitch centres, like euroclassical music with its major and minor modality, but also including all other modal practises. Similarly, the word *atonical* denotes music that is composed with pitch elements, but does not convey a sense of tonic centrality. An example of atonical music is what has been typically identified as 'atonality', epitomised in the works of the Second Viennese School and serialism.[31] Perhaps ironically, a large body of post-serial music, such as Lachenmann's *musique concrète instrumentale*, happens to be atonal in the more etymologically appropriate acceptation suggested.[32]

Other interesting label is *pantonality* (Reti, 1958), proposed to precisely recognise tonal relationships in sequences of pitches, intervals and chord sequences with changing tonics, without appealing to the structural implications of tonality that we

[31] Although it seems that Schoenberg himself disliked the term, according to Whittall (2011).

[32] Most of the novel vocabulary presented in this section is borrowed from Tagg (2014). We point the reader to his writings for a thorough explanation of the concepts introduced (especially tonal/atonal and tonical/atonical), accounting for etymological, lexicographical and otherwise musicological reasons.

describe in the following block, and even not conveying a key centre in any large scale sense (Drabkin, 2012). In this sense, pantonality acknowledges other ‘tonal’ practises such as free atonalism or even twelve-tone composition, as alternative methods to organise pitch relationships, although most significantly embraces other Modernists approaches to tonality, such as the parallel chord streams of Debussy or the polytonal practises of composers like Stravinsky or Casella. *Polytonality* has been an issue of discussion too. In theory, polytonality is the presence of more than one tonality (tonic and mode) operating at the same time. However, authors like Van der Toorn (cited in Krumhansl, 1990; Tymoczko, 2002) have questioned the possibility of perceiving two keys simultaneously, favouring complex-scale interpretations in relation to Stravisnky’s *Petrouchka Chord* (an aggregate of Cmaj and F♯maj). Krumhansl (1990, pp. 226–239) offers an empirical discussion on this issue (using the Petrouchka chord as source to her experiment) with results suggesting that listeners can actually recognise the importance of both tonal centres. In our opinion, the fact that the polychord under consideration presents two triads one tritone apart (dividing the octave most neutrally in two equal intervals) plays an unacknowledged role in the experiment, and we look forward to experimental results with less neutral intervallic relations (e.g. major seconds). In this same context, Tymoczko (2002, p. 83) brings in the term *polyscalarity*, as a conceptual midterm acknowledging Stravinsky’s intention of using two different modes without falling into the perceptual puzzle. This notion of multiple scales operating simultaneously will be of utility in explaining some tonal configurations in EDM, arising from the combination of different musical sources. However, in the following blocks, we return to more restrictive notions of tonality, in order to explain the basic workings of euroclassical music and rock modality.

2.2.2 Major/Minor Duality in Euroclassical Music

In her research on tonality perception, Krumhansl (1990) conceives tonal music as indicating a musical organisation around a reference chroma, where harmony plays an important role in establishing such sense of tonical centricity. In this acceptation, primarily monodic music, such as gregorian chant and many manifestations of folk music are excluded, even though they obviously present similar scalar constructions and a clear reference to a tonic note. These ‘other’ tonical manifestations, are normally referred to as *modal*, denoting a somewhat vaguer definition of the relationships between a pitch-class set —a mode— and the tonal centre it suggests. Nonetheless, Krumhansl’s dual definition of tonality corresponds to the original usage of the term, first appearing at the beginning of the Nineteenth Century, intended to establish a difference with previous polyphonic practises (Hyer, 2012).

The fundamentals of euroclassical major/minor tonality can be distilled from three influential facts, according to Whittall (2011). First, Rameau (1971)[1722] systematised the principle of inversion, by which chords are composed of stacks of thirds, and defined in terms of their root and type independently of their lowest note (typically perceived as the supporting note of the chord). With this operation, Rameau made equivalent all possible orderings of the same chord set (i.e. {ceg} and {gec}), something novel at that time. This equation allowed, by the beginning of the Nineteenth Century, to substitute figured-bass indications with Roman numerals associated with chords, defining a closed chord vocabulary connected with each key. Last, Hugo Riemann (1903) systematised in his theory of functional harmony, the fundamental role of *tonal functions* in establishing a tonality. Tonal functions are mainly assumed by the tonic (I), dominant (V) and subdominant chords (IV), whose intervallic distances one fifth above and below the tonic, respectively, provide an equidistant tensional arrangement balanced around the tonic chord.

Thus, euroclassical tonality is essentially grounded in a sense of musical directionality, obtained by the succession of the basic tonal functions, and materialised in chord progressions and cadences. These two elements are the actual narrative forces of euroclassical music. Cadences represent arrival points, interruptions of the rhythmic flow associated with the structural organisation of the music, and taxonomised almost as if they were rhetoric figures. The choice of chords in a sequence, on the other hand, determines the character, mood and directionality of a given excerpt, greatly connected with its modality. The major modality is essentially defined by the ionian scale, with its characteristic intervals, chords and tonal functions. However, minor modality shows a bit of resistance when inserted in the narrative scaffold of euroclassical functional harmony, at least, when assumed as a natural diatonic minor, assimilated with the aeolian scale presented in Figure 2.6.

Figure 2.9a presents a major *perfect cadence*. This simple four-chord structure, summarises the basic workings of tonal functional harmony: A tonic chord (I, representing tonal stability) progresses towards a subdominant chord (IV, normally associated with a state of intermediate tonal tension); then, the subdominant function continues onto the dominant chord (V, the maximum exponent of functional tension), which finally resolves back into the tonic. Naturally, these tonal dynamics are often extrapolated to different chords and sequences. However, the tonal forces expressed in this example are always present in euroclassical composition. In contrast, Figure 2.9b is presents a minor cadence rarely seen in euroclassical music, although it essentially presents the same structure. The three minor chords used (i, iv, v) are derived from the aeolian scale. The fundamental 'problem' of this cadence, regarding tonal harmony, is located in the 'tone' resolution (g$\rightarrow$a) in v$\rightarrow$i, contrasting with the semitone movement present in the perfect cadence shown in Figure 2.9a, where it lays a great deal of

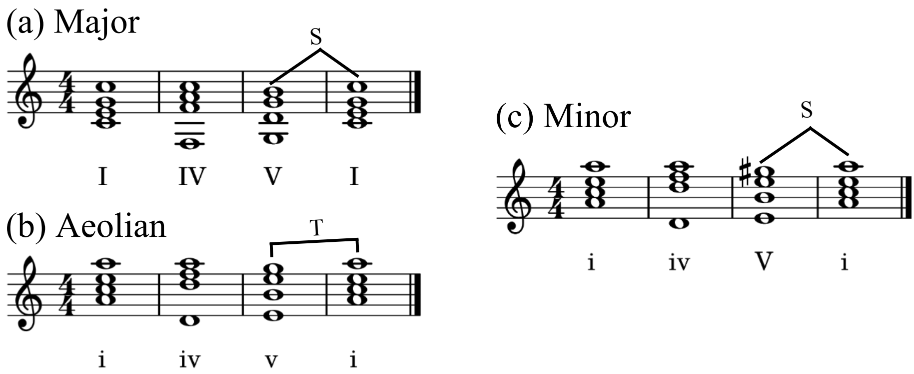


FIGURE 2.9: Cadences for various modal configurations. The major perfect cadence (a), an aeolian cadence (c), and the more frequent minor cadence (c), with a major V chord. We provide four-part realisations to illustrate the normative voice-leading in euroclassical practise.

the dominant tension. Due to the lack of the ‘leading tone’ ($\natural\hat{7}$), a v can hardly be considered as a dominant function, thus dismantling the essence of tonal functional narrative. Alternatively, Figure 2.9c presents the typical euroclassical minor cadence, where v is substituted by a major chord ‘borrowed’ from the parallel key of A major (V), re-establishing the missing semitonal resolution (g$\natural\rightarrow$a). This operation reflects well the bias towards major modality present in euroclassical tonality.

Another indicator of the difficulty to fit the minor modality within the tonal-functional building can be found in the variety of scales to explain it. Figure 2.10 shows three variants of the A minor scale. On top, the aeolian scale represents the natural tones, obtained by rotation of the diatonic set, and representing the closest counterpart to the C ionian scale. A common variant is the so called *harmonic* scale, which basically raises the seventh degree, creating a semitone interval with the tonic ($\natural\hat{7}$–$\hat{1}$). Unfortunately, the harmonic scale introduces a melodic problem, since the interval $\flat\hat{6}$–$\natural\hat{7}$ comprises of three semitones, providing an unwanted ‘exotic’ sound to this scale. In order to overcome this problem, the *melodic* scale in the bottom presents an alternate pattern, borrowing the $\natural\hat{6}$ and $\natural\hat{7}$ degrees from its ionian homonym, in order to be able to resolve to the tonic by semitone, and avoiding the augmented interval present in the harmonic scale. However, since resolution is no longer needed when leaving the tonic, the descending pattern of the minor melodic scale is essentially the aeolian mode. Theoretically at least —and according to their informative names— the harmonic scale is used to compose chord sequences in minor modality, complying with the impositions of tonal harmony; the melodic scale, complementarily, is used melodically to convey a minor feeling without alien intervals. In practice, however, the aeolian scale also plays a role in euroclassical harmony.

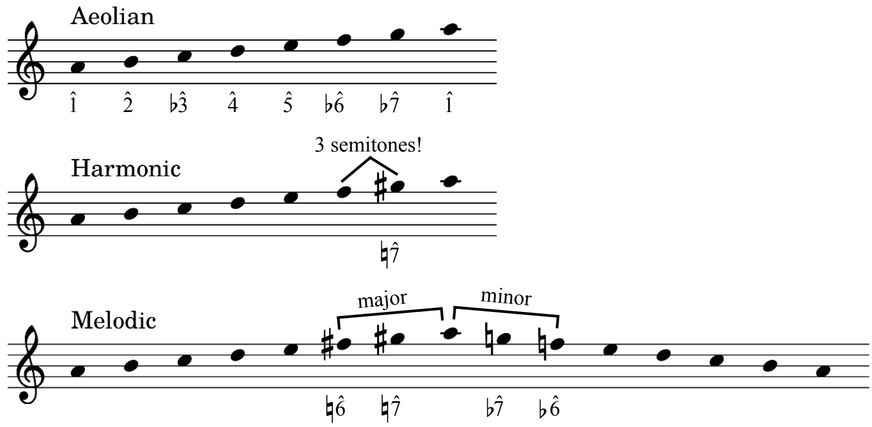


FIGURE 2.10: The three typical variations on the minor scale: A aeolian, A minor harmonic, with its raised seventh degree ($\natural\hat{7}$), and A minor melodic, with different ascending and descending patterns.

Figure 2.11 shows the triads derived from the aeolian (dark grey) and harmonic (light grey) scales. The chords depicted in black, mostly representing subdominant functions, are found in either scale. The differences arise on the degrees that can potentially convey a dominant feel, namely the third, fifth and seventh degrees. The harmonic scale brings in two chords borrowed from the major modality (V, vii°), plus a new chord type, an augmented triad ($\flat$III$^+$), which is hardly used in real music. Additionally, the aeolian scale contributes three chords: $\flat$III, which is extensively used, $\flat$VII, which is only occasionally used, and v, rarely found in euroclassical tonality.

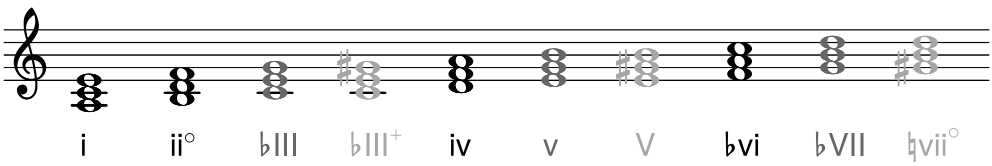


FIGURE 2.11: Triads of two combined minor scales. Chords in darker grey represent triads originating in the aeolian scale, whereas the ones in lighter grey come from the minor harmonic scale. The affected scale degrees are the third, fifth and seventh, potentially associated to tonal-dominant functions.

2.2.3 Key Relationships

In the previous block we have already suggested some possible relationships between different modes and keys. We have noted the transfusion of chords from a major modality into its homonym minor key, for the purposes of tonal resolution. Two homonym keys are said to be *parallel*, sharing the tonic ($\hat{1}$), subdominant ($\hat{4}$) and dominant ($\hat{5}$) degrees, while presenting differences in the other so-called modal degrees (e.g. C major and C minor). These common elements, and especially the sharing of tonal centre, make these keys perceptually related despite the amount of different tones in their respective scales is apparent (4 degrees). Other common connection is established between two keys that, although differing in tonic note, share the notes of their respective diatonic sets. This relation occurs, for example, between the keys of C major and A minor, whose respective modes (C ionian, A aeolian) are identical regarding their pc-set $\{\mathcal{E}024579\}$. These two keys are said to be *relative* to each other, and transitions between them in the course of a musical piece are extremely common. Comparing elements between pitch-class sets is a simple and effective method to assess the degree of similarity or 'closeness' between two keys, relating their *diatonic distance* with the number of common pitches between the two keys, as has been observed by music theorists (e.g. Schoenberg, 1974; Lerdahl, 2001). Significantly for tonal harmony, keys sharing six-out-of-seven pitches are located a fifth apart, what establishes a powerful connection between the tonal chords and the tonal *regions* of the dominant and subdominant. Distance relationships between different keys, chords or pitches have motivated the creation of various *tonal spaces* throughout the History of Music (Lerdahl, 2001). The 'circle of fifths' or 'regional circle' is a well-know of such tonal spaces, illustrating the relationships between relative and neighbouring keys. Figure 2.12 shows two of the first regional circles published, by Heinichen (1728) and Kellner (1732). While previous representations showed an alternation of major and minor keys (the first regional circle known is attributed to Diletzky, 1679) Kellner seems to be the first to separate major and minor keys in two concentric circles.

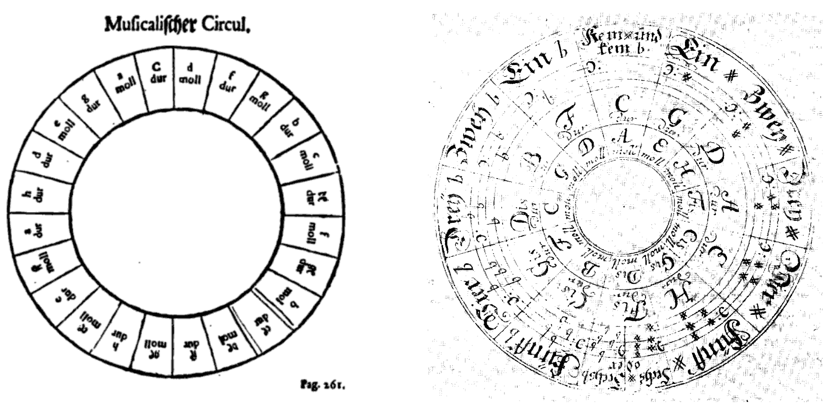


FIGURE 2.12: Heinichen's (right) and Kellner's (left) versions of the regional circle. Whilst Heinichen's diagram alternates between major and minor relatives, Kellner's representation is the first to order relative degrees in two concentric circles.

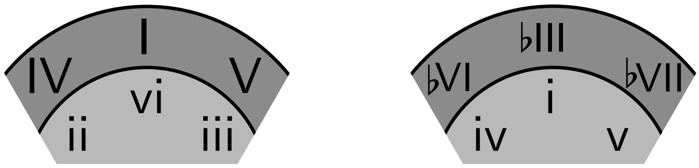


FIGURE 2.13: Sections of the regional circle, in Roman numerals and relative notation, expressing the principal neighbouring relationships for major (left) and minor (right) keys.

A detail of the regional circle is presented in Figure 2.13, abstracted into relative notation for major (left) and minor (right) keys. In both cases, the most common key relationships are represented in their respective figure. A major key is typically related to its dominant, subdominant and minor relative (vi) regions. Alternatively, minor keys mostly relate to their major relative (♭III) and subdominant, although excursions to other neighbour regions are not uncommon. As a general principle, the further two keys are in the circle of fifths, the further they are in terms of tonal similarity, and inter-key distance can be simply measured by counting the number of fifths between two tonics, as suggested above (Lerdahl, 2001).

Key relationships are essential to euroclassical tonal dynamics. In the course of a piece, music typically evolves through various tonal regions, establishing temporary deviations from the initial key. This process of digression from one key to another is called *modulation*, and it is arguably one of the most expressive devices in Western classical music. Most typically, music modulates to neighbour keys, especially to the dominant region (in major) and the relative major (in minor). Modulations can be abrupt and surprisingly unexpected, however, it is frequent to prepare the ear for the new key by means of a modulation process, consisting in a moment of temporary perceptual —or analytical— ambivalence. This ambivalence is typically achieved by the use of 'pivotal chords', that is, the common chord vocabulary of both departure and arrival keys, thus providing fairly soft transitions. Furthermore, modulation processes tend to culminate with a cadential process, reassuring the new tonal region.

Figure 2.14 illustrates a simple modulation from C major to D major, two steps apart in the circle of fifths. After establishing the key, the second half of the example initiates a modulation process through chords that are common to both keys (pivotal chords Emin and Gmaj), allowing a double interpretation of the fragment until the unequivocal new dominant chord (Amaj) appears in the second last bar, softly completing the modulation to D major.

Up to this point, we have attempted to cover the very basic materials of tonality, from an euroclassical music-theoretical perspective. We have seen how the division of the octave into various interval patters creates different scales, and how from these scales

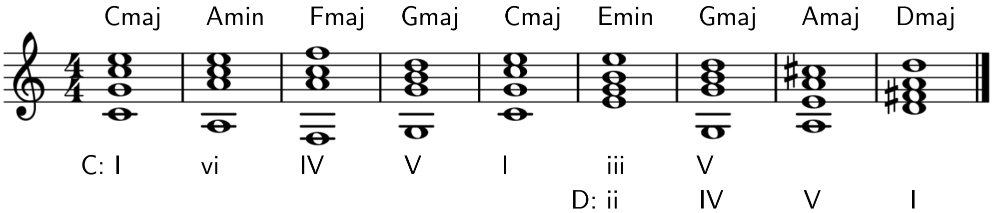


FIGURE 2.14: Simple modulation process using two pivotal chords (Gmaj, Emin) to proceed softly from C major to D major.

different chord vocabularies are obtained. We have also seen that chords are sequentially organised to provide a sense of key, and that multiple keys are combined in the course of a musical composition. In a proper sense, the notion of tonality, comprises each and all of these layers of musical information, operating simultaneously in order to convey a final sense of tonality, as suggested by the drawing in Figure 2.15. This is at least the opinion of theorists like Schenker or Schoenberg (1974), who proposed that a musical composition embodies one single tonality, structurally articulated by other secondary keys. Tonality, in this acceptation, appeals directly to the organisation of large musical structures, controlling any of its constitutive elements in an almost fractal metaphor.

2.3 Modal Practises in Popular Music

The description of euroclassical tonality provided in the previous block must be regarded as a practicable simplification of what is otherwise an extremely sophisticated tradition. Our intention was to establish a common ground, upon which other tonal practises —closer to and including our object of study— can be discussed and understood. Furthermore, this will help understanding some methodological decisions regarding the design and evaluation of key-finding systems, as discussed in Chapters 3 and 4, respectively. In the following paragraphs, we describe tonal and modal aspects in pop and rock music, to the extent that are deemed useful in developing subsequent chapters, since tonality in EDM is more akin to these genres. Throughout this dissertation, we refer indistinctively to these musics with the agglutinating term 'popular music', as it is common in scholar literature (e.g. Middleton, 1990; Moore, 2003; Tagg, 2014). The term 'popular music' is a polysemic and somewhat polemic acceptation, which has been defined either in relation to the music publishing industry (measured in terms of sale-rates and media presence) or to the essential class struggle, theorised both from a top-down elitist view ("inferior music", "music for the masses")

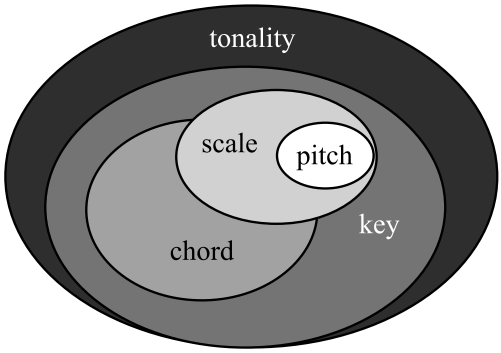


FIGURE 2.15: Hierarchical organisation of the constitutive elements of tonality. Pitches are the most basic elements of tonality, arranged melodically and harmonically, to suggest a particular musical key. In euroclassical tonality, the perception of a given key is mainly achieved with specific chord sequences and cadences. Furthermore, the overarching impression of a governing tonality is reached through narrative processes across various related key regions.

and a bottom-up leftist populist standpoint ("music of the people") (Middleton, 1990, pp. 3–7). Throughout this dissertation, we use the term to define a musical ground clearly differentiated from both the Western 'written' repertoire and traditional or 'folk' music, including pop and rock, but also other Western —and predominantly anglophone— manifestations, such as soul, funk, reggae or metal. Given the focus of this dissertation on electronic dance music, we have deliberately excluded it from this category. This is, in the first place, a simple methodological decision, which will allow us to differentiate EDM's tonal practises from those present in other popular music styles. However, this division is further grounded on two differential facts, already discussed in the introductory section to EDM (1.2.1). On the one hand, the essentially instrumental and accumulative nature of EDM contrasts with most popular music styles, which are predominantly sung and typically arranged in strophic structures. On the other hand, EDM seems to impose new production, consumption and distribution schemes, away from rock's stardom system and the record-sale markets which are typically associated with some of the definitions of 'popular music' (Tagg, 1994; Middleton, 1990).

2.3.1 The Extended Present

Popular music —in the restricted meaning that we have just suggested— is predominantly sung, and the impositions of cyclical verse structures have had a definitive role in shaping the essential formal layout of most popular musics. Regarding their tonal construction, Moore (1992) remarks that the formally strophic nature of popular song forced harmonic movement to be arranged as cyclical chord progressions, returning to the initial chord at the beginning of each verse, in contrast with the tonal linearity

aeolian	♭VI - ♭VII - i	(A♭maj - B♭maj - Cmin)
mixolydian	I - ♭VII - IV	(Cmaj - B♭maj - Fmaj)
ionian	I - V - vi - IV	(Cmaj - Gmaj - Amin - Fmaj)

FIGURE 2.16: Some common chord sequences in popular music, in Roman numeral notation and rendered in C. The degrees selected, denote quite specifically a diatonic mode (from Moore, 1992).

of euroclassical music (1992, p. 81). These cyclical sequences, typically *chord loops* comprising three or four chords (Figure 2.16) are one of the most salient structural elements of popular music, and have been studied and taxonomised in detail, according to modal and intervallic characteristics, by Moore (1992) and Tagg (2014, pp. 401–455).

A related consequence of the verse-chorus alternating structure, together with the repetitive nature of harmonic loops, separates further these musics from the euroclassical tradition, where modulation is the main organiser of the musical flow. Contrarily, modulation as a linear process is rare in popular music, and a great deal of the repertoire tends to remain in a single key for a whole song (as we statistically show in Chapter 4.1). However, modulation is not alien to popular music styles, but it is normally performed differently. Although it is still common to use pivotal chords, new tonal regions are normally not reassured via traditional cadential processes (Moore, 1995, p. 193). Besides, shifts to a different key without a previous preparation are frequent in transitioning between verses and choruses. These shifts are typically associated with aspects of emotional pitch intensification (for example, by progressing from a verse in minor into its relative major in the chorus), rather than to macro-structural tonal-functional organisation (Doll, 2011, par. 3). A similar operation has been described by Temperley (2011) as *scalar shift*, implying changes in the scalar pattern (mode) throughout a song, while maintaining a common tonal centre. In this sense, Moore has observed that an "important difference between modal and tonal is the assumption of span. There seems to be no a priori reason why we should assume that a mode operates throughout a song" (Moore, 2012, p. 71). A last example of the intensional —rather than structural— role of key changes in rock is epitomised in what Everett has denominated the "truck-driver's modulation" (Everett, 2004, p. 14), consisting in the successive modulation upwards by seconds, carrying no necessary implication of return (also Moore, 1995, p. 193).

For these reasons, in opposition to the extensional design of euroclassical music, with its supra-structural organisation around one single tonality, the musical experience of songs is identified with *intensional* aesthetics, intimately linked with what Tagg has denominated the *extended present*, "lasting roughly as long as it takes a human being

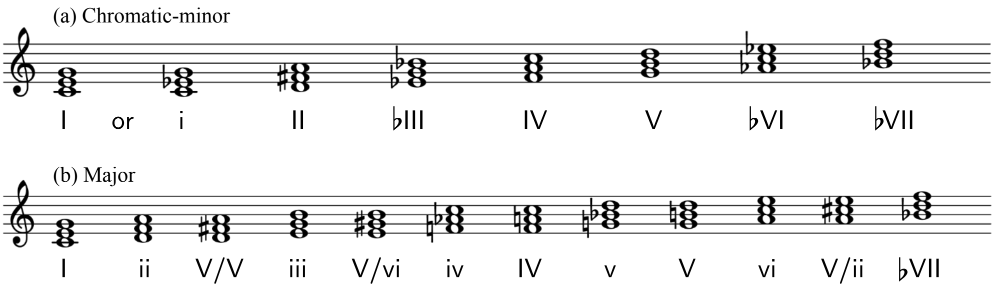


FIGURE 2.17: (a) 'Minor-chromatic' and (b) 'major' palettes by Stephenson (2002). While the first is a 'majorisation' of all the triads in an aeolian mode, his major mode is essentially a mixolydian with additional chords borrowed from the parallel minor (iv, v) and neighbour keys as secondary dominants.

to breathe in and out, or the duration of a long exhalation, or of a few heartbeats, or of enunciating a phrase or short sentence" (Tagg, 2012, p. 282), that is, corresponding to the short-memory span that a regular listener holds for interpretive purposes, aligned with harmonic loop repetitions, or the span of a verse or a chorus in a song.[33]

2.3.2 Rock Modality

Various sources provide different explanations regarding the formation of rock modality. For example, Everett (2004), considers up to six different tonal systems, taking into account principles of voice leading and harmonic structure. As a matter of fact, the first approach he describes is that of (a) common practise tonality, coexisting with other diverging approaches such as (b) diatonic modality, allegedly under the influence of traditional music styles. His third category represents a state of (c) relaxation of the principles of functional harmony and voice leading within the realms of euroclassical or diatonic modality, whilst the fourth system comprises of (d) musics evolving from the blues, radicalised in his fifth category, including (e) non-functional harmonisations of the pentatonic scale. Last, Everett's sixth system embraces other mainly chromatic practises, in which tonal centres progressively loose their syntactical function. Stephenson (2002, pp. 88–96), alternatively, proposes a threefold taxonomy, including aeolian harmony, a chromatic-minor system, which comprises of major triads over the degrees of the aeolian scale, and an 'extended' major mode, including major and minor triads over the degrees of the mixolydian scale, illustrated in Figure 2.17.

[33]The differentiation between 'intensional' and 'extensional' forms comes from Chester (1970). 'Extensional' denotes musical developments over larger periods of time (e.g. a Sonata form), whereas 'intensional' characterises musical developments akin to repetition. See also (Tagg, 2014, pp. 356–257).

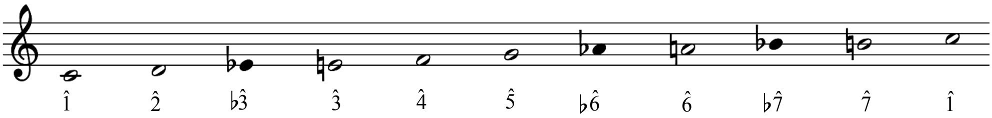


FIGURE 2.18: Rock's 'supermode' (Temperley, 2001, pp. 258–264), resulting from the combination of the ionian pattern (white noteheads) with the characteristic degrees of rock's flat-side common modes: mixolydian, dorian and aeolian (black noteheads). This results in an almost chromatic scale, only missing $\flat\hat{2}$ and $\sharp\hat{4}$.

From the various approaches to rock modality, Moore (1992, 1995, 2012) condenses in simple terms the harmonic role of diatonic modes (principally ionian, mixolydian, aeolian and dorian) in the conduction of the bass, whose notes are most likely in root position, considering the diatonic chord types as irrelevant for the expression of the mode (Moore, 2012, p. 73). This way, modal expression becomes an intermediate ground between melodic and harmonic thinking, with vague reminiscences of the parallel mixtures in the music of the French impressionists (e.g. Debussy), where the chord types assume more of a 'sounding' quality, rather than 'functional'. This view is supported by the study on heavy metal harmony by Lilja (2009), who attributes a major quality to all power chords (two-tone aggregates consisting of tonic and fifth, thus neither major nor minor), reinforced by highly distorted guitar sounds.

In summary, chords on any scale degree are often of major type, or, in the case of power chords, they tend to be perceived as such, given the reinforced harmonic series of the root and its fifth (Lilja, 2009, pp. 102–114). On the other hand, the root scale degrees tend to abandon ionian modality, favouring 'flat-side' modes, such as mixolydian, dorian and aeolian, as supported by the various theoretical observations. Everett's fifth system, for example, includes songs 'harmonising' the minor pentatonic scale, with power chords or major triads. Similarly —from a 'major-centric' perspective— the 'supermode' proposed by Temperley (2001, pp. 258–264) shown in Figure 2.18, represents an attempt to include the flattened degrees ($\flat\hat{3}$, $\flat\hat{6}$, $\flat\hat{7}$) characteristic of the dorian, mixolydian and aeolian modes into the ionian set, something that can be also interpreted as a merge of the parallel major and natural minor modes. One important thing these 'flattened' modes have in common is the absence of the leading $\natural\hat{7}$ degree, quintessential to tonal functional harmony, as we have seen in the previous section. From this follows that the V-I harmonic resolution does not play such a primary role in popular music harmony, superseded by subdominant relationships (Stephenson, 2002, p. 113). Consequently, the minor harmonic scale, which was used to provide the aeolian with a leading tone, is less common, insofar as the aeolian mode stands as popular music's minor modality *par excellence*. Figure 2.19 presents the six-chord sequence of Jimmy Hendrix's "Hey Joe", which illustrates some of the common traits of rock modality: (a) the root notes come from

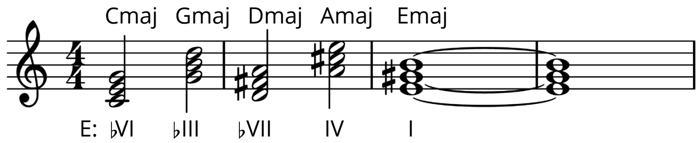


FIGURE 2.19: Chord cycle of Jimmy Hendrix's "Hey Joe".

the E aeolian mode (although they can be ambiguously seen as the C pentatonic scale {cdega}); (b) all the chords played are major (and originally distorted) and (c) the sequence presents a 'round of subdominants' proceeding clockwise in the circle of fifths towards Emaj.

All of the theories presented in this section suggest that the euroclassical distinction between major and minor modality in popular music is, to say the least, questionable, and this fact should be acknowledged in key-recognition systems, as we will address in the following chapter. However, authors coincide that the type of modality just discussed is just one of the 'possibilities' of popular music harmony, coexisting with more normative euroclassical harmony as well as with blues-based influences.

2.3.3 Blues, Pentatonicism and Dominant Seventh Chords

The influence of blues patterns in popular musical styles has been acknowledged by most authors writing on popular music harmony. However, its influence is particularly noticeable in the formation of rock'n'roll, blues-rock and early metal genres (Lilja, 2009, pp. 30–35). Blues is typically expressed over the *minor pentatonic* scale, although, as in other folk and world musics, it is often embellished with additional notes, falling out of the scale and of equal temperament. Figure 2.20 shows, in white noteheads, the scale of C minor pentatonic, presenting an intervallic pattern that avoids semitones (removing the differences between the three 'minor' modes, aeolian, dorian and phrygian, laying on the second and sixth degrees). Black noteheads, and especially $\sharp\hat{4}$, represent the so-called *blue notes*, idiomatic passing notes recurrent in melodic patterns and riffs. The other two blue notes ($\natural\hat{3}$ and $\natural\hat{7}$) are normally present in the harmonic structure, typically a idiosyncratic sequence involving major chords I, IV and V. As a matter of fact, blues harmony is more commonly conveyed with 'dominant-seventh' chord types, four-note sets originally appearing over the fifth degree of the ionian scale (or the first mixolydian degree).

Up to this point, we managed to organise our explanation without the need to discuss new chord types, since both euroclassical music and rock are essentially triadic. However, other musical genres such as jazz, use 'embellished' chords to define its characteristic harmonic sound, influenced by the language of French impressionists,

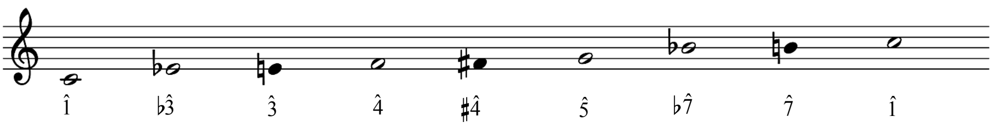


FIGURE 2.20: Minor pentatonic scale (white notehead) with additional *blue-notes*.

tin-pan alley and blues alike. These embellished chords are typically obtained by simply piling up more thirds together, according to the same constructive principles of triadic tertian harmony, obtaining chords of four, five and more notes that add their characteristic 'colours' to the triads.

Figure 2.21 illustrates some of these tertian constructions, typically named after the interval between the root and the highest note in the stack of thirds. On the left, the tetrads Cmaj7 and G7 are shown, as a result of piling up four consecutive thirds upon the roots of the C ionian collection (Figure 2.21a). Note that their intervallic pattern differs slightly: the first, indicates a major seventh interval from the root (maj7), whereas the single '7' implies a 'minor (dominant) seventh'. Similarly, Figure 2.21b shows Amin9 and Dmin9, two five-note chords built with the notes of the diatonic collection. These extended chords are used in other musical besides jazz. The min9 chord, for example is one of the most common pc-sets in house music, as we will see in Chapter 5. However, dominant sevenths are particularly abused throughout most tonal styles, from euroclassical music to blues, where it stands as the basic chord type in most traditional sequences, involving I7, IV7 and V7, although freed from their dominant tonal functions. These tentrads, instead, can be seen as the 'consonant' continuation of power chords and major triads, aligning with the acoustic properties of the root's harmonic series (Lilja, 2009, p. 135). Figure 2.21c shows yet another common type in blues-rock, a C7♯9, representing the fertile coexistence of melodic minor pentatonicism ($\sharp\hat{9} \approx \flat\hat{3}$) over major/dominant harmonies.

2.3.4 Tonal Ambiguity

Modal harmony is prone to naturally introduce musical ambiguity, promoted by the rotational nature of the modal system (remember that C ionian, D dorian, G mixolydian and A aeolian share the same pitch-class set). This peculiarity, together with the absence of dominant-tonic cadences in the normative sense, calls for other strategies in establishing a potentially ambiguous tonal centre. The main perceptual marks for the disambiguation of the tonic are typically provided by the alignment of tonic chords with hypermetrically strong positions[34] together with other aspects such as *persis-*

[34] A hypermeter can be thought of as a 'measure of measures', normally grouping blocks of four bars, as a continuation of the metrical hierarchy of the 4/4 time signature (Stephenson, 2002, pp. 56–60).

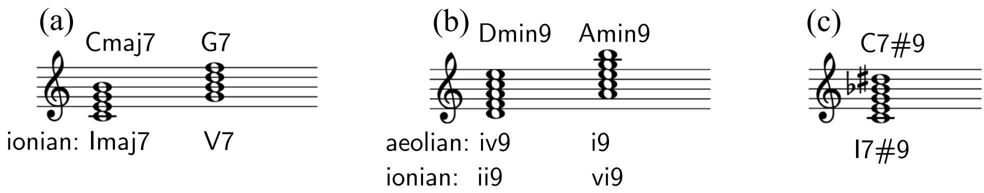


FIGURE 2.21: Typical chords containing more than three notes, namely 'sevenths' and 'ninths.'

tence (length) and laterality (initial/final chords) (Moore, 2012, p. 75). Figure 2.22 presents three different chord progressions, obtained by rotating the same chord sequence. In these examples, the disambiguating factor between the perceived tonics are mainly attributed to the metrical arrangement of the chords in each progression. Tagg (2014, pp. 421–450) has referred to this type of ambivalence as "bimodal reversibility", an operation by which the same sequence can be heard in two modes simultaneously, especially between ionian-mixolydian and aeolian-phrygian. Other naturally bimodal sequences are often found in so-called *harmonic shuttles*, consisting in ongoing oscillations between two chords of similar duration, in which cases, the preference of a tonic chord over the other might become a totally irrelevant issue. In addition to this, some traditional styles present a particular musical interaction that can not be univocally perceived from a single tonal centre, but as a shared tonicality between two chords in the same progression. These *bimodal sequences*, a term coined by Vega (1944), according to Tagg (2014, p. 436), convey a sense of 'horizontal' tonical ambiguity as the validation of two tonal centres in relation to the same scalar material, rather than as the simultaneous operation of different tonics that the notion of polytonality implies. This type of bimodality is particularly frequent between relative keys (I/vi or i/♭III), and is common in folk musics from Ecuador, Cuba or Argentina (Béhague & Schechter, 2012; Tagg, 2014).

Yet another source of tonal ambiguity comes from the so-called 'harmonic-melodic divorce', by which the harmonic sequence and the melodic arrangement do not neces-

(a)	*ionian:*	|: I	|	| IV	| V	:|	(Cmaj - Fmaj - Gmaj)
(b)	*mixolydian:*	|: I	|	| IV	| ♭VII	:|	(Gmaj - Cmaj - Fmaj)
(c)	*'major'*	|: I	| II	| V	| I	:|	(Fmaj - Gmaj - Cmaj - Fmaj)

FIGURE 2.22: Potentially ambiguous chord loops in ionian, mixolydian and Stephenson's 'major' palette. The sequences are obtained by rotating the same chord progression. The original sequences have been transposed to present exactly the same chords. The chord sequences belong to (a) The Beatles' "Here Comes The Sun", (b) The Kinks' "Lola" and (c) "Mr. Spaceman" by The Byrds. The sequences are taken from Moore (1992).

sarily express the same modality, as a relaxation of the melodic expression with regard to the harmonic structure (Stephenson, 2002; Moore, 1995; Temperley, 2007b). As we have seen, this could be effect of power-chord metal structures, major harmonisations over aeolian patterns or the minor/dominant interaction in blues-derived styles. This is the case, for example, in Led Zeppelin's "Whole Lotta Love", collected in the corpus by Temperley & De Clercq (2013), which will be discussed in Chapter 4. The tonal centre of the song is clearly E. However, authors annotate the tonic chord differently as Emaj and Emin. In our humble opinion, the actual tonic chord is a thirdless power chord. And this is exactly the point. As listeners, we could be more inclined towards the 'major quality' implied by the harmonic series of E5, according to Lilja's thesis (the instrumentation is, after all, a distorted electric guitar). Or perhaps, we could perceive more prominently the minor pentatonicism suggested by the riff (or is it mixolydian?). We could also concentrate on the vocal melody, where g♯ ($\hat{3}$) appears often, although sometimes considerably lowered as to be perceived closer to g♮ ($\flat\hat{3}$); or after all, we would be better accepting the multi-faceted and ambiguous nature of rock's modality, with influences as diverse as euroclassical tonality, folk-song modality and blues pentatonicism, in an otherwise extremely unique manifestation of tonal organisation.

2.4 Pitch and Tonality in EDM

As stated in Section 2.3, throughout this thesis we treat EDM as a musical genre differentiated from other popular music styles, mainly for methodological reasons. However, yet another indicator of its 'different' nature could be given by the visible isolation of the topic in popular music theory. As Doehring observes, musicological analysis has contributed far less to the study of EDM than other disciplines, arguing that EDM falls out of the reach of musicological enquiry, because it "cannot claim to be accepted by the dominant definition of music" (2015, p. 134). This fact is only worsened by the lack of scores, common instruments, symbolic compositional operations and widespread record circulation.[35]

In any case, the most comprehensive musicological studies address aspects of rhythm and meter (Butler, 2006; Danielsen, 2010), since they are doubtlessly the most prominent elements in EDM. Formal aspects have been covered by Spicer (2004), Garcia (2005) and Solberg (2014); and a few publications have attempted the analysis of complete tracks, raising specific methodological questions (Ratcliffe, 2013; Doehring, 2015), and presenting timid considerations of melodic and harmonic

[35] As Doehring himself points out, most EDM music is released in vinyl and published in short batches of around 200 copies, to be normally distributed among DJs, and therefore hardly accessible for the regular audience or the scholar.

features, in relation to other sonic, technological and procedural aspects (Ratcliffe, 2013, sec. 6). Therefore, most references to tonal habits in EDM are inserted incidentally in works addressing other musical aspects. For example, in analysing Andrés's "New For U", a successful 2012 house track, Doehring observes that

> "most of the chords have alterations we know from a lot of styles of popular music. The main theme [...] is a pentatonic scale on A minor that starts over a Dmin, which thus becomes a Dmin9 ." (Doehring, 2015, p. 144)

Ratcliffe writes in a similarly descriptive prose about "Chimes" by Orbital,

> "This material appears to have been constructed using a technique common to Detroit techno and early forms of EDM, whereby a sampled chord is assigned to the notes of a keyboard and then played/sequenced as melodic material." (Ratcliffe, 2013, s. 6)

These two fragments, certainly suggest a deliberate compositional working of harmonic aspects, either associated with previous musical styles in the case of Doehring, or to particular techniques characteristic of certain EDM genres in the text by Ratcliffe. In a similar vein, in one of the first publications drawing attention towards electronic dance music, Tagg (1994) devotes a paragraph to describing 1990's rave music in terms of its tonal idiosyncrasies. Tagg does not ascribe a prominent role to the bass layer, describing bass riffs as simple, made of repeating "notes under the overlying chord (usually a triad) or cycling stepwise round it." However, he concedes some relevance to the harmonic-filler, which normally consists of power chords or triads without extensions, played as syncopated stabs, with piano-like envelopes. Although this characterisation is not generalisable to other types of EDM, Tagg observes that,

> "The tonal language of rave music also shows some interesting traits. Whereas 'R&B dance' uses a lot of disco's major and minor seventh sonorities and whereas 'dance rap' sticks to the basically percussive backing tracks of rap music in general, European and North-American techno-rave seems to go in a big way for the aeolian and phrygian modes, not as harmonic padding for blues pentatonicism, but as straight sets of minor mode triads or bare fifths without much trace of a seventh, let alone ninth, eleven or thirteenth. [...] No internationally popular music of this century has shown such a leaning towards these modes." (Tagg, 1994, p. 215)

With his description, Tagg locates techno's tonal language in a unique position, not only with regard to the house music described by Doehring, but also with regard to all other Western major popular styles, expressing a perceptible fascination for the abuse of phrygian modality in techno-rave music. A similar observation is generalised in the following comment by Spicer:

> "An emphasis on dissonant tritone and semitonal relationships seems to be a characteristic of the harmonic language of many techno tracks: for example, also on *Music for the Jilted Generation*, Prodigy build the main groove of their "Full Throttle" around another oscillating two-chord vamp, I-♭II, featuring phrygian mixture. (Spicer, 2004, p. 54)

Whether this is characteristic of techno, truth is that we do not know about the sources that Spicer considered to make his claim (that techno emphasises tritone and semitonal relationships). On another track by Prodigy, Spicer continues,

> "While "Break and Enter" is most definitely in G♯, this tonality is by no means projected in a conventional manner. [...] The first two of the pitched riffs illustrate the oscillating two-chord vamp that governs most of the main body of the track: a G♯min7 chord moving to a tritone-related Dmaj triad [...] suggesting instead a kind of locrian mixture wherein the dominant chord is build on the lowered fifth scale degree." (Spicer, 2004, p. 54)

Furthermore, in the following page, Spicer acknowledges a sense of "conflicting modality, for example, G♯ aeolian against G♯ locrian" (2004, p. 55), which seems to call for the term polyscalarity (Tymoczko, 2002), introduced in Section 2.2.1.

These various comments make reference to harmonic language, popular music, various chord types, melodic-harmonic interaction, particular scales and even, polytonality. After all, it seems natural that the heterogeneity of EDM incorporates a wide variety of approaches, from the most conservative loans from other musical styles, such as jazz, soul and even euroclassical music —EDM is appropriative by definition— to more adventurous and 'unique' configurations. However, the excerpted quotations do not clarify much about tonality in EDM. First, most of them are expressed incidentally in works covering other aspects. Furthermore, with the exception of Tagg and Spicer's generalisations, all other observations correspond to analyses of individual composition, and no author has claimed any prescriptive meaning for their descriptions.

2.4.1 A New Tonal Framework

Wooller & Brown (2008) have already signalled that musicological analysis might have overlooked the significance of tonality in EDM. They acknowledge that limitations of the analytical power of traditional methods might have concealed potentially novel practises, claiming that tonality is an important creative parameter in EDM, when conceived and observed in more open-ended terms.

For example, they detect that EDM tracks sometimes convey an apparent lack of tonic, and that tonal ambiguity and the coexistence of non-tonal voices with tonal layers are common territories. However, in most scenarios, they recognise a clear tonic and modality, typically pentatonic, minor or mixolydian (2008, p. 93).

On the other hand, their most interesting contribution, is a conceptual framework to think of tonality in a more open ended way, characterising tonal practises in EDM by describing the horizontal and vertical interactions of pitched materials, for which they define four different tonal attributes, providing an ample list of examples within their paper.

1. The *rate of tonal change* relates to the amount of activity across the tonal layers. At one extreme of this attribute the music consists on a one-tone drone or a reduced pitch-class set that does not change over time, whereas the other end represents mostly atonical music (i.e. music in which it is difficult to find a sense of tonal centre.

2. The concept of *tonal stability* is related to the notions of *tonal implication* and *tonal ambiguity* proposed by Temperley (2007c) referring to the strength of the tonic feel (what we have defined elsewhere as tonicality) as much as to the recognisability of specific modes or scales.

3. The *pitch-to-noise ratio* attribute intends to establish a bridge between the timbral predominance of most EDM with the potential tonal information carried within. With this descriptor, Wooller & Brown are able to differentiate between tracks that consist only of noisy, untuned and/or distorted sound materials at one end, and tracks with pure sinusoidal tones on the other.

4. Last, the *number of independent pitch streams* refers to the counterpointal density of the music, as identifiable voices operating independently and simultaneously.

Authors like Ratcliffe (2013) seem to adhere to this analytical framework to observe tonal interaction within EDM tracks. However, its open-endedness, although an extremely useful departure point, does not target specific traces, as they could be used, for example, in an automatic key estimation system. In any case, this proposed framework seems to break with the reductionistic explanations found in the literature, as much as with other preconceptions inherited from either euroclassical tonality or rock modality. However, this apparent openness is somehow contradicted by the proliferation of software offering automatic key analysis for DJs (some of which are described and evaluated in Chapter 4.3), invariably based in the euroclassical binary taxonomy, especially supported by a well-known mixing technique, as we discuss in the next block.

2.4.2 Harmonic Mixing

DJ mixing can be seen both in terms of *simultaneity*, (by layering diverse sound sources together to create a new whole) and *progression*, that is, the sequential arrangement of different musical moments in order to create an engaging experience throughout the DJ set. The notion of *harmonic mixing* originates in the second one, as a conceptual extension of the practice of beat-matching between consecutive records in a set, in order to guarantee soft transitions between tracks. Analogously, choosing tracks with tonally related keys seems to smooth the transition between them, supporting the ideal of an uninterrupted DJ set.

Companies like *Mixed-In-Key*[36] —the industry standard in key detection software— have developed a didactic —and marketable— narrative around the craft of harmonic mixing, proposing the so-called *Camelot Wheel* shown in Figure 2.23, a colourful reworking of Kellner's regional chart from 1737 (compare it with Figure 2.12), labelled 'by the hours' rather than by chroma names, since there are twelve potential tonal centres, just as there are 12 hours.

As we have seen in Section 2.2.3, the circle of fifths is a music theoretical construct that arranges musical keys in intervals of fifths. In the equal-tempered system, this circle assumes enharmonic equivalence (f♯ ≡ g♭), so that adding twelve consecutive fifths leads back to the initial pitch, completing the chromatic circle. The importance of this arrangement is that keys one fifth apart share most of their diatonic scale, as we have seen, with six shared tones between neighbouring regions, what has been proven to be correlated with perceptual similarity (Krumhansl, 1990; Lerdahl, 2001) and had a great influence in the development of modulation as a discursive technique in euroclassical music. In the Camelot Wheel, major keys (labelled with a B) are represented in the outer circle, whereas the minor relative keys (labelled with an A) are depicted along the same radius, in the inner circle. 'Time differences' between keys represent their relatedness.

The marketable didactics of the Camelot Wheel spread in websites[37,38] and publications (Vorobyev & Coomes, 2012) where tips and tricks are published to underline the importance of a well, harmonically aware, mix. For example, they advise beginners to start of by mixing tracks with the same number (8A and 8B, parallel keys) or moving

[36] http://www.mixedinkey.com
[37] http://harmonic-mixing.com
[38] http://camelotsound.com

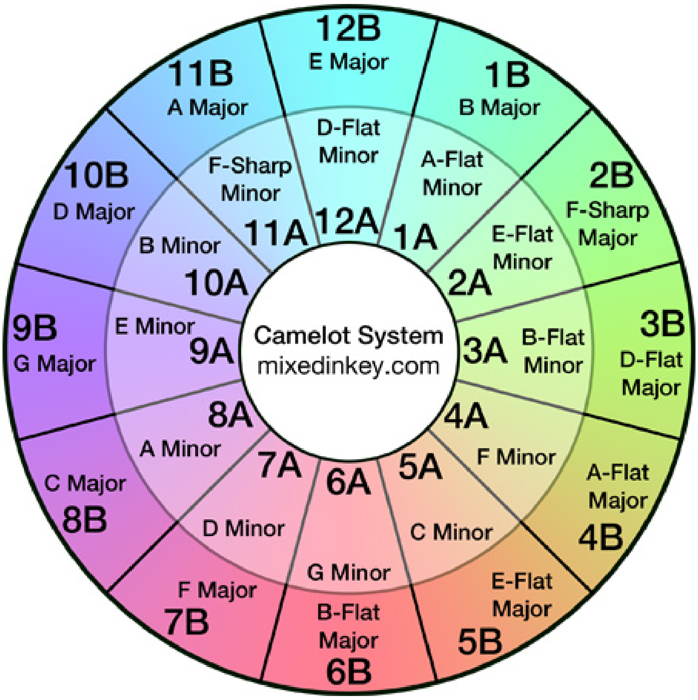


FIGURE 2.23: The so-called *Camelot Wheel*, a double circle of fifths arranged with hour labels, designed to simplify harmonic mixing (compare with Figure 2.12).

one 'hour' in either direction in the circle (neighbouring keys).[39] For an energy boost, they advise to turn 'two hours clockwise' (ascending tone change)[40] or mix the keys of the tonic and mediant (I, iii) "diagonally on the wheel from 8B to 9A" or vice versa (Vorobyev & Coomes, 2012).

Recently, harmonic mixing has attracted attention in the MIR field, and the last couple of years have seen publications addressing harmonic mixing from psychoacoustic perspectives. For example, Gebhardt et al. (2015, 2016) measure the perceptual roughness to determine the compatibility between two different tracks, whereas Bernardes et al. (2017a) proposed a metric that uses inter-key distance and sensory consonance, integrated in a visualisation tool to guide users in their mixes. However, in the light of Wooller & Brown (2008) considerations, the binary division into major an minor modalities, promoted by harmonic mixing technologies, seems, to say the least, inappropriate for some EDM subgenres. Furthermore, if the assumption that most EDM is essentially minor, as suggested by statistical studies of popular music (Schellenberg & Von Scheve, 2012), but especially, by the figures that will be provided in Chapter 4, a finer differentiation into phrygian and aeolian modes could become

[39]http://harmonic-mixing.com/HowTo.aspx

[40]http://harmonic-mixing.com/EnergyBoostMixing.aspx

useful for tonal characterisation in EDM. On the other hand, monotonic or difficult tracks might as well be better characterised just by considering the tonic note, as proposed by Temperley & De Clercq (2013) for rock music, potentially identifying practises with varying rates of tonal change.

***

In this chapter, we have discussed the basic workings of tonality, from its consolidation during the Common Practise Era to more recent tonal developments in Western popular music. We have shown that there has not been much research regarding tonality in EDM, probably (un)motivated by the general belief that pitch is either structurally unimportant or a mere appropriation from other musical styles, including pop, jazz and euroclassical music. However, the analytical framework proposed by Wooller & Brown (2008) suggests that there might be genuine tonal practises in EDM, but they require to be examined under a different light. In this sense, we would like to recall what it was said in Section 2.2.1: If we consider tonality in any of the restricted meanings discussed in this chapter, we will possibly be blinded before any attempt of discovery. On the contrary, if we regard the term 'tonal' as a laboratory in which all 'music made with tones' can be conceptualised and understood, we are likely to be surprised with some tonal configurations found in EDM. But before we reach that stage in Chapter 5, the next two chapters introduce relevant background in MIR, addressing the topic of computational key estimation and other related methodological aspects.

Chapter 3

Tonality and Computers

"Like our bodies and like our desires,
the machines we have devised are possessed
of a heart which is slowly reduced to embers."

W. G. Sebald, *The Rings of Saturn* (1995)

The main goal of this chapter is to delineate the scientific terrain over which we have grounded our computational approaches to studying tonality in EDM. As explained in Chapter 2, tonality has been a principal actor in most Western musical practices, including euroclassical, jazz and most types of popular music.

Quite naturally, this predominance has been mirrored in the interest of the scientific community, addressing the study of tonal aspects from a variety of disciplines, including cognitive psychology, artificial intelligence and information science. In particular, the challenge of computational key estimation from audio, has motivated abundant research in the MIR domain, which is exceptionally condensed in two doctoral dissertations specifically addressing the topic (Gómez, 2006a; Noland, 2009).

Gómez (2006a) starts her discussion with an extended review of literature related to tonal induction and symbolic key finding, before describing various approaches in the audio domain, broadly grouped into transcription-based and pitch-distributional, and respectively dissected in a bottom-up fashion. Additionally, Gómez presents an interesting report on the various adaptations of theoretical profiles to operating in the audio domain. On the other hand, Noland (2009) organises her report around a taxonomy of tonality models used in various key estimation algorithms, dividing her narrative into (a) psychoacoustic models, (b) tonal hierarchies, (c) pitch spaces, (d) preference-rule systems and (e) machine learning approaches. Furthermore, Noland presents a comparison of low-level signal processing methods, and analyses the benefits and shortcomings of the *chromagram* as as pitch-class summarisation representation. Consequently, we have tried to organise this chapter in ways that could complement the detailed reviews in the mentioned works.

In the following section, we start by succinctly discussing the role of tonal hierarchies in tonality induction —the cognitive processes involved in key determination— upon which an important number of key-finding algorithms has been grounded. We continue discussing some early methods of key finding in Section 3.2, developed in close connection with the cognitive hypotheses presented, and operating on symbolic representations of music, We dedicate the last section of this chapter (3.3) to review relevant key estimation algorithms in the audio domain.

However, we recall that the main aim of this research is the adaptation of existing models of tonality estimation based on music-theoretical inspection of EDM, which seems to present unique tonal characteristics unseen in other musical styles. Accordingly, the descriptions contained herein are intentionally directed towards aspects that will become useful in the achievement of our goals.

3.1 Tonal Hierarchies and Pitch Distributions

> "Results of psychological studies indicate that Western listeners, even those without formal instruction, have extensive knowledge of typical tonal and harmonic patterns. However, contrary to traditional assumptions, at least some aspects of this knowledge are acquired without extensive experience and training." (Krumhansl, 2004, p. 266.)

The process of key determination seems to be a faculty that most listeners are reasonably capable of, as the quote by Krumhansl suggests. It is commonly accepted that humans induce tonal aspects from a musical stimuli based on previously acquired *tonal hierarchies*. Although the details regarding the hierarchical organisation differ, evidence of tonal hierarchisation has been found across different cultures and different musical habits. Furthermore, according to Krumhansl & Cuddy (2010), tonal hierarchies (a) are musical 'facts' that characterise different musical styles, (b) represent statistically significant patterns of the music they relate to, and (b) have a psychological reality.

Krumhansl's series of 'probe tone' experiments stand amongst the first experimental evidences of the existence of tonal hierarchies. The essence of the probe tone experimental method is to present a subject with a musical context (normally a scale, a melody or a chord sequence), asking her to rate the suitability of a proposed continuation to the initial context. Using the probe tone method, Krumhansl and collaborators

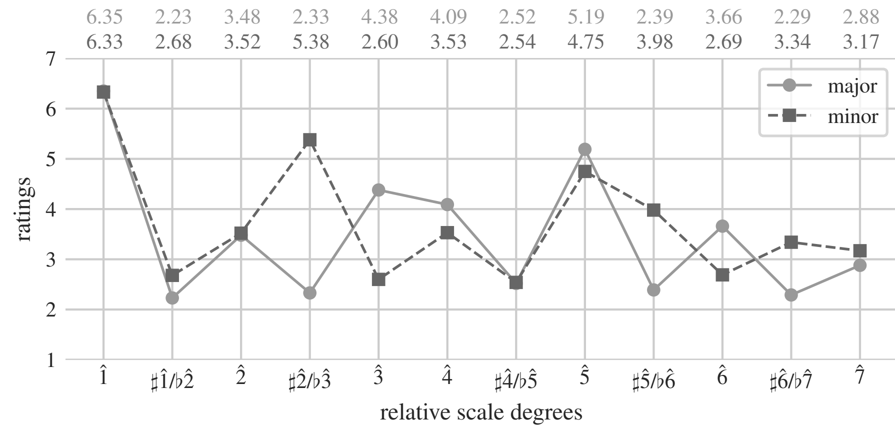


FIGURE 3.1: Major and minor *probe tone* profiles by Krumhansl & Kessler (1982).

have studied aspects such as tonal completion (Krumhansl & Shepard, 1979) and tonal context (Krumhansl & Kessler, 1982), over which further study relating tonal hierarchies with inter-key distance or tonal consonance has been grounded. A summary of this body of research can be found in Krumhansl (1990).

Figure 3.1 shows the contextual probe tone ratings from Krumhansl & Kessler (1982), in which subjects were asked to rate, in a scale from 1 to 7, the appropriateness of a proposed tone within a predefined tonal context. Although they conducted their experiments in the tonal region of C, they claim that experimental results revealed that these findings can be extrapolated to any other tonal centre, and accordingly, figures throughout this manuscript tend to represent relative scale degrees rather than specific chroma's.

Krumhansl & Kessler's profiles present intriguing correlations both with music theoretic observations and statistical descriptions of several musical corpora (Krumhansl, 1990, pp. 66–75). In these *key profiles*, the tonic note stands as the most important degree, after which the tonic triads of the respective modes appear. This *tonal hierarchy* is followed by the respective scale degrees, whereas the chromatic steps are situated at the bottom. This hierarchical division of musical pitch in relation to a tonal centre is of great importance in music cognition, and draws from Meyer's observation that humans develop their appreciation of musical style as statistical processes through listening to music (1957). Similarly, Krumhansl argues that "tonal hierarchies might be acquired through experience with the musical style, particularly through internalizing the relative frequencies and durations with which tones are sounded" (1990, p. 77). Therefore, according to Krumhansl's theory, we determine the key of a musical segment through a process of pattern matching between previously acquired tonal hierarchies and the particular pitch distribution of a given musical piece.

tonic:	$\hat{1}$											
diad:	$\hat{1}$							$\hat{5}$				
triad:	$\hat{1}$				$\hat{3}$			$\hat{5}$				
modal:	$\hat{1}$		$\hat{2}$		$\hat{3}$	$\hat{4}$		$\hat{5}$		$\hat{6}$		$\hat{7}$
chromatic:	$\hat{1}$	$\sharp\hat{1}/\flat\hat{2}$	$\hat{2}$	$\sharp\hat{2}/\flat\hat{3}$	$\hat{3}$	$\hat{4}$	$\sharp\hat{4}/\flat\hat{5}$	$\hat{5}$	$\sharp\hat{5}/\flat\hat{6}$	$\hat{6}$	$\sharp\hat{6}/\flat\hat{7}$	$\hat{7}$

FIGURE 3.2: Lerdahl's major 'basic tonal space' establishes a five-level hierarchy of pitch-classes given a tonal centre. (1988, p. 321)

From a music-theoretical perspective, Lerdahl (1988) studied the relationship between single pitches, chords, and keys with an algebraic model —in contrast with most 'geometrical' tonal spaces— of which the simplest representation is the 'basic tonal space' shown in Figure 3.2. Lerdahl's basic space establishes a five-level hierarchy for all pitch-classes given a tonal centre. The five levels correspond with progressively fainter indicators of tonal context, as represented by the (a) tonic note, (b) tonic diad (power chord), and (c) tonic triad, the (d) diatonic set, and finally, the (e) chromatic collection. Lerdahl himself points at the remarkable correspondences of his theoretical model and to Krumhansl & Kessler's experimental profiles (1988, p. 338).

In a similar vein, the 'Spiral Array' proposed by Chew (2000) explicitly attempts to incorporate the multi-levelled structure of tonality, representing pitches, chords and key relationships in a unified geometrical space. In simple terms, Chew's Spiral Array is an extension of the *tonnetz*, a planar representation of key and chord relationships dating back to Euler's times, and re-signified by Hugo Riemann (1903) to express chord and key relationships in tonal functional harmony. In Chew's model, the circle of fifths proceeds linearly along an ascending helix, while major thirds appear at the vertical alignment of pitch nodes. These basic intervallic relationships allow Chew to identify chord types with various planar configurations between pitch nodes, and, similarly, to associate specific keys to the distance-minimising point between its main chordal surfaces (i.e. tonic, dominant and subdominant), as represented in Figure 3.3. Besides its theoretical interest, Chew's spiral model has been used in the context of key-finding algorithms, both in symbolic (Chew, 2000, pp. 99–106) and audio domains (Chuan & Chew, 2005b).

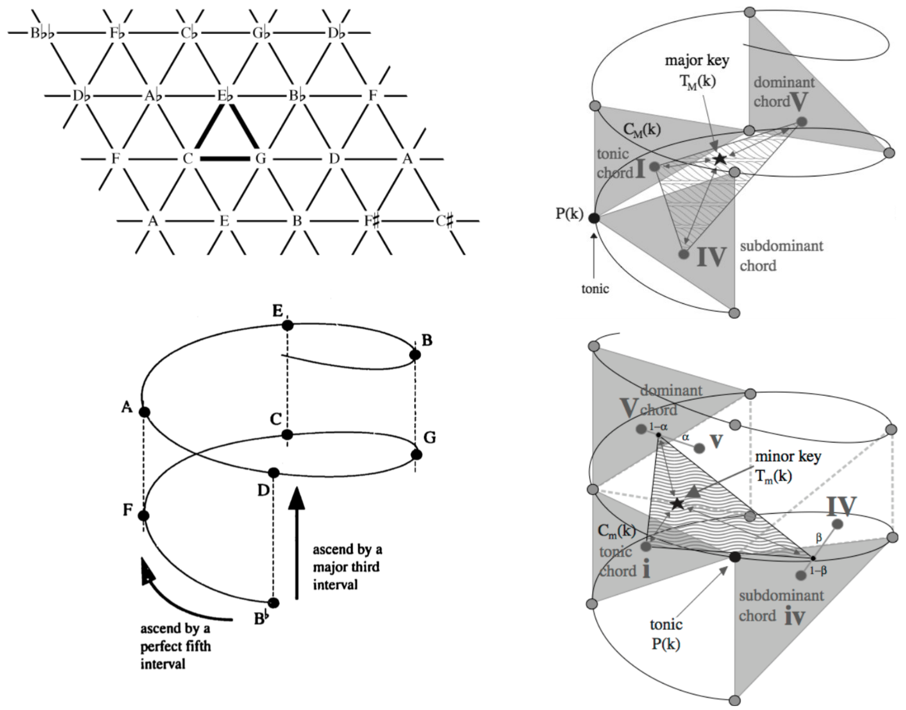


FIGURE 3.3: Simple *tonnetz* plane (top-left) representing major-third and perfect-fifth intervals along its axes. All other representations correspond to structures in the Spiral Array: translation of the tonnetz space (bottom-left), and pitch-chord-key distance relationships for major (top-right) and minor (bottom-right) modalities (Chew, 2000).

3.2 Symbolic Approaches to Key Identification

In this section we discuss key estimation methods which operate over symbolic representations of music, typically MIDI files, providing encoded sequences of pitch heights and durations, as shown in Figure 3.4. However, pitch-event representations do not necessarily imply the recognition of higher-level musical objects, such as chords and their progression, not to mention a sense of tonicality suggested to a particular listener.[41] The psychological reality of these higher-level structures is one of the aspects that symbolic approaches to key determination seek to illuminate. As a matter of fact, the essence of tonal analysis —from music cognition to musicological enquiry— resides in unveiling the relationships between decontextualised collections of objects (pitches, aggregates, sequences) providing them with a meaningful explanation at some level. Formalising these enquiries into computer programs has

[41] Although musical scores do write a key signature at the beginning of the staff, this has little effect on the actual perception of tonality, simply serving as a ‘deciphering’ code (i.e. ‘key’ in its original acceptation) of the note symbols.

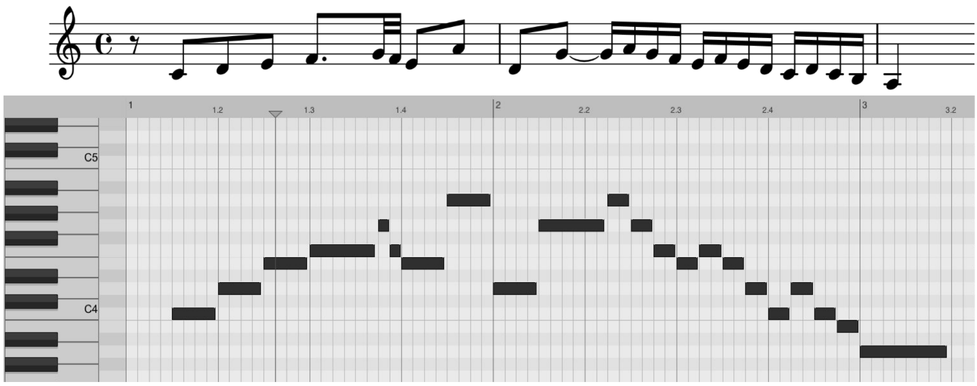


FIGURE 3.4: Symbolic representations of the first fugue subject from J.S. Bach's *The Well-Tempered Clavier*, in Western musical notation (above) and in a 'piano-roll' view (below), showing the height and duration of MIDI events.

proven fruitful in elucidating problems across several domains, from musicological questions about style (e.g. Cope, 1991) to cognitive research about our perception of music (e.g. Krumhansl, 1990; Temperley, 2001).

3.2.1 Rule-Based Early Methods

A number of early rule-based approaches have been proposed, originating in multidisciplinary studies in the fields of music psychology and artificial intelligence. For example, Longuet-Higgins & Steedman (1971, cited in Temperley, 2007a, p. 51) proposed a method to estimate the key on monophonic melodies, in the contexts of ionian and minor-harmonic scalar patterns. Their algorithm operates sequentially and on an event basis, discarding the keys not accounting for the totality of the pitches in the melody at each new step. This system appeals to a second rule in case the algorithm runs out of possibilities (e.g. if a modulation introduces chromatic tones) or there is more than one choice left (i.e. the melody consists of less than seven pitch classes), by looking at the first note to infer the key from it. Holtzman provided a similar method, by observing basic tonal marks, such as the presence of the elements of the tonic triad; and Chafe focused on melodic and rhythmic accents to detect tonal cues at important metrical positions (1977, cited in Krumhansl, 1990, p. 77).

3.2.2 Pattern-Matching Algorithms

As an attempt to assess tonality induction, Krumhansl and Schmuckler (Krumhansl, 1990, pp. 77–81) modelled a key-finding algorithm mimicking the pattern matching process between learnt tonal hierarchies and a musical stimuli. In their algorithm, they measure the total duration of each pitch class in an observed segment, creating a twelve-dimensional vector comparable to the probe tone ratings by Krumhansl & Kessler. The obtained distribution (sometimes referred to as 'input vector') is then compared pairwise with the major and minor experimental profiles, rotated to the 12 possible pitch classes. This comparison is computed as the Pearson correlation coefficient r,

$$r = \frac{\sum_{i=0}^{n-1}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=0}^{n-1}(x_y - \overline{x})^2 \sum_{i=0}^{n-1}(y_i - \overline{y})^2}} \tag{3.1}$$

where n expresses the length of the matching vectors (representing the 12 pitch classes), and $\overline{x}$ and $\overline{y}$ the sample mean of each profile,

$$\overline{x} = \frac{1}{n}\sum_{i=0}^{n-1} x_i \tag{3.2}$$

providing a single value in the range between -1 and 1, where a result of 1 identifies the two profiles as identical, and -1 implies that profiles are exactly opposite. The highest correlation value from the 24 measures obtained (2 key-profiles $\times$ 12 rotations) is taken as the key of the fragment.

This 'template matching' approach, is still regarded as one of the most successful methods for key identification, and has been implemented in both symbolic- and audio-processing scenarios with a number of variations. Temperley (1999) proposed a few improvements over the Krumhansl-Schmuckler method, including adjustments in the profile ratings, the correlation method, and the input-vector calculation, enabling the assessment of tonal evolution over time —thus potentially detecting modulations. Temperley proposes a few corrections in the original ratings, "arrived at by a mixture of theoretical reasoning and trial and error" (1999, p. 74). In Temperley's opinion, the new profiles provide a more faithful account of pitch distributions in euroclassical music by treating chromatic and non-modal degrees equally in both modes, by shifting the prominence of the aeolian $\flat\hat{7}$ towards the leading-tone from the minor harmonic scale ($\natural\hat{7}$), according to euroclassical normative practice (Figure 3.5). However, Krumhansl describes the subjects of her experiment as "university students of diverse musical backgrounds" (1990, p. 21), what might explain the coexistence of both lowered and natural sevenths in her experimental profiles, likely denoting the subjects' exposure to all sorts of popular music styles, besides euroclassical music.

Furthermore, during his revision of the rating profiles, Temperley (1999) proposed other modifications. For example, he observes that the Pearson correlation formula can be substituted by the dot product $\sum xy$, slightly simplifying the process. He argues that normalising both vectors for their mean and variance has no effect in the

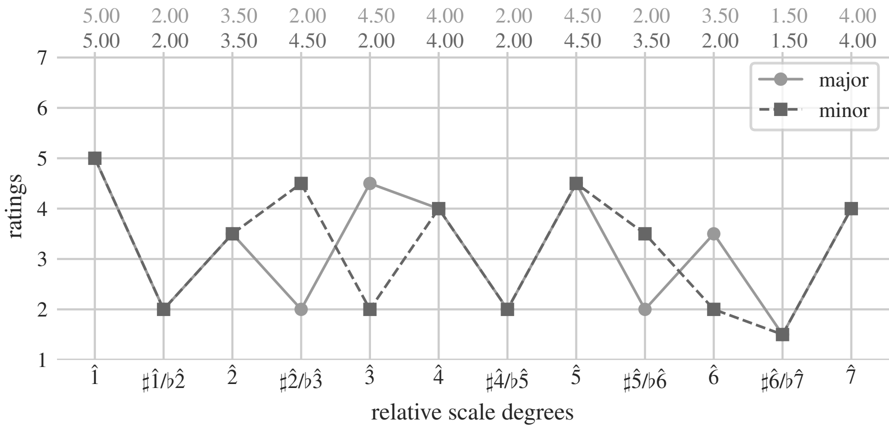


FIGURE 3.5: Modification of Krumhansl & Kessler's key profiles by Temperley (1999). Note that the values for chromatic and non-modal degrees are treated equally in major and minor profiles, and that ♭$\hat{7}$ has the lowest weight of all twelve degrees, below ♭$\hat{2}$ and ♯$\hat{4}$.

result, since the input-vector is the same for al 24 keys, although he acknowledges that normalising both key profiles might avoid biases towards one or other modality. However, of more relevance is the proposed alternative to the durational weights imposed by Krumhansl and Schmuckler. Instead, Temperley divides the musical stream into shorter segments, creating an activation 'flat' profile with the pitch classes present on each segment —acknowledging 'inspiration' from the previous model by Longuet-Higgins & Steedman. An argument supporting this "flat-input/weighted-profile" given the short analysis segments, is that the fewer the pitches, the more likely they will fall on stable tonal degrees, as represented by the hierarchies encoded in the profiles, or otherwise, they might indicate a modulation, i.e., a change in the hierarchy. However, to prevent an excessive jitter in the output of the algorithm, Temperley imposes a penalty when keys differ between consecutive segments, mimicking the perceptual inertia of remaining in the same key until there is enough evidence of an actual key change.

The various methodologies presented so far were directed towards euroclassical music. More exactly, all the models discussed addressed short melodic fragments, and were evaluated using the 48 fugue subjects from J. S. Bach's *Well-Tempered Clavier*, one of the foundational works of euroclassical tonality. This is probably one of the motivations behind Temperley's critique of Krumhansl & Kessler's aeolian-biased minor profile. In a different publication, Temperley (2001, pp. 258–264) devoted a few pages to the problem of key identification in popular music, suggesting a new profile accounting for rock's diverging modality. Instead of dividing keys into major and minor, he creates a single 'supermode' profile, merging

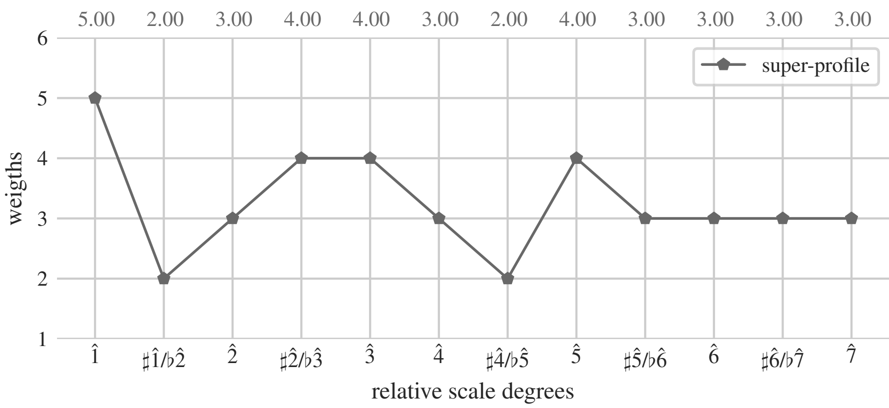


FIGURE 3.6: 'Superprofile' from Temperley (2001), calculated with Lerdahl's basic tonal space principle. Values are scaled as in Temperley (1999), to allow direct comparison.

the ionian and aeolian scales, and bringing together the 'colours' of rock's four common modes (ionian, dorian, mixolydian and aeolian), as it was discussed in Section 2.3.2. This 'superprofile' —the label is ours— can be obtained by applying Lerdahl's basic space, as shown in Figure 3.6. However, one of the main shortcomings of this approach, is that, while modality in rock accounts for various modes, and scalar shifts might occur in the course of a song, as we have seen, particular segments typically do not involve the degree of chromaticism suggested by this profile. In any case, Temperley's experiment brings out an issue about extrapolating specific models to cultural or musical domains that lay beyond the reach of the model. This exactly, has been acknowledged by the author and his colleague De Clercq in a more recent publication on rock harmony, observing that

> "most work on key estimation in popular music has identified keys as major or minor, following the common practice key system. However, we found in creating our corpus that it was often quite problematic to label songs as major or minor [...]. Thus, we simply treat a 'key' in rock as a single pitch-class." (Temperley & De Clercq, 2013 p. 194)

In later works, (e.g. Temperley, 2007a) makes a shift from cognitively-oriented profiles to corpus-driven distributions, highlighting important similarities between both approaches that can be taken as an argument supporting the statistical foundations of musical style. Figure 3.7 shows statistically derived profiles from two different musical collections, the Essen collection, comprising of over 6,000 monodies from European folk songs (Schaffrath & Huron, 1995); and the Kostka-Payne profiles, based on harmonic analyses of fragments of polyphonic euroclassical music from

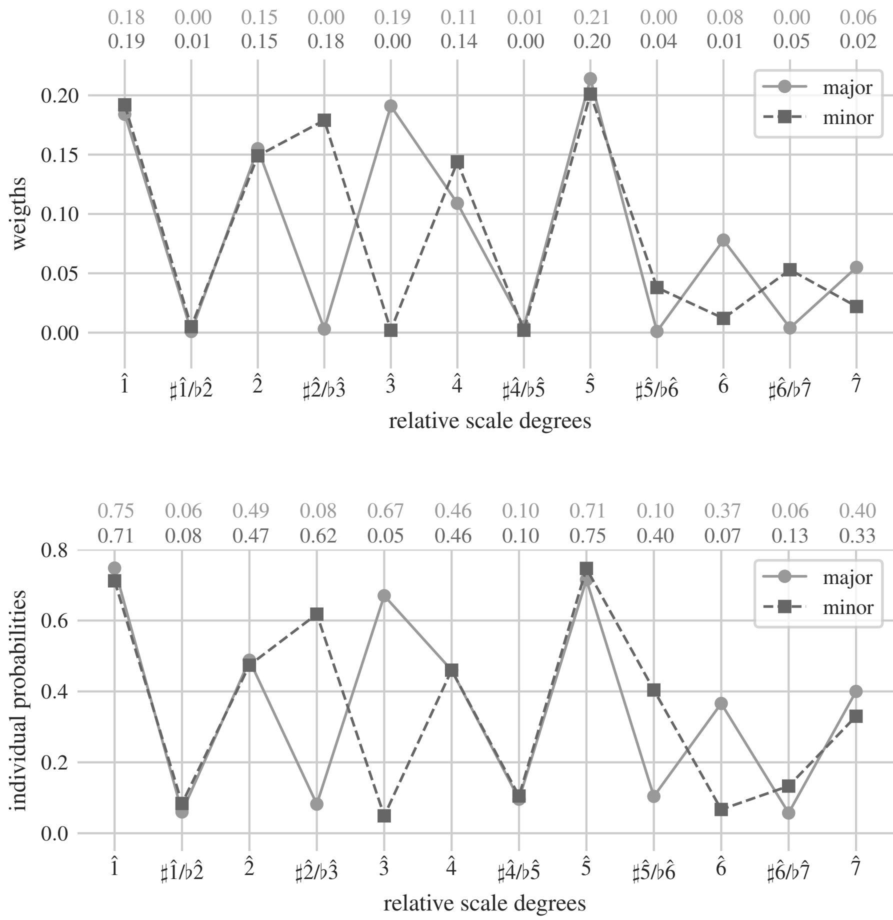


FIGURE 3.7: Statistical key profiles from the Essen (above) and Kotska-Payne corpora (below) from Temperley (2007a). The first one is obtained from single folk melodies, whereas the second is calculated from polyphonic music excerpts of euroclassical repertoire.

the Kostka & Payne Harmony Workbook (1995). It might be interesting to note that the Essen profiles are calculated as a single probability distribution, which could be reversely used as a generative Markov process of zero order, in which each new event is completely independent from previous events (actually Temperley calculated these profiles within a generative/analytical model). Conversely, due to the polyphonic nature of the corpus, the profiles extracted from the Kostka-Payne workbook represent the joint distribution of twelve different variables, corresponding to each scale degree given a particular tonal context. However, and despite their differences, these corpus-driven profiles show remarkable similitudes with those obtained experimentally: they emphasise the tonal organisation in hierarchies, suggesting that style-related variations occur in other modal degrees. For example, the minor profile of the folksong corpus resembles the original Krumhansl & Kessler profiles in that they present more energy in the aeolian $\flat\hat{7}$, what can be taken as distinctive of folk and popular music styles. On the contrary, Temperley's orientation towards euroclassical music is visible in his modification of Krumhansl & Kessler's profiles, which aligns with the profiles of the Kostka-Payne corpus.

In summary, despite the different variations and origins of key profiles, they all aim to represent the pitch hierarchies —and by extension, the pitch relationships— operating in music (whether acquired experientially or learnt from an analytical corpus). Therefore, differences in the weighting of specific scale degrees (e.g. $\flat\hat{7}$ vs. $\natural\hat{7}$), must certainly correspond to the statistical differences between musical styles (e.g. popular vs. euroclassical music). However, stylistic differences aside, the multiple variations of the key profiles might as well simply represent statistical 'noise', in which case key profiles could be reduced to Lerdahl's theoretical basic tonal space (Figure 3.2) as suggested by Temperley (2007a, p. 92).

3.2.3 Other Approaches

As we advanced in Section 3.1, one of the practical applications of Chew's Spiral Array is in determining the key of a musical piece. The 'Center of Effect Generator' algorithm (Chew, 2000, pp. 99–106), frames the task of key finding as a problem of distance-minimisation in the Spiral Array. Like in the method by Longuet-Higgins & Steedman, Chew's algorithm proceeds sequentially on an event basis. However, Chew's algorithm does not need to wait for completion of the analysis excerpt, providing a new estimate at each new step. As pitch classes unfold in the analysis, their respective duration weights are accumulated at each respective chroma *position* in the geometric space (Figure 3.3). In this fashion, the 'center of effect' C is calculated as the sum of all the past pitch positions p weighted by their durations d at any given event in time i ($j = 1$ represents the first note in the musical phrase):

$$C_i = \sum_{j=1}^{i} d_i \cdot p_i \tag{3.3}$$

In Chew's model, chords and keys have fixed positions in the geometrical space, just as much as chromas do: "a chord is the composite result, or effect, of its component pitches. A key is the effect of its defining chords" (Chew, 2001). Therefore, the key of the excerpt is simply calculated as the shorter Euclidean distance between the 'center of effect' C and each key position K, where $C = c_1, c_2 \cdots, c_n$, and $K = k_1, k_2 \cdots, k_n$, in an Euclidean n-dimensional space:

$$d(C,K) = \sqrt{\sum_{i=1}^{i_n}(c_i - k_i)^2} \tag{3.4}$$

In a similar vein, Shmulevich & Yli-Harja (2000) proposed a variation of the method by Krumhansl & Schmuckler to estimate keys in an accumulative fashion, by applying a smoothed sliding window, emulating the effect of short-term memory (making past events contribute less to the estimation at each subsequent step). They locate the 24 keys as points in an Euclidean space, separated by their inter-key distances (the correlation values between all possible keys), and convert the input vector into a spatial representation with multi-dimensional scaling. With this translation, as in Chew's model, the shorter Euclidean distance indicates the closest key.

We find that these two models provide interesting insights regarding the estimation of keys as continuous processes, what seems to stand closer to how a listener or musician operates in reality. In popular music theory, for example, pattern-matching approaches have being criticised exactly for not being able to consider the temporal and accumulative properties involved in musical perception, as expressed in the following quote:

> "In addition to making mistakes in determining keys, these methods [Krumhansl-Schmuckler and Longuet-Higgins] are all flawed in that they do not model correctly the process they are meant to explain. Human beings do not, before surmising the key of a musical passage, wait for the completion of a pitch source or wait for enough notes on which to base a comparison between durations and tonal strengths of pitches. A listener picks up clues from the very first sound she hears, interpreting it in relationship to her vast stores of tonal memories. [...] Although no satisfactory system has been developed of explaining how a key is perceived, the picture seems to be something like this: listeners perceive patterns that their musical memories teach them to associate with a particular key. The first notes heard, even the very first note, suggest as a tonality the key in which they are the most structural members; subsequent notes either confirm the original impression or supersede it with another." (Stephenson, 2002, pp. 31–32.)

3.3 Key Estimation from Audio

Nowadays, the problem of musical key identification has mostly shifted to the audio domain, as we explain in this section. Research problems are multiple, broadly concentrated in the multidisciplinary domain of MIR. The number of application

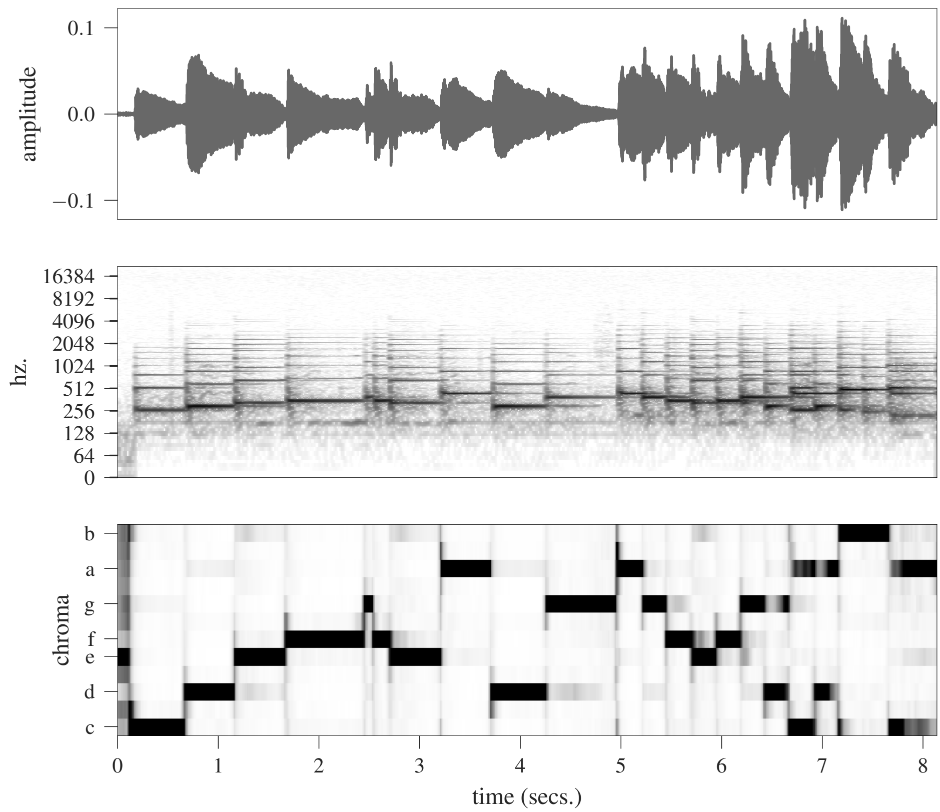


FIGURE 3.8: Various time representations of an audio file containing the first fugue subject from J.S. Bach's *Well-Tempered Clavier*, as recorded by Glenn Gould. The figure on top represents the raw audio signal, as a function of amplitude over time. In the middle, the same signal is represented as a log-frequency spectrogram, whereas in the bottom figure, the file has been transformed into a chroma representation using a constant-Q transform. Audio analyses were conducted with the librosa-python library (Mcfee et al., 2015) with ws = 4,096 and hs = 512 pt. Note that the chroma representation bears a close resemblance with the MIDI piano-roll in Figure 3.4.

scenarios of key characterisation of audio material, including library organisation, recommendation systems and music creation, is probably one of the main appeals towards the task.

With independence of the application domain, extracting pitch and duration information from an audio signal requires a few additional steps in order to transform the digital audio encoding into workable symbols or representations. Once this is achieved, many methods of tonal enquiry do not differ substantially from the ones discussed in the previous section —especially in the later stages of the determination process, as it might be suggested by the transformations shown in Figure 3.8. However, it is worth noting that a direct comparison with symbolic approaches would only be

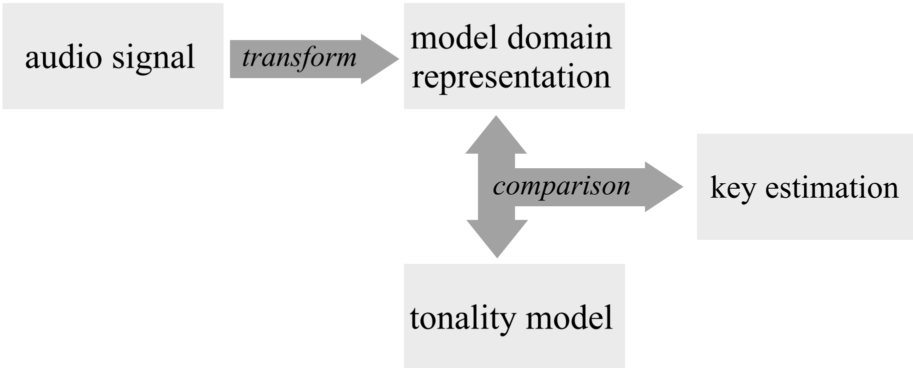


FIGURE 3.9: General structure of an audio key-finding system.

guaranteed by a proper transcription process, as pointed by several authors (Izmirli, 2005b; Peeters, 2006a) what has not been yet accomplished successfully for EDM and most polyphonic music. Therefore, in this section we focus in models that extract tonal information directly from the audio signal, avoiding the transcription process.

Figure 3.9 outlines the essential architecture of an audio key-finding system. As suggested, the first step necessarily consists in the transformation of the audio signal into a workable representation. This representation must be comparable in some way with a previously established model of tonality, be this an array of key profiles, a geometrical space or a different collection of descriptors. In any case, the results of this comparison (e.g. correlation) will determine the key chosen by the system, with or without additional verbose. Most methods perform each of these stages differently, but they all divide the estimation process into these structural steps.

3.3.1 Preliminary Assumptions

Digital audio signals typically comprise thousands of discrete points representing oscillatory pressure waveforms. Good quality audio formats —as represented by the CD standard— digitise audio signals at a sampling rate (R) of at least 44,100 points per second, offering a frequency range of up to 22,050 Hz, slightly above the human perceptive threshold. Besides, standard music distribution formats are normally stereophonic, containing two independent parallel streams of audio data, in order to store some spatial information within the digital representation. For the task of key estimation from audio —as much as for most other MIR tasks— stereo signals are typically merged into a single mono stream by summing the content of both channels, since tonal information neither originates nor depends on spatial information. Therefore, in the remainder of this section, all audio signals described must be assumed mono-aural, and initially sampled at 44,100 Hz.

Although differences in the methodologies to extracting pitch information from audio signals are noticeable, authors normally depart from general assumptions about the nature of harmonic signals. These properties, and the general ways in which they might be addressed in signal processing, are summarised in the following bullets.

- It is assumed that high-frequency components do not carry much information about pitch, and it is common to disregard spectral data above a threshold (think for example, that the highest musical note played by a piano or a piccolo flute corresponds to c8 $\approx$ 4,186 Hz).
- Furthermore, in the range below this threshold, higher partials of harmonic signals contribute less to the perception of pitch. Accordingly, spectral decay functions or spectral envelopes relating to this fact are typically implemented.
- Musical heights are organised and perceived according to logarithmic curves. The distribution of octaves across the frequency range is exponential ($f2^i$), and so it is the internal division into twelve perceptually equal semitones per octave ($\sqrt[12]{2}$). Consequently, the frequency range in spectral representations tends to be split into logarithmic units.

3.3.2 Template-Based Key Estimation Pipeline

The problem of tonal inference from audio has been typically split into two main tasks, key estimation and chord detection, although many authors have addressed both endeavours simultaneously (Catteau et al., 2007; Papadopoulos, 2010; Mauch & Dixon, 2010b; Ni et al., 2012). These two operations require a similar measure of tonal summarisation, generically referred to as chroma-feature or chroma vector, what makes the first steps in the processing pipeline of both tasks essentially equivalent. After all, chords and keys could be regarded as two hierarchical levels of the same of problem, only differing in scope and time-scale.

Given their suitability to model time-series statistics, Hidden Markov Models (HMM) have been one of the preferred techniques to approach chord and key detection endeavours. Sheh & Ellis (2003) were the first to apply a HMM for chord recognition from audio, followed by numerous other publications (e.g. Bello & Pickens, 2005; Papadopoulos & Peeters, 2007). Probabilistic models have also been applied in the simultaneous estimation of other contextual elements. For example, Papadopoulos (2010) uses a HMM to simultaneously estimate, chords, downbeat and keys; Mauch & Dixon (2010b) use a Dynamic Bayesian Network to jointly estimate bass pitch-class, in addition to the three mentioned parameters; and Ni et al. (2012) jointly predict chords, bassline and key.

Focusing on key estimation, approaches using HMM's have typically considered semantic units such as tonal regions (Chai & Vercoe, 2005) or harmonic sequences (Noland & Sandler, 2006; Papadopoulos & Peeters, 2009), although there exist models trained directly with raw chroma vectors (Peeters, 2006b), thus avoiding any pre-assumptions about musical context (instrumentation, style) and/or content (polyphony, harmony, melody, et cetera) beforehand. This type of end-to-end approach has been recently explored with neural-network models, in both chord- (Korzeniowski & Widmer, 2016) and key estimation environments (Korzeniowski & Widmer, 2017), achieving state-of-the-art results. However, most approaches to key finding are based on chromagram profile extraction and template matching. In the remainder of our report, we concentrate in this methodology, as it constitutes the main foundation of our own key detection methods, discussed in Chapter 6.

The basic pipeline of a profile-based tonality estimation method was first given by Fujishima (1999) in the context of a chord recognition system. Broadly speaking, template-based estimation methods usually convert the audio signal to the frequency domain by means of a fast Fourier transform or a constant-Q transform. The spectral representation is then folded into a chroma-based feature representing perceptually equal divisions of the musical octave, providing a measure of the intensity of each pitch class per time frame. For improved results, a variety of pre-processing 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. Figure 3.10 shows the signal flow of a template-based model with all its possible variations, upon which we have organised our explanation.

3.3.3 Time- to Frequency-Domain Conversion

The first step towards tonal analysis of audio signals is to be able to determine the evolution of pitched materials along the time axis. The Fourier transform allows to translate, without any information loss, the time domain into the frequency domain, typically using the discrete Fourier transform (DFT).

Furthermore, it is common to split the audio signal in sequential, sometimes overlapping fragments of short duration, in order to capture the temporal evolution of the signal, in a technique known as the short-time Fourier transform (STFT), and typically computed as a fast Fourier transform (FFT) with Cooley & Tukey (1965)'s method. At this stage, each audio frame is multiplied by a smoothing function of the same size, aimed at attenuating the edges of each data frame in order to remove unwanted spectral components originating in the slicing process. The FFT divides the signal's

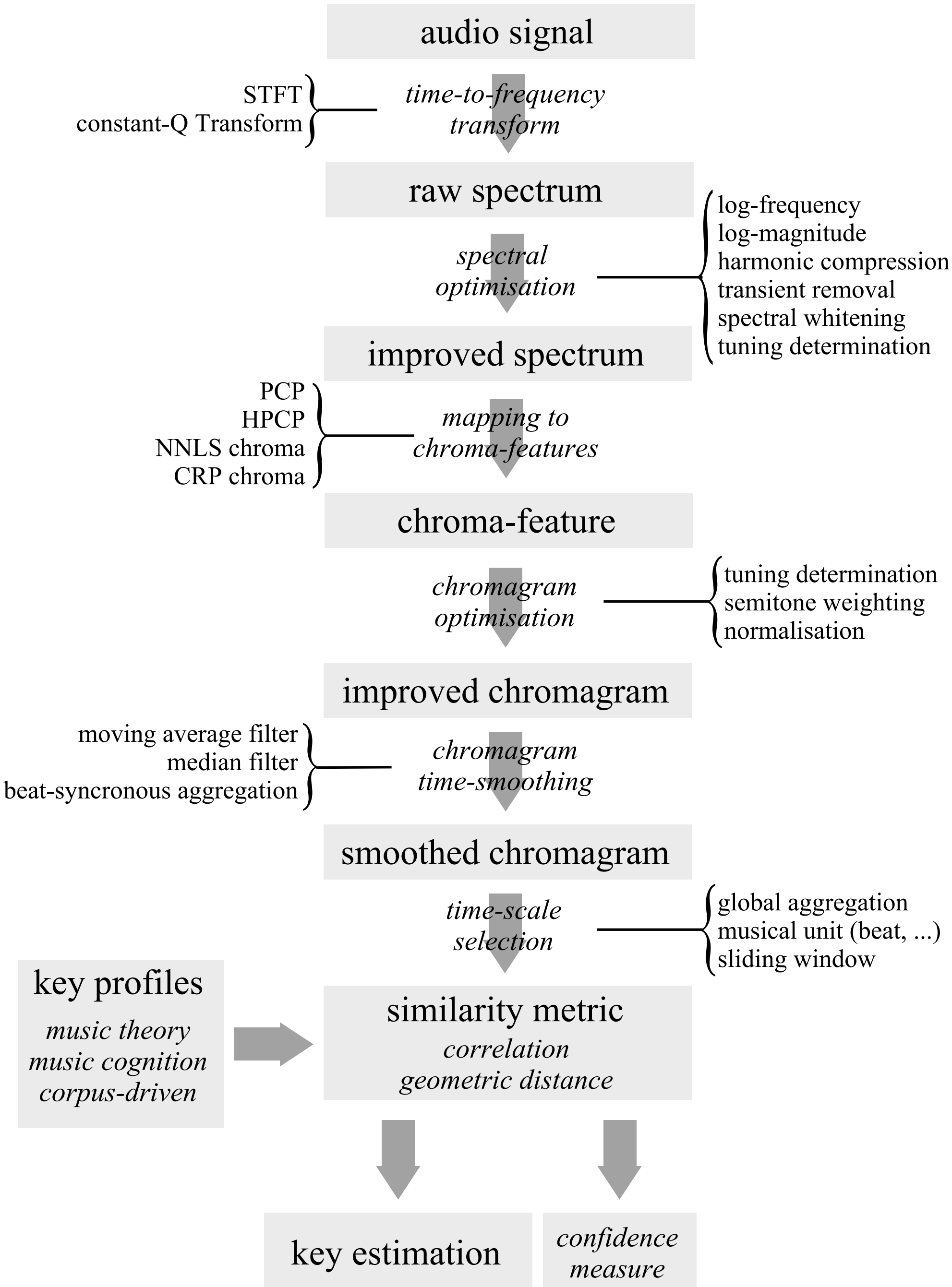


FIGURE 3.10: Processing pipeline of a profile-based system for key determination, with possible intermediate operations. In remainder of this chapter, we detail each of these stages.

frequency range into linear multiples (bins), with a frequency resolution df in Hertz given by

$$df = 0.5\frac{R}{ws} \tag{3.5}$$

dependent on the sampling rate R, typically expressed in samples-per-second, and the window sample-size chosen (ws). Lengthier windows have a finer frequency resolution at the expense of a lower temporal grain. This is often compensated by increasing the windowing overlap, or by downsampling the audio signal, providing a finer frequency resolution, accelerating the computational process, and in turn, discarding higher-frequency components. Noland & Sandler (2006) study the effect of downsampling and other digital signal processing (DSP) in the context of key estimation tasks.

Apart from this type of data reduction, linearly spaced spectrograms are often translated into logarithmic scales, both in terms of magnitude and frequency resolution, in order to provide a closer approximation to human perception. For example, Pauws (2004) uses an arc-tangent function to mimic the pitch-loudness curve of human perception, and Ni et al. (2012) propose a novel implementation based on a loudness-weighted chromagram.

An alternative computational transformation for tonality-related audio is the constant-Q transform (CQT), which roughly splits the signal's frequency range into a series of logarithmically spaced bandpass filters at a constant Q factor, expressed as the ratio between each filter's centre frequency cf and the chosen bandwidth bw (Brown, 1991). An efficient implementation of the Constant-Q transform is provided in Brown (1992), taking advantage of the FFT algorithm. The apparent superiority of the CQ transform over regular FFT methods lays in its finer resolution towards the lower frequencies, and its division of the frequency range into units closer to human perception, what might simplify some subsequent steps in the processing chain. However, both STFT-based and CQ approaches have been used indistinctly in the literature, achieving comparable results in key and chord estimation systems (Kelz et al., 2016).

In any case, the spectrograms obtained by either transform represent the signal in all its complexity, including periodic sounds from note attacks and percussive transients, together with the frequency components that presumably represent harmonics of actual pitches. With this in mind, a variety of techniques have been proposed to isolate tonally meaningful information. For example, Pauws (2004) uses a spectral-peak detection function to discard spurious non-harmonic peaks, and Gómez (2006a) uses a transient detection function to remove short noisy segments from the final chromagram aggregation. With a similar goal, Izmirli (2005b) applies a spectral flatness measure, zeroing windows with a flatter spectrum, and Peeters (2006b) implements a sinusoidal analysis/resynthesis model to reduce transient noise. More recently, harmonic-percussive source separation techniques (Fitzgerald, 2010; Driedger et al., 2014) have been used in chord and key identification tasks (Ueda et al., 2010a; Ni et al., 2012), arising as an optimal processing stage for popular musics with high percussive and transient content, such as EDM. Similarly, harmonic removal techniques (Lee, 2006), spectral whitening algorithms (Schwarz & Rodet, 1999; Röbel & Rodet, 2005) or slightly diverging methodologies (Klapuri, 2008; Mauch & Dixon, 2010a; Müller & Ewert, 2010) have been applied to neutralise the effect of equalisation and other timbral effects before the chromagram calculation (Gómez, 2006a).

Furthermore, the issue of tuning (i.e. that $\mathsf{a}_4 = 440$ Hz) is something that should not be taken for granted, since a good amount of music might fall out of this theoretical reference, including early music in lower tunings, orchestral recordings in slightly higher standards, or popular music simply out of the reference. With this in mind, some methods incorporate a phase of tuning determination over the spectrogram, although other approaches address this problem —and its potential correction— after the chromagram calculation. One common method to address this is by computing simple statistics over spectral components (Dressler & Streich, 2007), spectral-peak histograms (Zhu et al., 2005; Gómez, 2006a), or applying a "modelling error" for various tuning candidates (Peeters, 2006b).

3.3.4 Tonal Representations from the Frequency Domain

Shepard (1964) suggested a widely accepted description of pitch as a combination of two separate properties, *height* and *chroma*, which can be represented separately. This intuition is reflected in the music-theoretical notion of pitch-class, which discards the height dimension establishing the octave equivalence. Similarly, a chroma-based descriptor ignores height by mapping all octaves into a single chroma space. Despite the multiplicity of variants, chroma-features are typically derivations of STFT or CQ transforms, obtained by mapping or folding the spectral representation into an n-dimensional vector, representing the totality of available pitch classes.

Fujishima (1999) originally proposed the pitch-class profile (PCP) for use in chord recognition. The simplest conversion from a full-range spectrum into a chroma representation would be to add the energy contribution of each spectral component directly to its corresponding index in a twelve-dimensional chroma-vector. To improve this conversion, a number of spectral techniques were suggested in the previous subsection. Besides, there are other enhancements that can be inserted during the chroma-feature calculation.

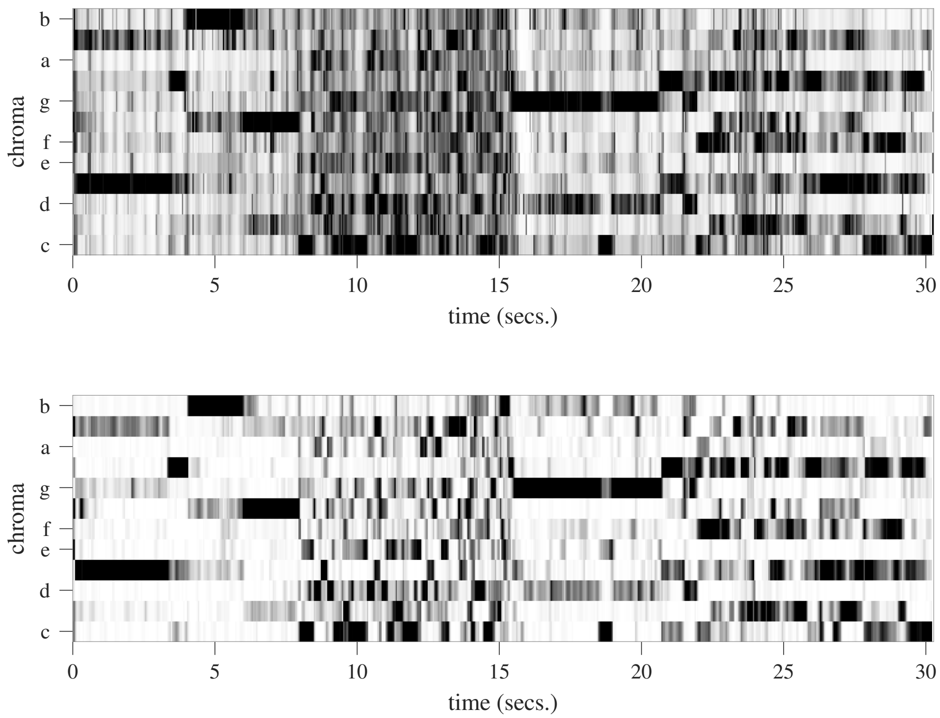


FIGURE 3.11: HPCP (above) and NNLS chromagram (below), from four random four-bar loops from a corpus of EDM tracks. Both analyses were carried in Sonic Visualiser, with the Vamp plugins developed by Gómez and Mauch, respectively, using their default settings. Spectral analysis parameters were set to a hanning window of 16,384 points with hop sizes of 2,048 samples.

A common variation of the PCP calculation is the so-called 'harmonic pitch-class profile (HPCP) proposed by Gómez (2006b), which proposes a few modifications over the original algorithm. Gómez limits the spectral peaks under consideration to the frequency range between 100 and 5,000 Hz, disregarding both the low- and high-ends of the spectrum, and making the computation slightly lighter. As an input parameter of her method, the number of spectral peaks to consider can be manually selected. Besides, the chroma-feature size is increased by a factor of three (3×12), obtaining a finer resolution of 36 divisions per octave. The major difference with the PCP method, however, lays in the 'folding' procedure of the spectral peaks into the chroma-feature. In Gómez's approach, each bin in the HPCP receives the contribution of various frequency components according to a weighting cos^2 function centred at the bin's frequency, with a window length expressed in semitones (defaulting to 1.333 semitones), which allows each frequency component to contribute in different proportions to various chroma-bins. According to Gómez, this procedure reduces the errors produced by inharmonic components in the signal.

The main advantage of finer chromagrams with 24 or 36 divisions per octave is that they can be used to detect and eventually 'adjust' the tuning of the chroma-feature, either by discarding bins falling out of the reference frequency (Harte & Sandler, 2009), or by applying median filters to shift the energy towards the actual semitone frequencies (Peeters, 2006b). Harte & Sandler (2009) proposed a tuning detection method based on a 36-step chromagram, which they subsequently convert into a 'tuned' 12-bin chromagram. Harte & Sandler accumulate the peak positions per semitone (with a resolution of 3 bins) and perform quadratic interpolation to find the semitone index with the maximum peak, which they regard as the tuning frequency. Afterwards, they create a 12-semitone chromagram with the peaks that are multiples of the estimated tuning frequency.

Other chroma-extraction procedures attempt to minimise the effect of timbre or harmonics in the final vector. One of the most successful approaches to overtone removal is represented by the non-negative least squares (NNLS) chroma' by Mauch & Dixon (2010a). In their approach, the authors detect the fundamental frequencies of pitches from a log-frequency spectrogram, which are then used as query entries in a manually curated dictionary of idealised note profiles, consisting only of pure harmonics, from which they derive the final 12-bin chromagram. Alternatively, Müller & Ewert (2010) proposed a timbre-invariant chroma-feature (the CRP chroma) using a discrete cosine transform to discard timbral information, as represented in the lower bins of an MFCC.

Additionally, and independently from the chosen chroma-feature, instantaneous chromagrams are often normalised with regard to the energy in the frame, to make them robust to dynamic changes, typically with L^1, L^2 or L^∞ norms, according to Cho & Bello (2014) and McVicar et al. (2014). Figure 3.11 shows a comparison of two of the described chroma-feature models, a HPCP (above) and a NNLS chroma (below), on four different random loops from a corpus of EDM.[42]

The final step in the chroma-feature calculation typically involves a smoothing function over consecutive individual chroma-vectors. Given that tonal units are normally of a duration longer than a single analysis window (comprising full beats, bars, and/or hypermeasures), groups of frames are often aggregated together, either with moving average filters (Fujishima, 1999; Lee, 2006), sliding median filters (Papadopoulos & Peeters, 2007; Harte & Sandler, 2009), or, particularly in popular music approaches, with *beat-syncronous* aggregation, averaging together chromas belonging to regular musical durations, typically obtained with beat-detector algorithms (Bello & Pickens, 2005; Mauch & Dixon, 2010b; Ni et al., 2012). It is also worth noting that Mauch & Dixon (2010b) proposed the calculation of a separate bass-chromagram addressing specifically the bass-layer, an approach also followed by Ni et al. (2012).

[42]The four tracks belong to the GiantSteps key dataset, which will be described in Chapter 4. Their unique file identifiers are, from left to right, 2018991, 2436276, 3005030 and 3415063.

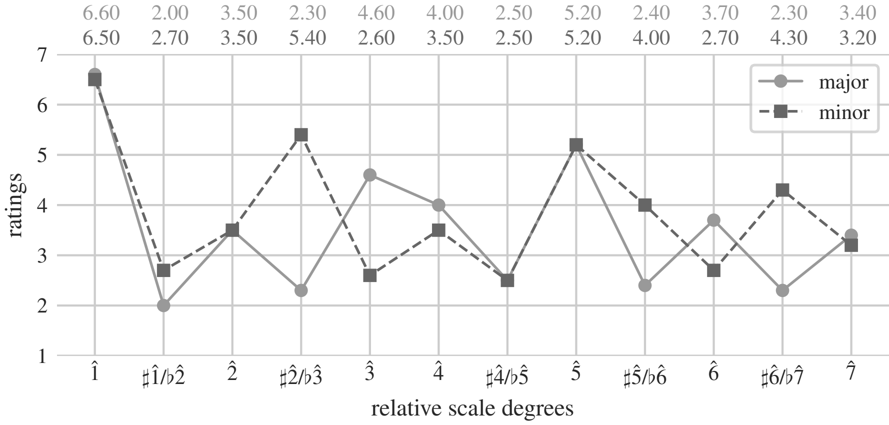


FIGURE 3.12: Modification of Krumhansl & Kessler's key profiles by Sha'ath (2011) for EDM.

3.3.5 Templates of Tonality for Audio

Adaptation of Theoretical Models

Given the complexity of audio signals, and despite the processing steps reported in order to obtain the clearest representation of pitch distributions, the theoretical key profiles discussed in Section 3.1 are often 'adapted' to account for the complexity of harmonic signals, in contrast with the simple symbolic representation of musical tones.

Early audio key-finding methods were directly based upon the profiles by Krumhansl & Kessler without any further transformation. It is the case of Pauws (2004), who instead tries to model the chromagram according to human auditory sensibility curves. Similarly, Sha'ath (2011) makes small heuristic modifications on the original profiles, in order to obtain better results in his corpus of EDM. As shown in Figure 3.12, the main differences between Sha'ath's profiles and the original probe tone weightings are a slight boost of the weight for the $\natural\hat{7}$ in major and a relatively significant increment of the subtonic ($\flat\hat{7}$) in minor. Other than these, the two profiles remain essentially identical. Although Sha'ath does not provide any musicological grounding for his modifications, he is actually favouring ionian and aeolian modalities.

However, other approaches perform modifications in order to incorporate the nature of complex harmonic signals in the tonality profile. For example, Gómez (2006b) adapts the Krumhansl & Kessler profiles to give account of (a) *polyphony* and the (b) upper partials of the fundamental tones. Gómez constructs her polyphonic profiles by adding together the respective weights of the tonic, subdominant and dominant chords at the degrees belonging to each chord. Additionally, she considers the first four harmonics of each scale degree —possibly because the fifth harmonic already falls out of equal temperament— and adds each harmonic contribution to the respective scale position weighted by a exponential decay factor s experimentally set to 0.6. A similar approach had been previously used in Purwins et al. (2000) accounting only for the third harmonic (i.e. a fifth).

An alternative method is suggested by Izmirli (2005b), who creates tonality templates in a twofold manner. First, combined templates are obtained as the multiplication of the flat and modified profiles proposed by Temperley (1999), obtaining a key profile with zeroes in the chromatic degrees. Then, spectra of single piano notes in range a_1–b_5 (with a decreasing function mimicking the less frequent occurrence of higher notes) are used to create 24 'spectral profiles' accounting for all major and minor keys, by multiplying them with the combined profiles, which are finally averaged chroma-wise into a final chroma template for each candidate key.

Statistical Profiles

Apart from the variations upon the cognitive and theoretical models described, other authors have proposed the construction of tonal profiles based on direct analysis of musical recordings, just as Temperley had done with the Essen and Kostka-Payne score collections (Figure 3.7). One of the advantages of statistical profiling from audio recordings, is that it prescinds of the adaptations required by symbolic models, since spectra of rich sounds are already embedded within the model representation. In this sense, these approaches bridge the fracture between the model (key profile) and the input vector (chromagrams) present in other strategies. With statistical distributions, the tonality model is typically represented by an 'idealised' or averaged chromagram, so the similarity assessment needs no further adaptation. Furthermore, chances of an optimal performance increase when the model and the input vectors are calculated with the same parameters and methodology, as already pointed by Izmirli (2005b) and Noland & Sandler (2009).

On the other hand, corpus-driven profiles are, at best, biased to specific musical genres rather that aiming at all-purpose solutions, since tonal and timbral features vary considerably among styles. A recent experiment by Korzeniowski & Widmer (2017) showed that merging corpora of various musical styles considerably lowered the performance for all the involved genres, something that can be easily pictured by mentally comparing a polyphonic keyboard fugue from the Eighteenth Century to an EDM track made with synthesisers, several layers of percussion, tonal glissandi, and spoken voices, 'contaminating' the tonal model.

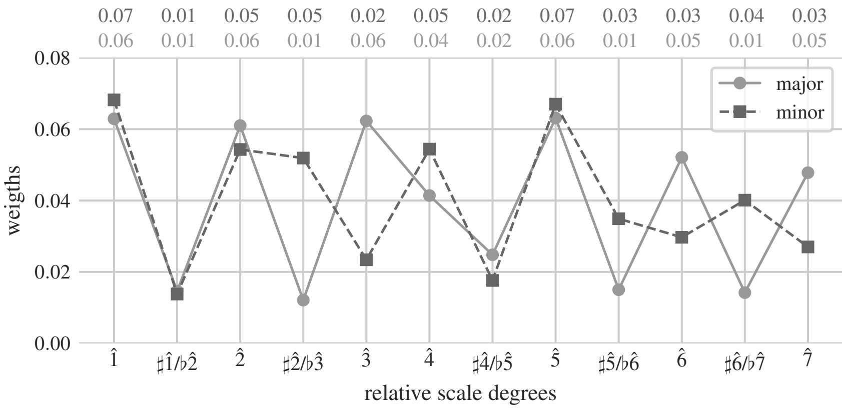


FIGURE 3.13: Key profiles from Bach's *Well-Tempered Clavier* (Noland & Sandler, 2007).

Gómez (2006a) already signals that corpus-based methodologies typically carry genre specificities, and that cross-stylistic evaluation poses some conceptual problems. Furthermore, she acknowledges —in 2006— that EDM presents a severe challenge for existing algorithms, an aspect to which we return in Chapter 4.

Amongst existing corpus-driven profiles from musical recordings, Noland & Sandler (2007) extracted major and minor models from Glenn Gould's recordings of the first book of Bach's Well-Tempered Clavier, comprising of 24 preludes and 24 fugues in all euroclassical keys. Noland & Sandler use a constant-Q transform that is subsequently folded onto a single-octave chromagram with 36 bins (3 per semitone), adding together the contributions of equivalent pitches across all octaves. After the chromagram calculation, they sum together all chromagrams in a given track, rotating the resulting vector so that the tonic of each piece is represented by the first bin in the vector. Each of the resulting profiles is weighted by the duration of each piece, and summed together with all other pieces with the same modality. The two resulting profiles, one major and one minor, are finally normalised so that values add up to 1.

With regard to models specifically addressing EDM, Faraldo et al. (2016a, 2017) derive statistical profiles from several corpora of electronic dance music, as it is detailed in Chapter 6.

3.3.6 Key Determination

Global vs. Local Analysis

Most approaches towards audio key identification have considered the global scope of complete musical excerpts, providing a single estimation label for each analysed item (e.g. Pauws, 2004; Izmirli, 2005b; Gómez, 2006b; Peeters, 2006b; Korzeniowski

& Widmer, 2017). This, as we have seen in Chapter 2, seems to be a good characterisation for euroclassical music, where, despite modulation, compositions are regarded as conveying a single tonality, ultimately expressed by the sequence of tonal regions throughout the piece. Authors attempting a global characterisation of euroclassical music, typically address the first seconds of the audio signal where the key of the piece is stablished unambiguously, in order to avoid falling into modulation processes, although there are approaches that explicitly address the detection of modulation cues (Purwins et al., 2001; Chai & Vercoe, 2005).

In popular music styles, the notion of tonality is somewhat laxer. The assumption of a unifying tonality does not necessarily apply, and different parts (e.g. versus and chorus) can convey different tonal centres without holding specific structural implications. Therefore, although considering the first seconds of a piece of popular might work for some items, it does not align conceptually with the nature of the music, which, tonal considerations aside, might as well start with un-pitched introductions, such as a drum pattern, unthinkable in euroclassical music. Perhaps with this in mind, Noland & Sandler (2009) attempted both local and global characterisations, by modelling short-term harmonic sequences with a Hidden Markov model, what seems to fit conceptually with the tonal structuring of most popular songs.

In the domain of electronic dance music, on the other hand, the differentiation between local- vs. global-key estimations might appear as irrelevant, given the almost total absence of modulation processes, and its structural organisation based on accumulation rather than in alternating parts.

Confidence Measures

Approaches aiming at a global characterisation commonly adopt the Pearson correlation method described in Section 3.2.2. Besides the highest rank, indicating the chosen key of the fragment, some authors use the difference between the first and second correlation values as a measure of confidence —unambiguity— of the key estimate (Gómez, 2006b; Izmirli, 2005a). Other authors have demonstrated that a measure of cross-entropy provides comparable results (Temperley, 2007a; Temperley & De Clercq, 2013), and yet others have employed geometric distances as a similarity calculation. For example, Sha'ath (2011) uses a cosine distance to compare the input vectors to the tonality templates, and Chuan & Chew (2005b) consider the minimum Euclidean distance of chroma features in the Spiral Array as the key determining factor. However, distance-based methods (e.g. nearest neighbour) have been mostly employed in spatial representations or approaches with larger vocabularies, such as chord type dictionaries (e.g. Fujishima, 1999) or wider modal classes, in the context of world music analysis (Chordia & Senturk, 2013).

The global-key estimation process typically proceeds by averaging the entire analysed fragment (Gómez, 2006b; Faraldo et al., 2016a), although silent or flat chromagrams are typically left out of the computation process. Noland & Sandler (2006) and Izmirli (2005a), alternatively, consider individual key labels and confidence measures for each temporal frame, obtaining a global estimation by summing the confidence values of each key estimated.

Tonal Vocabulary

All the algorithms discussed in this section, whether operating at a global or a local scope, have a limited vocabulary of 24 keys (12 tonics $\times$ 2 modalities). While this might be optimal for the euroclassical tradition, where these modalities have their origin and genuine expression, there is evidence that this classification does not result appropriate for popular musics, where tonal centre identification might be more aligned with the natural ambiguity of some popular music genres (Temperley & De Clercq, 2013). This shortcoming could be addressed by, for example, increasing the verbosity of the algorithms, either by increasing the modal labels to the four rock modes, or by providing some details regarding other salient aspects, such as the major/minor modal ambivalence. We believe that EDM-oriented algorithms, could also benefit from this slightly diverging scenario.

In sum, in subsequent chapters, we attempt to adapt template matching methods to the musical and timbral particularities of EDM, that will be described in Chapter 5, deriving statistical profiles and expanding the tonal classifier to provide finer modal information in Chapter 6.

***

In this chapter we have described various scientific approaches to the identification of key, from its psychological reality to its materialisation in computational methods, both operating in symbolic and non-symbolic domains, with a particular focus on template-matching approaches. The choices of analysis scope and tonal vocabulary seem to be of great importance in the development and evaluation of key estimation methods, and there are signs indicating that these should differ when addressing different musical styles. However, to a great extent, these aspects are pre-determined by the availability or research corpora with the required degree of analysis. This and other methodological concerns are addressed in the next chapter, where we report on available corpora for tonal analysis as well as on regular evaluation conventions. Furthermore, we present a preliminary evaluation with state of the art methods, supporting the plausibility of a closer look to tonality in EDM.

Chapter 4

Methodology

"In that empire, the art of cartography attained such perfection that the map of a single province occupied the entirety of a city, and the map of the empire, the entirety of a province."

Jorge Luis Borges, *On Exactitude in Science* (1946)

The primary concern of this dissertation revolves around automatic key identification in electronic dance music. With that purpose, we embarked on a study to identify characteristic tonal practises in EDM, an endeavour that, to our knowledge, has only been addressed superficially (Wooller & Brown, 2008), and that is the object of Chapter 5.

One of the imperatives of computational musical analysis is the availability of a representative body of valid and reliable data, typically originating in human knowledge or empirical evidence. These corpora are subsequently used in the development and evaluation of proposed analytical methods. With this in mind, this chapter surveys existing music collections with computer-readable tonal information, that are accesible to the MIR community. Datasets with key annotations include euroclassical music (mostly relying on the habit of naming compositions with the key on the title), popular music, and a number of scattered labels of EDM from different websites. An important effort of our current research has gone into curating, collecting and analysing a corpus of electronic dance music, with the purposes of identifying specific tonal practises in EDM, developing and testing our research algorithms. One of the outcomes of such endeavour —the GiantSteps key dataset— has been already published (Knees et al., 2015) and referenced in a number of publications (Faraldo et al., 2016a, 2017; Bernardes et al., 2017b; Korzeniowski & Widmer, 2017), and it is conveniently introduced in the section dedicated to EDM test collections.

The second part of this chapter is devoted to discussing validation practises and other methodological aspects, regarding the evaluation of key-finding algorithms, whereas Section 4.3 offers a preliminary evaluation of available key detection systems with the described datasets. With this operation, we intend to make an argument supporting the need of analysis methods tailored to specific musical genres —EDM, in this particular— in dialogue with music-theoretical enquiry.

4.1 Music Collections

As reported in Chapter 3, tonality has been an active area of research in MIR, creating a demand of test collections complying with different research goals. Tonally-annotated datasets normally include one or more of the following marks: structural sections, keys (either globally or indicating key changes) and chords, although melodic annotations could be considered tonal observations too (e.g. Temperley & De Clercq, 2013).

Although the importance of well-formed corpora and test datasets for information research is capital, literature on the topic is not abundant, originating mainly in the areas of linguistics and speech processing (MacMullen, 2003). Regarding music information research, the issue of corpus formation has been been addressed by Peeters & Fort (2012) and Serra (2014), who makes an important differentiation between *test datasets* (a collection of annotated data to be used in a particular experimental framework) and *research corpora*, which normally include wider efforts devoted to capture essential aspects of a particular musical practise. Serra isolates five important criteria in the creation of research corpora: purpose, coverage, completeness, audio quality and reusability. We have condensed them in the following bullets, according to the needs of the current research.

- Availability of good quality audio paired with metadata labels (*quality, completeness* and *reusability*). This is not always possible due to copyright law infringement, what is sometimes compensated with an unambiguous reference to particular audio releases, or by offering alternative ‘encodings’ such as time-series of spectra or chromagrams instead of the actual sound files.

- Empirical evidence of the data labels (*purpose*, *reusability*). This is normally achieved through a process of manual annotation, preferably by more than one subject, especially in domains —such as key labelling— open to multiple interpretations. In some cases, the process of annotation is partially supported by an automated task, under the supervision of a human expert.

- Representativeness and significance of the collected data (*coverage*), which should represent in statistically significant terms the variety and specificity of a given repertoire. This is a crucial requirement, since observations stemming from the study of a given corpus, are necessarily contained within the boundaries of the provided data.

As we will see shortly, the available research datasets do not always fulfil these requirements. In any case, in the following paragraphs we report on the available corpora for western tonality-related MIR, grouped into three broad stylistic categories, namely euroclassical, popular and electronic dance music.

4.1.1 Euroclassical Music

Euroclassical music has been typically described as presenting a *main key*, major or minor, which is established at the beginning of a composition, abandoned through modulatory processes throughout its development, and typically recalled in the conclusion. This main or global key, is normally expressed in the title of the piece, or at least, in the score's key signature.

This simple fact has favoured —and conditioned— many key estimation methods addressing euroclassical tonal estimation, and most procedures normally consider only the first seconds of the analysed sound file (between 2.5 and 20). In this sense, virtually all euroclassical music repertoire could be used for key estimation, and, as a matter of fact, different authors tend to use different musical sources by merely considering the key in the title as ground truth. For example, Pauws (2004) uses a combination of keyboard music from Bach, Shostakovich and Chopin, whereas Izmirli (2005b) takes a random selection from the Naxos Records streaming service,[43] and Peeters (2006a) benefits from a database of "European baroque, classical and romantic music." It has been an extended practice to use Johann Sebastian Bach's *Well-Tempered Clavier* (Pauws, 2004; Gómez, 2006a; Noland & Sandler, 2007), for it presents an even distribution of all 24 major and minor keys, plus it is considered one of the fundamental oeuvres laying down the foundations of euroclassical tonality.

For this research, we occasionally take advantage of an in-house test collection, the 'Classical DB', previously compiled by Gómez (2006a). This dataset contains 881 audio tracks comprising keyboard, chamber and orchestral music from the Common Practise period, labeled after the key in the title of each piece or movement.

[43] http://www.naxos.com

4.1.2 Popular Music

In the MIR community, most efforts towards the creation of tonality-related corpora have been directed towards popular music. This is probably an effect of several circumstances, such as the lack of written scores providing additional metadata and the wide interest in automatic chord recognition tasks.

The largest accomplishment in this direction is materialised in the Million Song Dataset (Bertin-Mahieux et al., 2011), a collection providing an extensive list of audio-feature descriptions, including tonal information.[44] However, the metadata accompanying each entry is algorithmically extracted by *The Echo Nest* and lacks human validation.[45] Furthermore, the lack of available audio makes this corpus unsuitable for training and evaluation endeavours. Another relevant contribution, conceived as an enduring effort in which annotations are expected to grow both in number of items and annotated parameters, is the SALAMI project (Smith et al., 2011), providing structural annotations for over 1,400 recordings from various sources, including the Internet Archive[46] and other published datasets, like the RWC dataset (Goto et al., 2002, 2006). The structural manual annotations are complemented with additional audio descriptors, also taken from *The Echo Nest*. Among the additional features, a global-key estimate is provided, but again, it is inferred algorithmically.

A related compilation, with relevant tonal information, is the McGill-Billboard dataset (Burgoyne et al., 2011), a collection of 742 unique songs from US billboard charts, containing popular music hits from the period between 1958 and 1991. The Billboard datasets encodes, besides SALAMI-style structural annotations, metric and chordal information. Furthermore, although the authors are not permitted to release publicly the related audio, they provide timed chromagrams of the audio, and claim to be open to extract other features on demand. Recently, Korzeniowski & Widmer (2017) obtained a subset of this dataset with 625 global-key annotations, by discarding songs with multiple tonics or with ambiguous modality (less that 90% of tonic chords in the same mode), that we will use in subsequent experiments.

More modest in number of items, the Isophonics dataset results as the union of different forces around the Queen Mary University in London, gathering structural, metric and tonal descriptions of pop music (Mauch et al., 2009b). References with key annotations include the complete discography by The Beatles (180 songs transcribed by Harte, 2010), 18 songs by Queen (from 'Greatest Hits' compilations), Carole

[44] https://labrosa.ee.columbia.edu/millionsong

[45] The Echo Nest was a digital platform providing online automatic analysis of audio and musical features. In March 2016, it was acquired by Spotify.

[46] http://www.archive.org

A:	I vi | IV V |
In:	$A*2
Ch1:	$A*4
Ch2:	$A*3 I IV | I |
Ch3:	$A*3 I |
Br:	IV | I | IV | I | IV | I | V/V | V |
S:	[Ab] [12/8] $In $Ch1 $Ch2 $Br $Ch2 $Br $Ch3

FIGURE 4.1: De Clercq's harmonic transcription of The Penguins's *Earth Angel*, illustrating the syntax proposed by Temperley & De Clercq (2013).

King (7 tracks from her album *Tapestry*) and Zweieck & die Herzrhythmus-Combo,[47] presumably annotated or revised by Mauch (2010). The online repository,[48] provides reference to the exact audio releases used for the transcriptions, although it recommends to use the key annotations with care (The Beatles) and "moderate confidence".

Temperley & De Clercq (2013) provide harmonic and melodic annotations for a collection of 200 rock songs (RS), chosen after Rolling Stone magazine's selection of "500 Greatest Songs of All Time" (a continuation of the work initiated in De Clercq & Temperley, 2011).[49] The main goal of their annotations is to study statistical trends in rock music at various levels, so they propose a labelling framework that is interpretable as a series of rewrite rules, capturing essential structural traits without redundancy (like a repeating chord sequence), while simultaneously linking them to actual song renditions. Figure 4.1 shows one of such harmonic transcriptions. In this example, the two-bar chord sequence A (in Roman Numeral notation) is repeated twice in the introduction (In) and four times in Ch1 (the $ sign indicates the substitution operation). The next two choruses (Ch2, Ch3) present different cadential endings, and the chords in Br represent a novel bridge section. The last line S (song) is reserved to unveil the song's tonal center and the structure of the song, written as a sequence of substitution signs. This way, the authors guard themselves from making assumptions about the modality of a song (at least in dual terms) leaving the capacity to infer the modality to a particular parsing algorithm. This notation convention also allows the authors to write modulations relative to a central tonic. However, the main argument supporting this annotation method is the semantic limitation of a dual modal system given the particularities of rock's modal system, as we have discussed in Section 2.3.2.

[47] Definitely not a mainstream band. An online search did not provide much information about it, other than a online listening service to their —apparently only— album *Zwielicht*, containing the 18 songs transcribed by Mauch in this collection.

[48] http://isophonics.net/datasets

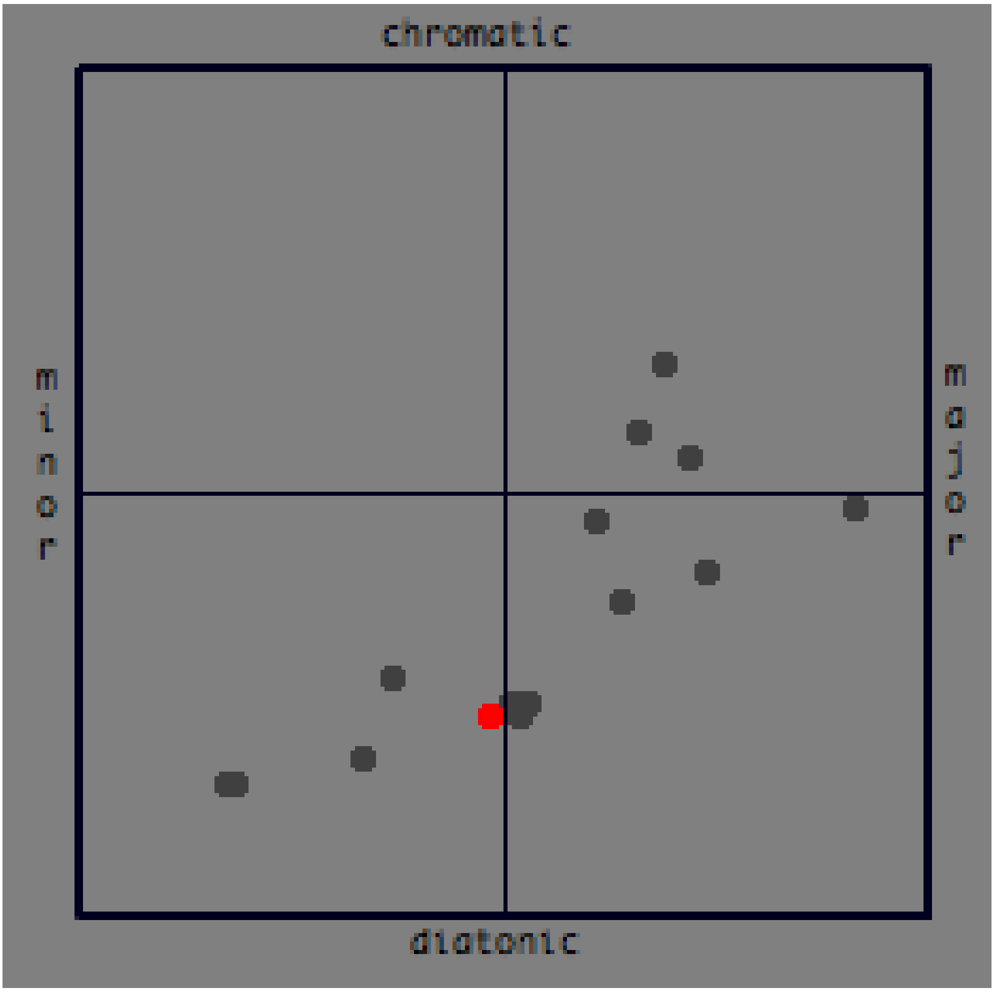
[49] Analysis data and computer programs to help parsing the corpus are currently available in the following website, and not in the one reported in the publication: http://rockcorpus.midside.com

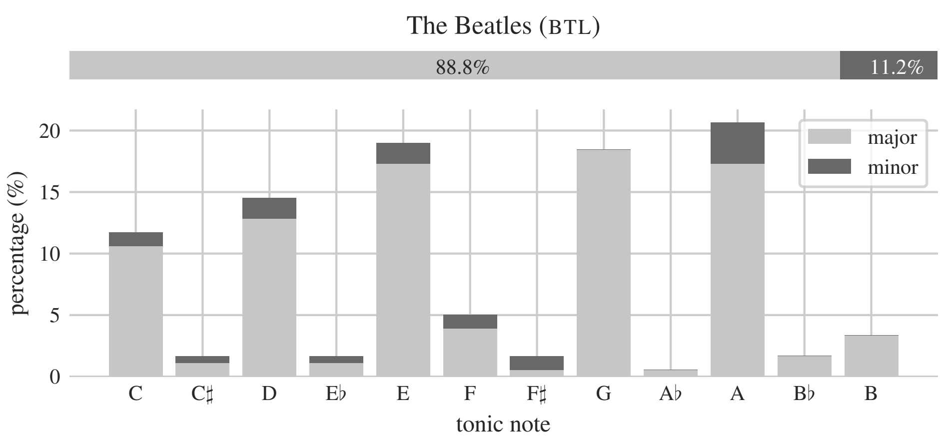
Other interesting resource has been published by Di Giorgi (2013), and comprises of the first five albums by Robbie Williams, totalling to 65 songs annotated with chords and key changes with four different modal variants (major, mixolydian, minor and dorian).[50] The 'major' label refers exclusively to the ionian scale; however, although the 'minor' label is in principle associated to aeolian (Di Giorgi, 2013, p. 21), it seems plausible that it could also denote other minor variants (harmonic) not discussed at all in the publication. In any case, Di Giorgi's research is oriented towards beat-aligned chord detection, and the diatonic modal frame proposed seems intended to extend the chord vocabulary in relation to a tonic triad (for example, ♭VII and ♮VII in a C context). Regarding chord identification per se, Barbancho et al. (2013) prepared a large experimental dataset with 275,040 piano chords and various degrees of polyphony. However, this data does not provide any contextual key information.

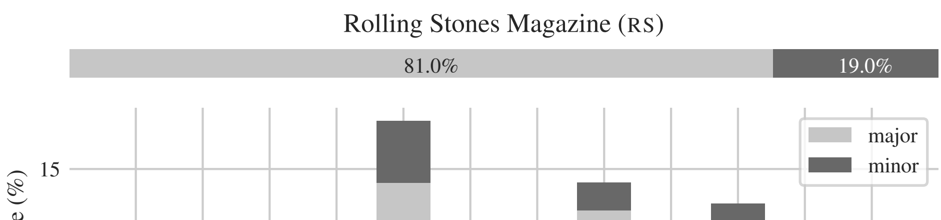
Figure 4.2 shows the distribution of keys in the three different pop music datasets. All the collections present a strong bias towards major modalities (85% on average) and tend to focus on natural tonal centres. This is particularly clear in The Beatles dataset (BTL), where most keys correspond to guitar open chords (Cmaj, Dmaj, Emaj, Gmaj, Amaj), what can be seen as indicative of the importance of the instrumental medium in the compositional process. As stated above, the McGill-Billboard key dataset (BB) comprises of 625 unambiguous tracks without key changes. That suggests that from the total of 742 entries in the corpus, the remaining 15.8% are either tonally ambiguous, or present at least one key change. Similarly, from the 180 songs by The Beatles, 21 songs ($\approx$ 11.3%) contain at least one key change, typically correlated with a structural change in the song, although we assume the single key reduction by Pollack (1999) for this statistics. A closer examination of Harte's annotations (Mauch et al., 2009a), reveals that seven tracks are annotated as modal variants (4 mixolydian, 2 aeolian and 1 'modal'). This brings up an important aspect of analysing popular music: Rock modality typically presents an array of scale variants beyond the ionian/harmonic euroclassical system, These variants, however, can be broadly grouped into major (ionian, lydian, mixolydian) and minor (aeolian, dorian, phrygian, harmonic), mostly according to the type of their tonic chord.

The RS dataset has been analysed independently by two experts with slightly divergent approaches. For example, one annotator (De Clercq) tends to annotate local key changes, whereas the other (Temperley) normally analyses them as applied chords. In total, Temperley annotates 31 tracks with key changes vs. the 35 by De Clercq, around 16.5% of all the songs, in which cases we have taken the predominant key as global. Despite these minor differences, the agreement between the two annotators is very high (93.3%, according to their paper). For example, in the case of Led Zeppelin's

[50] http://www.researchgate.net/publication/260399240


chromatic
minor
major
diatonic


The Beatles (BTL)
88.8%
11.2%
major
minor
percentage (%)
0
5
10
15
20
C
C♯
D
E♭
E
F
F♯
G
A♭
A
B♭
B
tonic note


Rolling Stones Magazine (RS)
81.0%
19.0%
major
minor
15

"Whole Lotta Love", referenced in Section 2.3.3, authors annotate the tonic chord differently, as major and minor (Emaj vs. Emin) showing an interesting disagreement in relation to the ambiguity of power chords and rock's major/minor modal merge, However, it is precisely for this type of songs that they have decided to report the key as the tonic note only. Other difficult entry, although much more 'classical', is Queen's *Bohemian Rhapsody*. After a start on B♭ major, the song evolves alternating regions of E♭ and B♭ (with a short bridge in A major). However, the song ends in a surprising change to F major. Therefore, although we have chosen E♭ major as the most prominent key (the longest region in duration) this type of song shows how approaches based on a single global-key annotation might say very little about the music under consideration.

4.1.3 Electronic Dance Music

Regarding EDM, the biggest effort in building a test dataset probably comes from Sha'ath (2011), who initially released a list of key annotations for a hundred tracks in order to develop his software *KeyFinder*. Recently, he expanded this list to a thousand entries with the help of three human experts, to which we refer as the KeyFinder dataset (KFD). Although access to audio files is not public due to copyright issues, the annotations with one global-key estimate per audio track, and a modal vocabulary of major and minor are freely provided in his website[51]. In the course of our research, we managed to obtain 998 of the total tracks as MP3 files at variable bitrates (128–320 KBPS), although there might be differences in some audio files with regard to Sha'ath's personal audio collection, especially when it comes to remix versions. Besides, a closer examination of the annotations reveals that around 200 entries in the collection represent other popular music styles such as reggae, rock or soul, including songs by artists like Aretha Franklin, AC/DC or Bob Marley.

Figure 4.3 shows the distribution of global keys in the KeyFinder dataset (KFD). It is notorious the great presence of minor keys (85.1%), contrasting with the modal distributions in euroclassical music or pop. Although we do not have numbers regarding the rate of modulations in this dataset, we are inclined to think that these will be less common than in other popular music styles, given that typical alternating 'verse-chorus' structures so common in pop-rock are essentially absent in many EDM subgenres (Garcia, 2005). Besides, Figure 4.3 shows a much more even distribution of tonics along the 12 chromas, compared to the other musical styles discussed, what might be indicative of production and creative techniques centred around synthesisers, digital tools, and eventually keyboards, in contrast with the guitar-centric distribution of keys in popular music datasets.

[51] http://www.ibrahimshaath.co.uk/keyfinder/KeyFinderV2Dataset.pdf

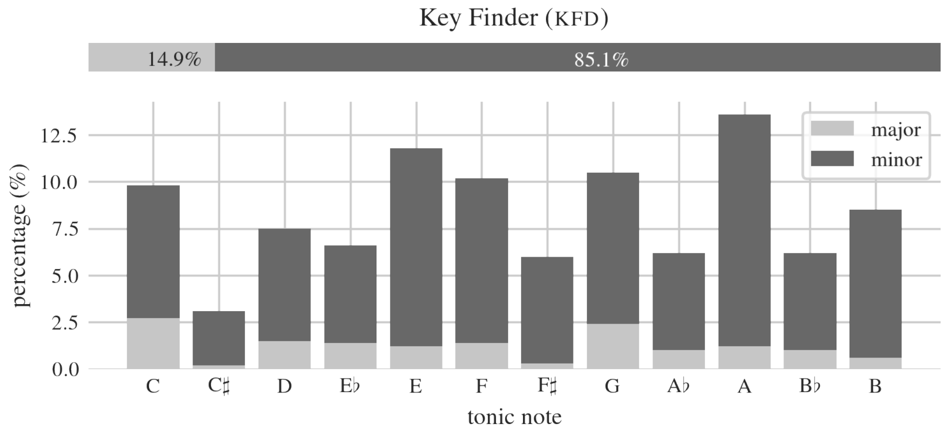


FIGURE 4.3: Distribution of global keys in Sha'ath's KeyFinder dataset (KFD).

Other labelled sources on the internet come from DJ magazines and online software reviews. For example, the web platform *Djtechtools*, has a series of entries reviewing available key estimation software in 2009,[52] 2012,[53] 2014[54] and 2015.[55] Similarly, DJ Endo annotated a collection of EDM tracks with the same purpose in 2011 (ENDO^{A})[56] and 2013 (ENDO^{B}),[57] what somehow underlines the interest in key estimation in the EDM and DJing communities. We will come back to some of these resources shortly, as we have merged some them into the GiantSteps key dataset, described in the following paragraphs.

The GiantSteps Key Dataset

The GiantSteps project[58] helped us become aware, amongst other things, of the need of better tailored algorithms for applied MIR in music production environments. However, due to a more or less systematic lack of analytical ground truth in EDM, we embarked on the recollection of empirical data that we could use in our development process. This endeavour materialised in the creation of the two so-called GiantSteps datasets, comprising tempo and global-key annotations, respectively. Both data collections are already publicly available, and were initially described in Knees et al.

[52] http://djtechtools.com/2009/11/02/key-analysis-software-smackdown
[53] http://djtechtools.com/2012/01/26/key-detection-software-showdown-2012-edition
[54] http://djtechtools.com/2014/01/14/key-detection-software-comparison-2014-edition
[55] http://djtechtools.com/2015/11/16/key-detection-software-comparison-2015-edition
[56] http://blog.dubspot.com/dubspot-lab-report-mixed-in-key-vs-beatport
[57] http://blog.dubspot.com/endo-harmonic-mixing-key-detection-analysis
[58] http://www.giantsteps-project.eu

(2015), from where we extract most of the content for this block. However, in the reminder of this dissertation, we only consider the GiantSteps key dataset ($\textsc{gs}^{\textsc{k}}$).[59]

As mentioned elsewhere, it is common for a certain type of DJ to organise her collection with simple tonal information (i.e. with global keys). DJs often obtain their tracks online in music stores like *Traxsource*,[60] *Junodownload*[61] or *Beatport*,[62] which are designed to facilitate DJ's creative workflow by selling music labeled with genre, tempo and key information, as well as release dates, record labels or remix artists, besides other regular tags. Beatport is one of the most popular of such online services, providing two-minute previews for each entry in their database. Each item, typically each single track, is described in an individual web page, where related metadata in JSON format, together with the 96 KBPS MP3 audio preview, can be easily obtained from the source code, providing an interesting resource for audio and semantic MIR. Unfortunately, Beatport's key and tempo metadata are algorithmically determined, thus becoming useless for training and evaluation purposes. However,

> "until late 2014, Beatport allowed its customers to provide feedback on tempo and key information via a link on their website, pointing to a dedicated online forum. In this forum, users would post their corrections in free-form text using natural language. We performed a complete web crawl of this user forum in May 2014. At the time of the crawl, there were 2,412 comments available, of which 1,857 contained a direct link to a track on the Beatport website. From the link to the track, we downloaded the complete metadata record in JSON format using web scraping techniques. From this, we also extracted the associated style descriptor for statistical reasons." (Knees et al. (2015))

From all the posts containing a link to a specific track, we safely filtered those that could point to other popular key estimation algorithms such as *Mixed-In-Key*[63] or *Melodyne*,[64] searching for key labels in the remaining ones. After this process, we obtained a total of 404 key corrections, of which 15 were duplicates, and 1 track was no longer available, leaving us with a total of 388 tracks with key annotations. A detailed explanation of the process of information extraction from the Beatport forum can be found in Knees et al. (2015).

[59] As it will be explained in Chapter 5, in the course of our research we performed a revision of this dataset. That is the reason to denote this dataset with a superscript $^{\text{k}}$, standing for 'key' or 'Knees' (first author in the publication in which the dataset was made publicly available). In subsequent chapters we will introduce the revised GiantSteps+ dataset ($\textsc{gs}^{+}$), containing corrections to the original annotations as well as a greater detail of modal specification.

[60] http://www.traxsource.com

[61] http://www.junodownload.com

[62] http://www.beatport.com

[63] http://www.mixedinkey.com

[64] http://www.celemony.com/en/melodyne/what-is-melodyne

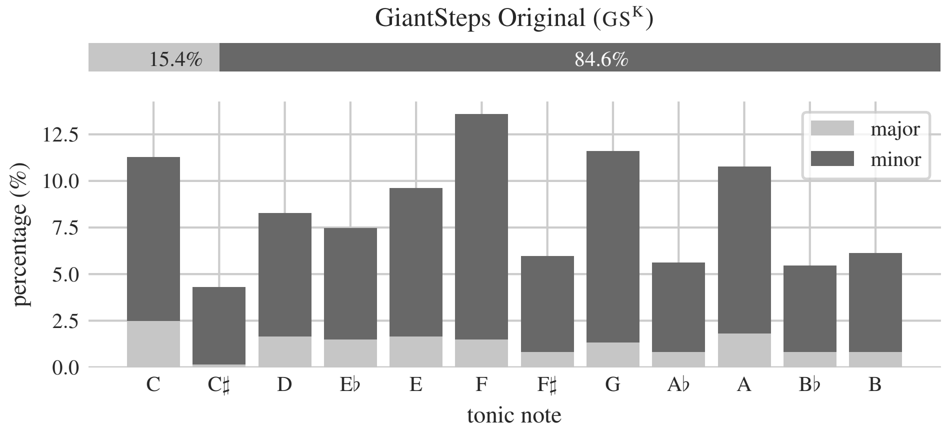


FIGURE 4.4: Distribution of major and minor keys in the original GiantSteps dataset.

In order to enlarge the collection, we decided to incorporate other scattered labels from EDM magazines and blogs. In particular, the analyses by Endo mentioned in previous paragraphs conveniently annotate various Beatport resources: ENDO^{A} consists of a list of 100 tracks provided as a GIF image file. This image contains 99 items (one of which is a duplicate) with artist name, song title, his key label, and the predictions of the Mixed-In-Key software and Beatport. We used OCR software to convert this list to a spreadsheet in order to obtain the human labels and access to the audio excerpts from the Beatport website. Using a simple script we retrieved the metadata of the candidate tracks from Beatport. When artist and title matched perfectly, the track and key label were assigned together, whereas cases with multiple candidates (e.g. with different remix versions), were assigned a correct label manually. This process allowed us to obtain 92 out of the unique 98 tracks in Endo's list. In his second report (ENDO^{B}), Endo makes a more exhaustive comparison between seven different key estimation applications. The new track list holds a total of 119 entries. 19 references direct to YouTube videos, while other seven tracks are listed without links or Beatport key tags. Excluding these 26 items, we were left with a batch of 93 additional songs with manual labels and direct links to the Beatport samples. As a last resource, we looked at the annotations used in the Djtechtools's 2014 showdown mentioned above, conducted on 60 tracks.

With all of these sources added together, we obtained a merged dataset with 633 labelled tracks, 29 of which were duplicates among the different sources. In these duplicate cases, the different sources agreed on the reported key, providing evidence of the reliability of our approach. In total, we gathered a global-key dataset of 604 two-minute EDM excerpts, as it is currently published.[65,66] A simple evaluation of the Beatport key labels, revealed that only 29.5% of the keys provided in the website are correct, according to our annotations.

Figure 4.4 shows the distribution of the tracks in the original GiantSteps dataset, arranged by tonal centre and modality, presenting a similar distribution compared to KFD, and with 84.6% of the items in minor modes.

4.1.4 Summary of Music Collections

Figure 4.5 shows a comparison of the key distributions in the three musical genres reported. The Western classical music dataset (Classical DB) presents a small prominence of major keys ($\approx$ 63%), which conforms to assumptions about euroclassical tonality (e.g Krumhansl, 1990, pp. 66–75). The three combined pop music collections, increase the bias towards major modes ($\approx$ 84%), with a concentration of tonal centres in pitches corresponding to the natural pentatonic scale. The modality distribution between pop and EDM is almost inverse, with only 15% of the total number of items in major. Here, the distribution across the twelve chroma is slightly more even, although the combination of the GS^{K} and KFD increases slightly the presence of natural tonics, just like in the other two genres.

To conclude this section on music collections, Table 4.1 shows a summary of the datasets discussed containing key information, together with their number of entries, musical genres covered, and quality of the available audio data. We can see that almost half of the datasets provide a single key estimate per audio item, whereas the other half provides structural key annotations. Regarding the vocabulary used, most datasets are annotated in a twofold major/minor modal vocabulary, although both BB and RS datasets are transcribed in richer ways, permitting to obtain additional modal information.

[65] The GS^{K} dataset is hosted in Github (https://github.com/GiantSteps/giantsteps-key-dataset).

[66] Johannes Kepler University provides a descriptive portal of the two GiantSteps datasets (tempo and key) plus some simple evaluation results (http://www.cp.jku.at/datasets/giantsteps).

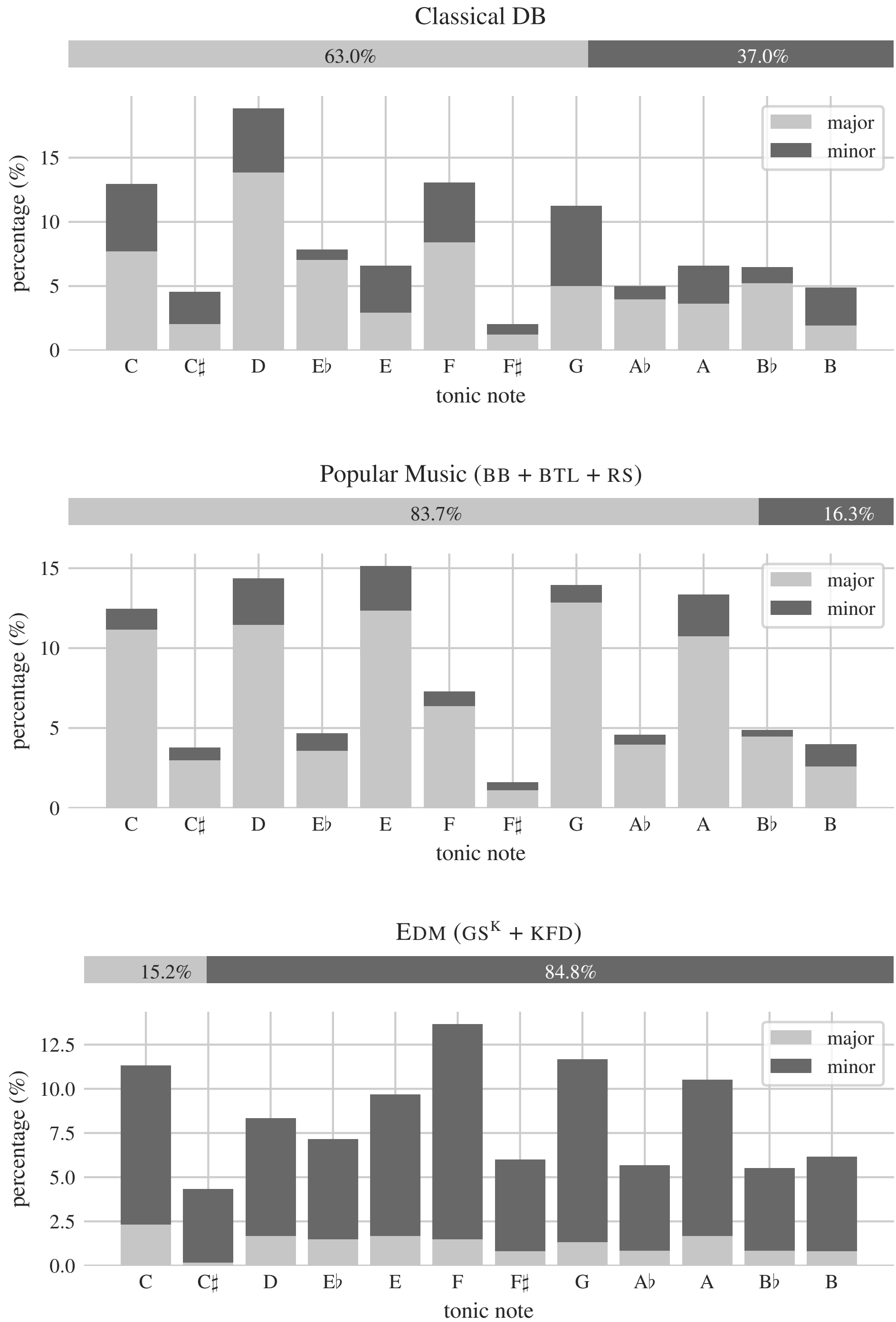


FIGURE 4.5: Joint distribution of keys in different musical genres. From top to bottom: Euroclassical (Classical DB), popular music (BB + BTL + RS) and EDM (KFD + GS^{K}).

					labels			
name	*abr.*	*style*	*tracks*	*vocabulary*	*global*	*changes*	*format*	*related publication*
Beatles	BTL	popular	180	*majmin*[a]		•	FLAC	Pollack (1999); Mauch et al. (2009a)
Billboard	BB	popular	625	*tonic + chords*		•	FLAC	Burgoyne et al. (2011)
ClassicalDB	CDB	classical	880	*majmin*	•		MP3@var	Gómez (2006a)
GiantSteps	GS[K]	EDM	600	*majmin*	•		MP3@96	Knees et al. (2015)
KeyFinder	KFD	EDM	998	*majmin*	•		MP3@var	Sha'ath (2011)
Isophonics	ISO	popular	223	*majmin*		•	FLAC	Mauch et al. (2009a)
RWilliams	RW	popular	65	*diatonic modes*		•	FLAC	Di Giorgi (2013)
RS200	RS	popular	200	*tonic + chords*		•	FLAC	Temperley & De Clercq (2013)
WTC	WTC	classical	96	*majmin*	•		FLAC	Noland & Sandler (2007)

[a] with occasional diatonic modes.

TABLE 4.1: List of publicly available datasets with key annotations, indicating the number of items, modal vocabulary, musical style and temporal scope of the key labels.

4.2 Evaluation Methods

In the computational study of certain musical styles, the tasks of tonic identification and mode recognition can be isolated as separate problems. This is a regular practise, for example, when approaching the computational study of some Non-Western musics, like Turkish Makam music (e.g. Karakurt et al., 2016) or several Indian traditions (e.g. Gulati et al., 2014). One of the reasons for this conceptual separation of the scale pattern and the underlying tonic can be found in the normally larger range of possible modes given a single tonic, and in the essentially monodic quality of many these traditions. However, in Western musics —and especially in euroclassical music, with only two basic modes— a tonal centre can hardly be seen in isolation with the modality it prescribes, for it normally is associated with a tonic chord (already suggesting a certain mode), and a set of relationships with other chords and neighbouring keys. This is probably the reason why most authors have proposed evaluation strategies that try to capture subjective aspects of tonal perception, like the close interplay between nearby keys. For example, C major is typically perceived as being closer to A minor (although it does not share neither tonic or mode) than to D major (sharing mode and only one second apart) or C minor (sharing the tonic note). Therefore, it is a common practise to report some of these 'acceptable' errors besides the ratio of correct tonic, mode and key. Gómez (2006a), for example, details the correct joint estimation of tonic and mode (key) as a measure of the accuracy of her system, but also provides further details regarding the percentages of correct modes, semitone errors (as potential tuning errors), as well as errors related by dominant, relative and parallel key relationships. Pauws (2004) provides similar details, although he adds together the just mentioned errors (plus a subdominant error) as a measure of the accuracy of the system.

4.2.1 The MIREX Scoring System

The Music Information Retrieval Evaluation eXchange (MIREX)[67] is an international initiative born to evaluate advances in music information retrieval among different research centres, by quantitatively comparing algorithm performance using test datasets that are not available beforehand to participants (Downie, 2008; Downie et al., 2010). Since 2005, it is celebrated on a yearly basis, as a special event taking place during the International Society for Music Information Retrieval Conference (ISMIR).

[67] http://www.music-ir.org/mirex/wiki/MIREX_HOME

Over the years, authors have detected flaws and problems in different MIREX evaluation tasks (Salamon & Urbano, 2012; Hu & Kando, 2012; Davies & Böck, 2014; Scholz et al., 2016), although there still seems to be room for discussion and revision of the evaluation strategies, given it is a community-driven initiative. Regarding the audio key-finding task, however, the test dataset and evaluation criteria would have remained the same since the first edition in 2005,[68] if it was not for the recent incorporation of the GS^{K} dataset (Knees et al., 2015), introduced previously in this chapter.[69]

The Mirex05 key dataset (the test collection that has been used in all MIREX editions so far) comprises of 1,252 euroclassical music pieces rendered from scores onto monoaural uncompressed audio files with a MIDI synthesiser.[70] The ground truth is taken from the title of the works, since, as explained above, it was a regular habit to name compositions according to formal and tonal descriptors (e.g. Mozart's *Symphony No. 40 in G minor* or Beethoven's *Piano Sonata No. 8 in C minor*).

Regarding the evaluation procedure, the submitted algorithms must provide a single label indicating the *tonic* and the *mode* of each audio file. Tonic notes can include any of the twelve chromas, whilst modality is limited to a binary output of *major* or *minor* only. However, since the perception of key is considered to be contextual and slightly subjective, the evaluation system imposes a ranking by which related keys, such as relative or parallel keys, or those by a distance of a perfect fifth, are weighted and summed together into a composite weighted score. The weighting of neighbouring keys is, to say the least, misleading, and the task's webpage does not make any further clarification regarding the 'perfect fifth' distance. As a matter of fact, looking into the computer code used in the MIREX evaluation,[71] reveals that the evaluation algorithm only regards as a 'positive' error the dominant-as-tonic mislabelling. This biased weighting seems to be a bug in the algorithm, which has been corrected for the 2017 edition to punctuate equally both ascending and descending fifth relationships.[72] The weighting values for the different errors are presented in Table 4.2.

[68] http://music-ir.org/mirex/wiki/2005:Audio_Key_Detection

[69] However, the use of the GS^{K} dataset is not reported on the official wiki website, where there is only reference to the 'Mirex05' dataset.

[70] According to their website, two different synthesisers were used for the first edition in 2005 ("Winamp synthesised audio and Timidity with Fusion soundfonts", in http://www.music-ir.org/mirex/wiki/2005:Audio_Key_Finding_Results), yielding slightly different results. Subsequent editions seem to have omitted the Winamp files, and they only provide error percentages for a single database synthesised using Timidity.

[71] https://github.com/ismir-mirex/nemadiy/blob/master/analytics/trunk/src/main/java/org/imirsel/nema/analytics/evaluation/key/KeyEvaluator.java

[72] This information was revealed in a personal communication with Johan Pauwels, the person responsible of the evaluation task, and reflected in the 2017 results webpage.

	error types				
	correct	*fifth**	*relative*	*parallel*	*other*
weights	1.0	0.5	0.3	0.2	0.0

TABLE 4.2: MIREX key-finding evaluation error-weighting system. Between 2005 and 2016, the fifth error only accounted for dominant errors. Subdominant errors were given a score of 0.

The participant algorithms are run over the initial 30 seconds of each audio track, discarding the rest of the audio information to prevent the interference of modulation processes in the estimation of the principal key. In our view, although a convenient solution when analysing euroclassical music, it remains questionable whether this prescription should apply to other musical genres such as pop music or EDM, since key changes are not characteristic of these types of music.[73]

Brief Discussion of the MIREX *Results*

Tables 4.3 and 4.4 show the evaluation results with the Mirex05 dataset for all the algorithms submitted to the competition since its origin until 2016.[74] Results are taken from the corresponding MIREX webpages, with the exception of the 2005 edition, where we have used the results for the Timidity database —instead of the two provided— for the sake of comparability.

What becomes evident at first sight, is the temporal gap between the first edition in 2005 and the second in 2010. From 2010, the evaluation has been run on a yearly basis, even though in 2014 and 2015 there is only a single, recurrent candidate. In order to show a more realistic adjustment with the number of novel submissions, we present resubmissions in a smaller font size. From this reduction, we can see clearly that the first edition in 2005, attracted the largest number of participants (6, one author sending two variants of the same algorithm), followed by 2010–2011 with 4 independent participants. Last, in 2012 and 2016 there where 3 independent contenders, a rate that seems to be preserved in the 2017 edition.

A look at the scores reveals that the results from the first competition have not being improved in subsequent editions. For example, if we set an arbitrary boundary at 0.85 MIREX points (in bold font), we find that five from the seven methods submitted in 2005 (which have been mentioned and discussed in varying degrees in Section 3.3)

[73] Again, according to an informal conversation with the task captain, the evaluation on the GS^{K} data was carried on the full two-minute excerpts, although this detail is not provided in the results webpage.

[74] Results for the 2017 competition are not published as of 10th Nov. 2017 (music-ir.org/mirex/wiki/2017:Audio_Key_Detection_Results). However, the 2017 competition seems to use additional datasets (ISO, RW and BB, summarised in Table 4.1).

	submission details			MIREX *weighted errors and overall score*					
year	*contributors*	*name*	*code*	*correct*	*fifth*	*relative*	*parallel*	*other*	*score*
2005	Chuan & Chew (2005a)			.7228	.0759	.0543	.0176	.1294	.7806
	Gómez (2005)	*start*		.8259	.0351	.0343	.0160	.0887	**.8569**
	Gómez (2005)	*global*		.8107	.0583	.0471	.0184	.0655	**.8577**
	Izmirli (2005a)			.8698	.0335	.0248	.0144	.0575	**.8969**
	Pauws (2005)			.8259	.0184	.0551	.0264	.0743	**.8569**
	Purwins & Blankertz (2005)			.8466	.0575	.0168	.0168	.0623	**.8838**
	Zhu (2005)			.7701	.0527	.0375	.0264	.1134	.8129
2010	Pauwels et al.	*ELIS/DSSP KeyChordExtractor*	PVM2	.6933	.1677	.0751	.0080	.0559	.8013
	Peeters (2010)	*ircam key detection*	GP8	.7500	.1046	.0367	.0319	.0767	.8197
	Rocher et al. (2010)	*Simbals key 1*	RRH1	.4864	.2173	.2157	.0072	.0735	.6612
	Rocher et al. (2010)	*Simbals key 2*	RRH2	.3530	.2843	.0248	.2029	.1350	.5432
	Ueda et al. (2010b)	*UUOS*	UUOS	.6534	.1502	.0879	.0136	.0950	.7575
2011	Bandera et al. (2011a)	*AKD2011-pdfs*	DTBS1	.7340	.1094	.0655	.0080	.0831	.8100
	Bandera et al. (2011b)	*AKD2011-pdfs-modes*	DTBS2	.7524	.0982	.0439	.0152	.0903	.8177
	Khadkevich & Omologo	*key1*	KO3	.5056	.2061	.0942	.0639	.1302	.6497
	Khadkevich & Omologo	*key2*	KO4	.4657	.2117	.0551	.0671	.2005	.6014
	Pauwels et al.	*ELIS/DSSP KeyChordExtractor*	PVM2	.7252	.1358	.0719	.0104	.0567	.8168
	Peeters (2011)	*ircamkeymode-1.2.1*	GP1	.7500	.1046	.0367	.0319	.0767	.8197
	Rocher et al.	*Key-Simbals*	RHR1	.2428	.0447	.4225	.0080	.2819	.3935
	Ueda et al. (2010b)	*Audio Key Detection*	UUOS2	.6534	.1502	.0879	.0136	.0950	.7575

TABLE 4.3: MIREX evaluation results on the Mirex05 dataset for key-finding algorithms submitted between 2005 and 2011. Resubmitted algorithms are shown in smaller font size *(source: http://www.music-ir.org)*.

	submission details			MIREX weighted errors and overall score					
year	*contributors*	*name*	*code*	*correct*	*fifth*	*relative*	*parallel*	*other*	*score*
2012	Jansson & Weyde (2012)	*Zweiklang Profiles*	JW1	.7372	.0799	.0559	.0343	.0927	.8008
	Pauwels et al. (2012)	*ircamkeychord-keyindependent*	PMP4	.6230	.1749	.1286	.0112	.0623	.7513
	Pauwels et al. (2012)	*ircamkeychord-keytheoretic*	PMP5	.6542	.1837	.0911	.0112	.0599	.7756
	Pauwels et al. (2012)	*ircamkeychord-keyclassical*	PMP6	.7348	.1134	.0615	.0128	.0775	.8125
	Peeters (2012)	*ircamkeymode-1.3.2*	GP5	.7500	.1046	.0367	.0319	.0767	.8197
	Tzanetakis (2012)	*MarsyAudioKey*	GT4	.2284	.2021	.0855	.1584	.4657	.3588
2013	Cannam et al. (2013)	*QM Key Detector*	CF3	.8267	.0599	.0272	.0176	.0687	**.8683**
	Pauwels & Peeters (2013)	*ircamkeychord-key*	PP5	.7348	.1134	.0615	.0128	.0775	.8125
	Peeters & Cornu (2013)	*ircamkeymode-1.3.2*	GP4	.7500	.1046	.0367	.0319	.0767	.8197
2014	Cannam et al. (2014)	*QM Key Detector*	CN1	.8267	.0599	.0272	.0176	.0687	.8683
2015	Cannam et al. (2015)	*QM Key Detector*	CN2	.8267	.0599	.0272	.0176	.0687	.8683
2016	Bernardes & Davies (2016)	*INESC Key Detection*	BD1	.7260	.1406	.0583	.0160	.0591	.8170
	Cannam et al. (2016)	*QM Key Detector*	CN1	.8267	.0599	.0272	.0176	.0687	.8683
	Faraldo et al. (2016b)	*fkey*	FJH2	.6342	.1829	.1102	.0128	.0599	.7613
	Faraldo et al. (2016b)	*fkey-edm*	FJH3	.6070	.2404	.0367	.0543	.0615	.7491

TABLE 4.4: MIREX evaluation results on the Mirex05 dataset for key-finding algorithms submitted between 2012 and 2016. Resubmitted algorithms are shown in smaller font size *(source: http://www.music-ir.org)*.

surpass that mark. On the contrary, this criterion is only met once in later editions, by the method proposed in Cannam et al. (2013), based on the work of Noland & Sandler (2007). The best performing algorithm on the 'mirex 05' dataset is the one proposed by Izmirli (2005a), based on templates elaborated from sampled piano notes, weighted with the flat profiles mentioned in Temperley (1999). It is followed closely by the algorithm by Purwins et al. (2000), based on constant-Q profiles correlated with probe tone profiles by Krumhansl & Kessler, although most submissions from 2005 yield very similar results. If we lower the boundary to .80 points, however, we can find at least two methods in each year's competition reaching or surpassing this value, implying at least 70% of correctly classified instances.

In any case, the MIREX evaluation results should only be read in the narrow musical context they represent: 30 second excerpts of euroclassical music, rendered to audio from musical scores. Therefore, they say little about how the submitted methods would perform on actual musical recordings, or in other musical styles, at different exposure times. In any case, at least from 2015, the submitted algorithms can be evaluated comparatively, on a dataset of two-minute real audio recordings, representing a body of electronic dance music, as shown in Table 4.5. At first glance, it is already noticeable that the performance decreases considerably compared to the results in Tables 4.3 and 4.4. It is remarkable the decrease of the QM Key Detector, from over 82.67% correctly classified instances to a bare 39.7%, practically halving its performance. We attribute this drop to at least two factors: first —and most importantly— the GS^{K} dataset contains real audio excerpts, as opposed to audio synthesised from MIDI scores. This introduces aspects beyond the scope of transcription, mostly of timbral or spectral nature, with clearly difficult the detection process. Second, this decrease in performance suggests that EDM represents an actual challenge to the key estimation task, perhaps indicating that the models and assumptions of tonality present in submitted algorithms do not reflect well the range of tonal practises in EDM. This is at least partially suggested by the two different variants by Faraldo et al. (2016b), whose only difference lays in the profiles used. *Fkey* uses the profiles by Temperley (1999), whereas *fkey-edm* uses tonality templates derived from a corpus of EDM. We will return to this algorithm in Chapter 6, since it is one of the methods developed in the course of this research.

	submission details			MIREX *weighted errors and overall score*					
year	*contributors*	*name*	*code*	*correct*	*fifth*	*relative*	*parallel*	*other*	*score*
2015	Cannam et al. (2015)	*QM Key Detector*	CN2	.3974	.0480	.1325	.0430	.3791	.4697
2016	Bernardes & Davies (2016)	*INESC Key Detection*	BD1	.5530	.0662	.0977	.0381	.2450	.6230
	Cannam et al. (2016)	*QM Key Detector*	CN1	.3974	.0480	.1325	.0430	.3791	.4697
	Faraldo et al. (2016b)	*fkey*	FJH2	.3411	.0712	.1689	.0960	.3228	.4465
	Faraldo et al. (2016b)	*fkey-edm*	FJH3	.6209	.0613	.0662	.0563	.1954	.6826

TABLE 4.5: MIREX evaluation results on the GS^{K} dataset for key-finding algorithms submitted in 2015 and 2016. Resubmitted algorithms are shown in smaller font size *(source: http://www.music-ir.org)*.

4.2.2 Other Methodological Concerns

Evaluation Metrics

For each dataset and algorithm under analysis, we normally report the percentage of correctly estimated tonics and accurately estimated modes independently, together with the correct key estimation. Additionally, we describe typically 'acceptable' errors (neighbours, relatives, parallels), and provide a MIREX weighted score according to the weights in Table 4.2. It is important to note that although diverging from previous MIREX results and other popular MIR evaluation toolboxes such as 'mir_eval'[75] (Raffel et al., 2014), we consider fifth errors (neighbour keys) to include both ascending and descending intervals, provided that both tonics share the same mode. Therefore, the four possible mislabels in the 'fifth' category include 'I as V', 'I as IV', 'i as v', and i as iv'. We think this is a more neutral way of assessing this error, which, in our view, was biased towards euroclassical music in previous MIREX competitions. For example, dominant relationships (I-V-I) summarise the main directional force in euroclassical music, with almost every composition in major modality containing a modulation to the dominant region. In contrast, rock modality rarely presents this structure, superseded by predominantly subdominant relationships (I-IV-I) (e.g. Temperley & De Clercq, 2013).

Audio Quality

Up to this point, we have not explicitly discussed the quality of the audio files in the above mentioned datasets. In computational research, audio resources are necessarily digitised and stored in a computer or online server. Datasets with a high audio quality are normally transcoded from CD's in uncompressed PCM formats. However, uncompressed data takes a considerably larger memory space than other compressed formats. The FLAC file format (standing for 'free lossless audio codec'), is a compression audio format that provides the same quality as original uncompressed data at lower memory consumption, but not all decoders are FLAC-friendly. In the reality of music consumption, with increasing online music purchases and streaming services, the actual standard are so-called *lossy* formats, which reduce the amount of data by compressing or cutting frequency bands typically without much musical information, and in which the human hearing apparatus is perceptually weaker. An MP3 file at 320 KBPS is considered to be a good quality audio file, despite being encoded in a lossy format. For example, the Spotify streaming service distributes music in *Ogg Vorbis* format, an open-source alternative to MP3 encoding, at 96, 160 or 320 KBPS, depending on whether the streaming happens on a mobile device, a desktop computer

[75] http://craffel.github.io/mir_eval

or with a premium account.[76] While audio bit rates seem to vary substantially across resources, a sampling rate of 44,100 Hz seems to be the standard quality for most audio resources, from CD rips to lower quality compressions.

With this amount of variability, a good key estimation algorithm should expect to receive all sorts of data formats and qualities, especially if the algorithm is developed for a practical scenario and/or with a creative orientation. Fortunately, the datasets we have at hand reflect well the variety of formats and compression levels found in real world scenarios, as shown in Table 4.1. For example, the KFD includes lossy formats at various bitrates, as a side-effect of our gathering of tracks from various sources; most popular music datasets have been transcoded into FLAC directly from compact discs; the GS^{K} collection has been downloaded from online preview clips as low-quality MP3 files at 96 KBPS.

A potential problem of 'perceptual' codecs like MP3, is that they filter out high-frequency content, what could be detrimental in analysing specific musical genres, such as EDM, characterised by its high percussive content and saturated electronic timbre. According to Urbano et al. (2014), who evaluated the robustness of chroma features under various codecs and bitrates in a variety of musical genres, chroma features are very robust to encoding differences, even with bitrates as low as 64 KBPS. However, they advise to normalise the chroma vector in order to minimise the effects of lossy audio codecs, and observe that best results are usually achieved when training data is the same encoding format as the expected analysis data. In the last part of this chapter, we present a preliminary evaluation of the effect of audio degrading in the key recognition task.

Track Length

As discussed above, the MIREX evaluation has been typically carried on the initial 30 seconds of MIDI renders of euroclassical music scores. This follows an extended practice of performing key estimation in fragments of short duration at the beginning or end of a piece of music (before a "departure from" or after a "return to" the main key) (Pauws, 2004; Izmirli, 2005b; Peeters, 2006a; Gómez, 2006b). One of the motivations of observing the beginning of a piece of music is to avoid falling into modulations that can obstruct the global-key estimation task. However, modulation is not characteristic of EDM neither of pop music. Furthermore, as it will be shown shortly, our experiments suggest that computational key finding generally provides better results when analysing full-length tracks, something already noted by Pauws (2004), who discusses the classification accuracy for different analysis windows (from 2.5 seconds to entire pieces) at different time positions in the music signal.

[76] https://support.spotify.com/us/article/What-bitrate-does-Spotify-use-for-streaming

4.3 Evaluation of Available Resources

As mentioned elsewhere, the extended practise of harmonic mixing among DJs and producers, together with an increasing demand for automatic labelling and classification of ever growing music collections, are probably the main factors behind the proliferation of digital tools for key estimation in recent days. Some of these tools originate in academic research and are made available as part of audio analysis environments such as *Essentia*, an open-source library for audio analysis and description (Bogdanov et al., 2013) which includes a variant of the method by Gómez (2006a), described in Section 3.3. Similarly, the approach by Noland & Sandler (2007) is wrapped as a vamp plugin (the QM *Key Detector*) to be used within the analysis software *Sonic Visualiser*. These methods are normally regarded as general solutions, or targeted at euroclassical music at best.

On the contrary, commercially available methods are typically tailored to popular music and, with the exception of *Beatunes*[77] —a music player that incorporates analytical methods to create enhanced playlists— are mostly aimed at the production and mixing of EDM. Some of these solutions are offered as standalone applications with key analysis as their only —or main— purpose. It is the case of *Mixed-in-Key*[78] and *KeyFinder*,[79] a freely available piece of software by Sha'ath (2011). Additionally, key estimation methods are normally integrated into all-purpose DJing tools, as Native Instrument's *Traktor*,[80] Pioneer's *Rekordbox*,[81] *Serato*[82] or *Virtual DJ*.[83]

4.3.1 Competing Algorithms

In the following paragraphs we describe briefly the peculiarities of some of these solutions, summarised in Table 4.6, before proceeding with the evaluation per se. In particular, our evaluation will compare the QM Key Detector and Essentia's Key Extractor (which implements the method by Gómez (2006b)) to KeyFinder, Mixed-In-Key —the preferred choice among EDM producers— and Traktor, which is regarded as the quality standard in djing software. Bear in mind that commercial applications are black boxes, and so we can not learn much about their inner workings. However, we consider this comparison to be a valid indicator of the state-of-the-art when it comes to applied MIR in real life scenarios.

[77] http://www.beatunes.com
[78] http://www.mixedinkey.com
[79] http://www.ibrahimshaath.co.uk/keyfinder
[80] http://www.native-instruments.com/en/products/traktor
[81] http://www.rekordbox.com
[82] http://serato.com
[83] http://www.virtualdj.com

		key scope		
name	*abr.*	*changes*	*global*	*related publication*
Essentia Key Extractor	ES		•	Gómez (2006a)
QM Key Detector	QM	•		Cannam et al. (2016)
KeyFinder 2.3	KFA		•	Sha'ath (2011)
Mixed-In-Key 8	MIK		•	
Traktor 2.11	TK		•	

TABLE 4.6: key estimation algorithms used in this preliminary evaluation. We also show their analysis scope and indicate related publications where applicable.

Essentia's Key Extractor

Essentia[84] is a C++ framework with python bindings for audio signal processing and music information research developed at the Music Technology Group in Pompeu Fabra University (Bogdanov et al., 2013). It provides an ever growing collection of analysis and processing methods that users —namely programers— can combine and adjust according to their needs. Besides, Essentia comes with a number of default 'extractors', that is, predefined combinations of instructions to perform typical analytical tasks, aimed at less proficient users. Since most parts of the methods introduced in Chapter 6 are developed in Essentia, in this preliminary evaluation we include the output of *Essentia*'s key extractor, which is based in the method by Gómez (2006b) described in Section 3.3. For the most parameters, we have used the extractor's default settings, listed in Table 4.7. However, in order to provide a fairer comparison, we have increased the analysis window size from the 4,096 points by default to 16,384 points.

QM *Key Detector*

The QM *Key Detector* (QM) is based on the work by Noland & Sandler (2007) and available as a *Vamp* plugin written by Cannam et al. (2016) for *Sonic Visualiser*,[85] a user-friendly program that can perform a wide range of sonic analyses, aimed at researchers and musicologists. The QM Key Detector vamp plugin, uses key profiles derived from analysis of J. S. Bach's *The Well-Tempered Clavier I* (1722), with default window- and hop-sizes of 32,768 points, providing a key estimate every 10 frames. QM's output vocabulary is limited to a major/minor classification, plus an 'unknown' label, when the algorithm can not detect a specific key. We have processed all the files

[84] http://essentia.upf.edu
[85] http://www.sonicvisualiser.org

parameter	*value*
window size	16,384 pt.
hop size	2,048 pt.
window type	blackman & harris
minimum frequency	40 Hz
maximum frequency	5,000 Hz
maximum number of peaks	10,000
split frequency bands	✗
non-linear spectral transformation	✗
chromagram size	36 bins
chroma weighting type	squared cosine
chroma weighting size	1.333 st.
key profile	temperley (Fig. 3.5)
similarity	cross-correlation

TABLE 4.7: Essentia's key extractor configuration parameters.

using the *sonic-annotator* software,[86] with a script kindly provided by Chris Cannam, that reduces the multiple estimations to a single one by choosing the most prevalent as the global estimation, as submitted to the MIREX competition.[87]

KeyFinder 2.3

KeyFinder (KFA) is a free piece of software for OSX, whose only functionality is to analyse the global key of any imported audio file, writing the resulting estimation onto the audio file as an ID3 tag or in the filename. The method in previous versions of the software (1.26) was described in Sha'ath (2011), and allowed the user to manipulate some analysis parameters, such as the window size or overlapping factor, as well as to selecting different key profiles, even customised ones. Unfortunately, the current version (2.3) lacks any user configuration, becoming a completely hidden system.

Mixed In Key 8

Mixed-In-Key (MIK) is probably the most popular software when it comes to key detection for harmonic mixing, due to its acknowledged high performance. This piece of software analyses sound files in search of keys, tempo and cue points, writing

[86] http://vamp-plugins.org/sonic-annotator

[87] https://code.soundsoftware.ac.uk/projects/mirex2013/repository/show/audio_key_detection/qm-keydetector

the estimated metadata into the files or exporting it as CSV files. Mixed-In-Key occasionally reports multiple keys for a single track, however the exported annotation for each track is still one single label, separated with a slash (e.g. 'A / Am', 'G / Gm / Dm'). In such cases, the first label is always the one that takes the lengthier segment of the analysed audio. Additionally, some files are labelled as 'All'. After inspection, we found that 'All' labels typically refer to fragments with spoken voice or highly percussive segments, in any case with sparse pitch content, resulting highly neutral for harmonic mixing purposes (i.e. without a key).

Traktor 2.11

Traktor (TK) is Native Instruments' DJ and mixing software, possibly among the preferred solutions by professionals and amateurs alike. Intimately working with their own series of dedicated controllers, Traktor's visual metaphor reminds of a solid DJ mixer, allowing the user to perform typical mixing operations. It offers tempo and key analysis per file, storing the analysis results onto an NML file, Native Instruments' own XML dialect. Generally speaking, all the mixing solutions available (rekordbox, Serato, Virtual DJ, etc.) are very similar regarding their functionality and graphic user interface, with options to mix with one, two or four desks.

4.3.2 Evaluation Results

The remainder of this chapter presents a comparison of the algorithms just mentioned, in order to prepare the ground for the discussion of our own contributions in Chapter 6. We start with a preliminary validation of some of our methodological assumptions, namely, the preferred evaluation on full-length excerpts and the robustness to various audio formats and qualities. Although we are mainly interested in measuring their performance in EDM, we present additional results for popular and euroclassical music too. Except where noted otherwise, all the methods are tested with their default settings and the latests software versions as per October 2017. All the algorithms under consideration are only capable of a binary modal output. Therefore, these evaluations are carried considering a binary major/minor classifier, with a single key label per item, since only QM explicitly labels on a segment basis.

Audio Quality

Table 4.8 illustrates the effect of quality downgrading in three datasets from different musical styles (WTC, BB, KFD) an three different algorithms (ES, QM, MIK). We simply present the MIREX score —a proper evaluation follows in the next section—

		MIREX *scores*	
set	*method*	FLAC	MP3@96
WTC	ES	.9021	.9021
	QM	.8281	.8458
	MIK	.8698	.8552
BB	ES	.7304	.7296
	QM	.6565	.6550
	MIK	.7788	.7795
KFD	ES	.4551	.4543
	QM	.4557	.4519
	MIK	.7658	.7667

TABLE 4.8: Effect of audio quality degradation in various key-finding algorithms on datasets from different genres: euroclassical (WTC), popular (BB) and EDM (KFD). The methods tested are Essentia (ES), the QM Key Detector (QM) and Mixed-In-Key 8 (MIK).

on the original data and downgraded to MP3 files at 96 KBPS. The downgraded audio quality is chosen according to the format of the GS^{K} dataset. Our main intention is to confirm the observations by Urbano et al. (2014), validating the usage of lower quality data for research purposes. As can be seen in the table, the effect of downgrading is residual in all instances, and the difference is never greater than 0.018 points (QM on WTC, in which the algorithm performs slightly better on the downgraded sample). Mixed-In-Key presents a difference of 0.015 points in the WTC, but all other tests show minimal differences, below 0.004 points. We take this to justify that throughout this dissertation, we perform all evaluations using the original format of each dataset.

Track Length

In Table 4.9, we show the effect of selecting a shorter analysis period from the beginning of each audio file. Like in the previous measurement, we tested the same three methods on different musical styles, providing the MIREX weighted score for four different durations, 7.5, 15, 30 and 60 seconds, plus the score for the complete audio track. What is true in all scenarios is, despite the different algorithms and test collections, that all methods provide their best results when analysing the complete audio duration (with the only exception of MIK, that reaches an equivalent result when taking only 7.5 seconds of WTC). Besides, this experiment suggests that the musical genre has an influence in the results, depending on the analysis window. Similarly, the particular characteristics of each algorithm seem affect the performance at the various durations.

		MIREX *weighted scores*				
set	*method*	*7.5 s.*	*15 s.*	*30 s.*	*60 s.*	*all*
WTC	ES	.8385	.8656	.8438	.8188	**.9020**
	QM	.4917	.6052	.6198	.7333	**.8281**
	MIK	**.8572**	.8062	.7656	.7385	.8552
BB	ES	.5661	.6301	.6866	.7305	**.7304**
	QM	.4096	.4649	.5347	.5936	**.6565**
	MIK	.6498	.7102	.7533	.7675	**.7784**
KFD	ES	.3566	.3873	.3986	.4197	**.4551**
	QM	.2885	.3156	.3614	.4222	**.4557**
	MIK	.5422	.5995	.6610	.7253	**.7658**

TABLE 4.9: Effect of analysing the first n audio seconds with various key-finding algorithms on datasets from different genres (euroclassical, popular and EDM). The methods tested are Essentia (ES), the QM Key Detector (QM) and Mixed-In-Key (MIK).

The QM Key Detector originally provides an estimation every 10 analysis windows (of 32,768 points, roughly every 2.22 seconds), taking the longest averaged fragment as the global estimate. This is probably the reason why it presents the greater variability between the different durations, and that the improvement increment seems correlated with the number of estimated segments for all three datasets.

Similarly, ES and MIK present this incremental behaviour in pop and EDM. Regarding euroclassical music, both algorithms drop performance in the intermediate stages, regaining accuracy at the entire duration. We attribute this behaviour to the modulatory nature of the WTC corpus, which seems to introduce difficulties in the intermediate parts of each track, but also to peculiarities of the key-finding process. In the case of ES, the accumulative nature of the algorithm (which averages all the chromas together) probably makes the system favour other keys instead of the main key at the intermediate levels, regaining confidence with the reappearance of the main key toward the end of each piece. It is plausible to infer that MIK also participates of a certain accumulative procedure, although we can not be certain. One more observation can be drawn from Table 4.9. As already noted, MIK provides a 'no-key' estimation ('all') when it detects un-pitched excerpts. In very short excerpts (7.5 secs.), MIK is able to determine the key of WTC pieces as good as with the complete audio —actually slightly better— probably because the opening of each work is very clear key-wise. However, in BB and KFD, the scoring for the 7.5 fragments is considerably lower, compared to the entire track. In order to study this, Table 4.10 shows the percentage of 'no-key' estimations for the three datasets. As expected, WTC does not present any of such estimations, being essentially pitch-only music. However, 'no-key' labels appear as we look into popular music, and they become

set	method	% of 'no-key' estimations 7.5 s.	15 s.	30 s.	60 s.	all
WTC	QM	0.00	0.00	0.00	0.00	0.00
	MIK	0.00	0.00	0.00	0.00	0.00
BB	QM	0.64	0.16	0.16	0.32	0.16
	MIK	**2.08**	**1.12**	0.48	0.32	0.00
KFD	QM	0.70	0.60	0.70	0.40	0.30
	MIK	**6.80**	**5.70**	4.30	2.30	0.20

TABLE 4.10: Percentage of items with 'unknown' (QM) and 'no-key' (MIK) estimations produced by the different analysis durations and datasets.

significant in EDM, especially for MIK, reaching up to 6.8% on 7.5 second fragments. This is a possible explanation for the poorer performance on these styles using short time window (since for the moment, all the annotations are labelled with a key). Besides, this might tell us something about the nature of some EDM —and to a lesser extent, rock— tracks, probably starting with un-pitched materials, such as spoken voices, special effects or simply introductory drum patterns.

General Evaluation

We would like to close this chapter with a preliminary evaluation of the five algorithms described at the beginning of this section. Figure 4.6 shows the evaluation results for the different algorithms calculated independently for three different styles, including the WTC dataset, a combination of the popular music collections (BB, BTL and RS), and the merged KeyFinder (KFD) and GiantSteps (GS^{K}) datasets, containing mostly EDM. The highest scores concentrate in euroclassical music, where ES obtained the maximum rank, followed by MIK and TK. We observe a decremental drop in ES and QM as we progress to other styles, suggesting an increasing complexity in popular music and, especially, in EDM. It is important to note, however that this 'complexity' is surely not tonal, but most likely of a spectral nature: whilst WTC is keyboard-only music, pop-rock instrumentation typically includes guitars, percussion and vocals. Electronic dance music, on the other hand, is mostly created with electronic musical instruments, opening doors to sonically vaguer areas. From Figure 4.6, it can also be inferred that popular music presents less variability regarding the percentage of correctly estimated keys with the different algorithms, whereas EDM seems to pose a real challenge to methods such as ES and QM, initially tailored for euroclassical music using the key profiles by Temperley (1999) and Noland & Sandler (2007), as shown in Figures 3.5 and 3.13.

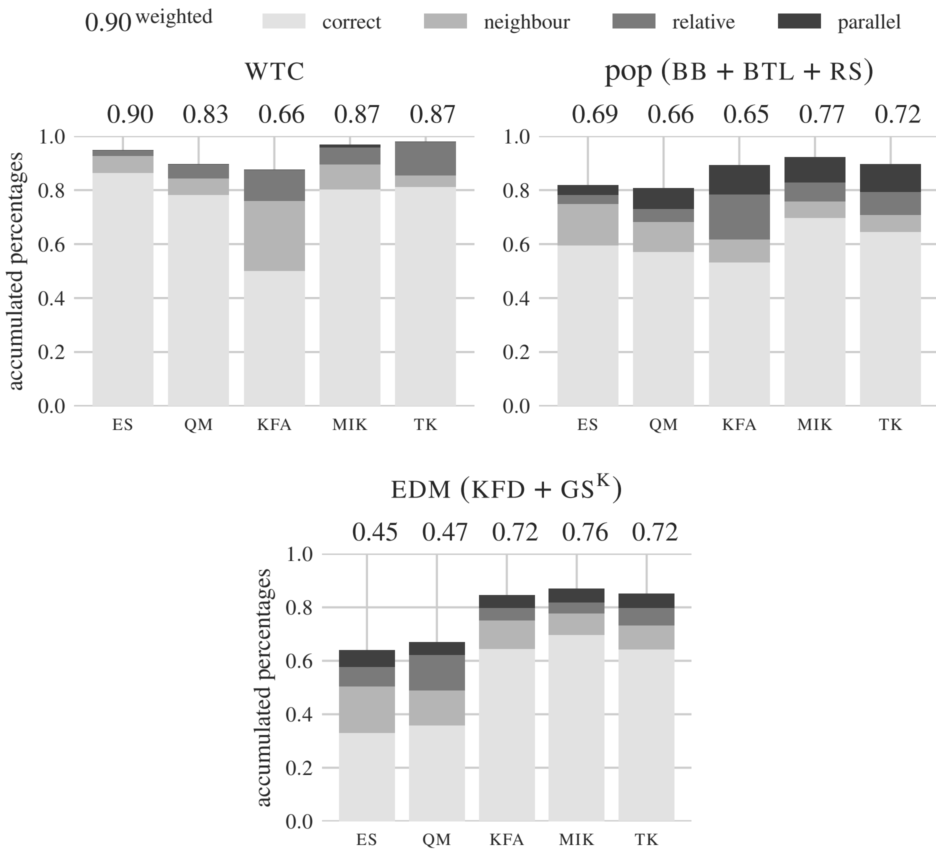


FIGURE 4.6: MIREX evaluation results of state of the art algorithms per genre. The bars show the accumulated percentages of commonly accepted error types (neighbours, relative and parallel), besides the percentage of correctly classified items. On top of each bar, we present the MIREX weighted score according to the weights in Table 4.2.

Additionally, Tables 4.11 and 4.12 provide more detailed numbers regarding the popular music and EDM datasets, respectively. In popular music, the less accurate combination appears to be KFA on The Beatles's music, with only less than half of the items correctly classified ($\approx$ 47%), in contrast with MIK, which provides the highest results for the 'rockier' datasets BB and RS ($\approx$ 71%). Algorithms seem to punctuate their best on the Billboard collection, something that is more noticeable by looking at the MIREX weighted scores. We attribute this to the fact that BB is a 'cleaned up' dataset, with ambiguous and modulating tracks removed from the test collection. The only exception to this pattern is the QM key detector, which provides better results on The Beatles dataset.

		correct items			*typical errors*				
set	*method*	*tonic*	*mode*	key	*fifth*	*relative*	*parallel*	*other*	score
	ES	.6704	.8576	*.6416*	.1488	.0288	.0288	.1520	*.7304*
	QM	.6448	.7952	*.5696*	.1120	.0528	.0752	.1904	*.6564*
BB	KFA	.6480	.6416	*.5488*	.0832	.1760	.0992	.0928	*.6630*
	MIK	.7936	.7840	**.7072**	.0656	.0704	.0864	.0704	**.7784**
	TK	.7680	.7456	*.6688*	.0640	.0832	.0992	.0848	*.7456*
	ES	.5500	.8611	*.5278*	.1056	.0444	.0222	.3000	*.5983*
	QM	.6778	.8000	*.6111*	.1111	.0333	.0667	.1778	*.6900*
BTL	KFA	.5722	.6167	*.4667*	.1111	.1889	.1056	.1278	*.6000*
	MIK	.7500	.7500	**.6389**	.0722	.0833	.1111	.0944	**.7222**
	TK	.6833	.6667	*.5611*	.0667	.1111	.1222	.1389	*.6522*
	ES	.5771	.8060	*.5075*	.2090	.0348	.0696	.1791	*.6363*
	QM	.6318	.7612	*.5373*	.1095	.0398	.0945	.2189	*.6229*
RS	KFA	.6716	.6517	*.5323*	.0746	.1095	.1393	.1443	*.6303*
	MIK	.8060	.7960	**.7064**	.0448	.0547	.0995	.0945	**.7651**
	TK	.7413	.7413	*.6418*	.0597	.0597	.0995	.1393	*.7094*

TABLE 4.11: Evaluation of state of the art algorithms with popular music datasets (BB, BTL and RS).

		correct items			*typical errors*				
set	*method*	*tonic*	*mode*	key	*fifth*	*relative*	*parallel*	*other*	score
	ES	.3878	.6463	*.3377*	.1794	.0591	.0501	.3737	*.4551*
	QM	.3767	.5792	*.3387*	.1413	.1293	.0381	.3527	*.4557*
KFD	KFA	.7084	.8737	*.6663*	.1062	.0341	.0421	.1513	*.7381*
	MIK	.7575	.8677	**.7054**	.0802	.0331	.0521	.1293	**.7658**
	TK	.7064	.8206	*.6543*	.0912	.0541	.0521	.1483	*.7265*
	ES	.4000	.6467	*.3183*	.1667	.0933	.0817	.3400	*.4460*
	QM	.4517	.5967	*.3900*	.1150	.1417	.6167	.2917	*.5023*
GS^{K}	KFA	.6617	.8167	*.6050*	.1117	.0667	.0567	.1600	*.6921*
	MIK	.7283	.8433	**.6800**	.0767	.0583	.0483	.1367	**.7455**
	TK	.6800	.7917	*.6250*	.0850	.0833	.0550	.1557	*.7035*

TABLE 4.12: Evaluation of state of the art algorithms with EDM datasets (KFD and GS^{K}).

In contrast, the results in EDM (Table 4.12) present more variability across the different algorithms. Both ES and QM (methods designed for euroclassical music) achieve the lowest scores for the two test datasets, as low as 32% correctly classified keys (ES on GS^{K}). On the upper side, Mixed-In-Key is again the algorithm with the highest scores (70% for KFD and 68% for GS^{K}), followed by Traktor and KeyFinder, whose performance should be taken carefully, since the 67% accuracy on KFD likely comes as an overfitting effect, given that the algorithm has presumably been trained with this dataset. To us, it came as a surprise that neither MIK, nor TK—supposedly professional solutions aimed specifically at EDM— managed to supersede the results obtained with the popular music datasets. Someone could argue that this could be caused by particularly challenging or unusually difficult datasets, in comparison to the average difficulty of EDM. However, we are more inclined to think that the actual challenges for key estimation in EDM lay in the wide range of timbral configurations, the omnipresence of percussive elements, and perhaps, an under-consideration of tonal configurations diverse from other musical genres. In any case, the ratio of $\approx 70\%$ accurately classified entries (for rock and EDM) leaves considerable room for improvement in popular music styles, and EDM in particular.

***

In this chapter, we have reviewed existing musical collections with key annotations. We have tried to address different musical styles —broadly labelled as euroclassical, popular and electronic dance music— in order to highlight differences between them, aligning with the idiosyncratic tonal practises of each style as discussed in Chapter 2. Furthermore, we have provided simple statistics for each genre, regarding the distribution of tonal centres and modality, introducing the original GiantSteps key dataset (GS^{K}), a collection of 604 EDM excerpts which constitutes the first contribution stemming from our research. Additionally, we have discussed typical evaluation metrics, proposing a minor variation regarding the evaluation of neighbouring keys, as well as defending computational key estimation using full audio excerpts, offering experimental support to our claims. We have also conducted a preliminary evaluation of key estimation algorithms, on datasets from various musical backgrounds.

State of the art methods present a perceptible variability across musical genres, typically achieving the best results on euroclassical music, despite the likely presence of modulation. We attribute the lower performance on popular music and EDM mainly to the timbral complexity present in these musics, but also to possible misconceptions of the tonal practises characteristic of these styles. This fact has already been pointed by (Temperley & De Clercq, 2013) regarding popular music. However, to our knowledge, there is no research illuminating whether this is the case in electronic dance music. Within this scenario, in the next chapter, we continue our narration with a statistical study of tonal practises in EDM, originating as an effort to gather additional data for our experiments. With such study we intend to shed some light about tonal configurations in EDM —including the use of reduced pitch-class sets, tonal ambiguity, as well as bimodal and atonical passages— that might be behind the generally impoverished performance of key estimation methods in this metagenre.

Chapter 5

A Study of Tonal Practises in EDM

"We were never musicians,
we're just collage artists"
Future Sound of London

In the previous chapter, we introduced the GiantSteps key dataset (GS^K), an effort to gather ground-truth from online resources that could be used to evaluate existing key estimation algorithms, as well as to develop and test our own. The main advantage of the GS^K dataset was its semi-automatic annotation procedure, relying on information from users' fora and contrasted labels across different sources, minimising considerably our labelling efforts.[88] However, such methodological approach imposed a few limitations too, namely, that the number of available audio fragments, the distribution between keys and genres, as well as the degree of annotation detail and its labelling confidence, were laid out of our operative power.

In order to address some of these restrictions, during our research, we initiated two additional recollective and analytical endeavours, which are described in detail in the current chapter. The Beatport dataset (BP), described in Section 5.2.1, was born as an attempt to enlarge the amount of available data, with a balancing criterion regarding the distribution of modalities and genres. Moreover, the GiantSteps+ dataset (GS^+) discussed in Section 5.2.2 provides additional pitch-class annotations, key changes, as well as other morphological descriptions of a sub-collection of 500 fragments from the original GiantSteps key dataset, motivating a revision of the initial key annotations. In both annotating efforts, we used an experimental labelling system that is less restrictive than the typical binary tagging, even though it maintains a certain degree of 'reducibility' to a binary vocabulary, allowing comparison with

[88] See Knees et al. (2015) or Section 4.1.3 for details.

other methods. Although this annotation strategy was developed in the course of our analyses, we discuss it in this chapter prior to the description of the datasets, in the following section.

As a natural extension of our analytical labour, in Section 5.3 we attempt a generalisation of the most characteristic tonal traces found throughout our study, with references to specific tracks from one of the collections described. The advantage of discussing examples from these datasets is double. On the one hand, the data is publicly and directly available, facilitating a deeper understanding —and criticism— of the descriptions herein contained.[89] In addition, the same audio collections have been subject to computational analysis during the development and assessment of our own key-finding methods, establishing an interesting dialogue between music-theoretical and engineering inquisitions.

5.1 A Lax Annotation Strategy for EDM Datasets

Throughout this dissertation we have made repeated references to the likely unsuitability of a binary modal vocabulary for almost any type of non-euroclassical music. This has been substantiated in Chapter 2, where we described the various modal systems operating in popular music, and the possibility of open-ended tonal practises in EDM. Furthermore, the results of our preliminary evaluation in popular and electronic dance musics seem to support this claim, given the impoverished performance in these styles when assessed under a binary classifier (see Section 4.3.2). Questions about the suitability of euroclassical binary tags have been raised in previous publications (e.g. Temperley, 2001; Gómez, 2006a) and addressed specifically by Temperley & De Clercq (2013), who annotate their rock music collection as tonic-only keys without any modal tag. However, Temperley & De Clercq provide chord annotations that could be parsed in various ways to obtain detailed modal implications. A similar strategy is used in the Billboard dataset (Burgoyne et al., 2011), and it is essentially inherent to any corpus containing chord-sequence annotations instead of single key labels. However, as suggested in Sections 1.2.1 and 2.4, chord sequences are characteristic only of certain EDM subgenres, mostly those under the influence of disco and pop, in flavours ranging from epic euroclassical progressions in trance music to highly sophisticated jazz sequences in deep house. Other styles, such as techno and its variants or breakbeat-driven subgenres, can make use of just a single chord, or no chords at all. Therefore, a strategy based on chord labels seemed a priori inappropriate to

[89] Amongst the materials accompanying this thesis, we provide scripts to download the audio directly from Beatport,[90] as detailed in Appendix C.

characterise EDM as a whole, and, on the other extreme, a single tonic-note annotation felt too sparse an indicator —although certainly the most appropriate regarding tracks presenting just a single tone— to characterise a complete EDM track or a hypermetrical loop. Besides, embarking in new analytical endeavours made us consider the utility of more detailed annotations, accounting for deviant modal practises, potentially ambiguous tracks or fragments without pitch or key at all, given the pitch scarceness found in some subgenres. In order to meet these requirements, we designed an open-ended annotation framework that could easily allow us to reduce detailed modal tags to a simpler binary modal dictionary —granting a comparison with other methods— providing analytical information useful beyond purely computational approaches.

Figure 5.1 shows the modal tree used to develop our annotation methodology, created after observation of the most common traits in the two corpora analysed, that will be described in subsequent sections. From left to right, the figure represents a progressive increment of detail regarding pitch information —except for the no-key branch. As we have insinuated in Section 2.2.1, a first differentiation between pitched and unpitched tracks could result useful when dealing with some subgenres (level 1 in the figure). Moreover, even when pitched elements are present in a piece of music, it should not be taken for granted that pitches are going to convey a sense of tonicality (level 2), although it certainly is the most likely scenario. Level 3 in Figure 5.1 presents the four principal labels in our proposed vocabulary, accounting for excerpts conveying or not major or minor modalities, without defining further modal peculiarities. These four tags should be regarded as the most basic forms of modal expression, as represented by a major or minor third interval over the tonic, by the absence of it, or by the presence of other indicators such as diminished triad, which might still produce a clear sense of tonicality. The fourth level in the figure represents a finer modal grain, providing differentiation between the diatonic modes and other types of scale. It is perhaps worth noting that the tag ‘other’ can be further divided into locrian (the diminished diatonic mode) but also single-tone configurations and, in general, tracks without a clear major or minor bias (i.e. with absence of tonic major or minor triad indicators).

Our lax annotation strategy uses the labels in Figure 5.1, keeping the same tree structure and vocabulary. It typically consists of a single row with one, two or three columns —depending on the annotation detail and type— separated by either a space, a tab or a comma. The first column contains the tonic pitch class. We established a convention to annotate atonal or atonical tracks with either an ‘`X`’ or an hyphen ‘`–`’, including an additional ‘`unknown`’ original to allow the annotator to isolate cumbersome cases in which any estimation would be little more than random. In order to be able to signal tracks deviating from standard tuning reference, we may

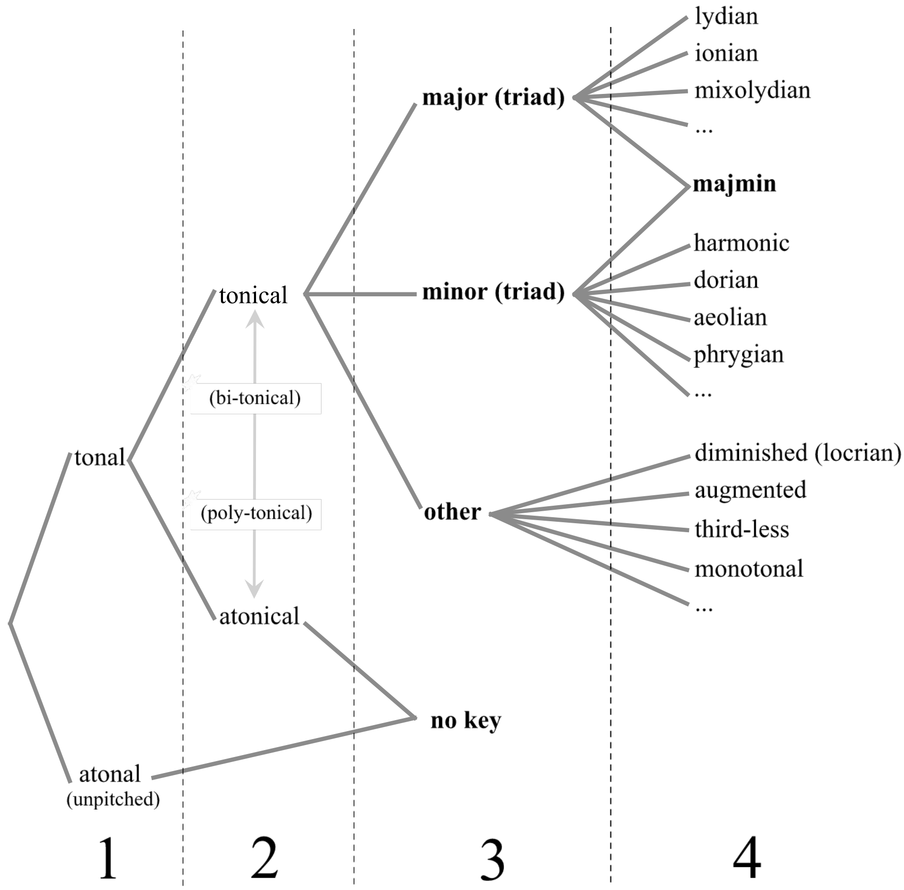


FIGURE 5.1: Modal arborescence in the Beatport dataset, from basic characterisation to complete modal specification.

append two special characters, a caret ('^') or an underscore ('_') to indicate a raised or lowered tuning from the pitch reference, respectively. In this way —although we typically discard this information in evaluation procedures— we can easily recognise 'detuned' fragments, which should be considered carefully in applied scenarios such as harmonic mixing.

After the tonic pitch-class annotation in the first column, the basic mode indicator takes the second field, consisting in one of three possible labels ('major', 'minor', or 'other'). The 'no-key' tag is redundant in the second field, since it is implied with the X label in the first column). At every next level, further annotation detail can be added, whether specifying a particular mode, a monotonic excerpt or an atonal track. Although there is no particular analytical motivation to limit the modal dictionary,

col. #1 'tonic'	*col. #2* 'basic mode'	*col. #3* 'detailed mode'
X		
X	atonical	
C	major	
C♯^	other	
G_	minor	phrygian
D♭	minor	harmonic
B♭	other	monotonic
unknown		

TABLE 5.1: Examples of annotation labels for our corpora of EDM. The first column indicates the tonic note (or a lack of it). It is normally followed by a basic modal indicator and, optionally, by a more detailed modal descriptor.

for programmatic reasons we have limited it to the following descriptors: aeolian, dorian, harmonic, ionian, locrian, major-♭6, mixolydian, pentamaj, pentamin, phrygian and phrygian-major.[91] Figure 5.1 presents a few annotation examples according to this simple procedure.

This annotation procedure, however, does not yet provide specific means to deal with modal ambiguity or tonal ambivalence, which in our view, constitute important aspects of tonality in EDM, as we will shortly explain. At first, we thought about including an additional 'ambiguous' label to help identifying tracks with open or multiple interpretations. However, considering ambiguity as a label on its own would typically hinter the actual particularities of tonally ambiguous tracks, by assigning them to the same placeholder. Therefore, we decided to indicate ambiguous tracks by annotating the main tonal forces involved in each track's ambiguity, separated by the special 'or' character '|' (e.g "A minor | C major", expressing bimodal ambivalence between relative keys, or "F minor | F major" indicating modal ambiguity over the same tonic F). With this operation —allowing multiple annotations for a single fragment— we gained some assets both in analytical and computational domains. First, we do not see any practical, computational or theoretical advantage in providing an interpretive disambiguation for intendedly ambiguous passages. Even more, we think that acknowledging the factors of ambiguity itself can shed more light over this expressive phenomenon, somehow neglected when attempting to disambiguate it, perhaps too subjectively. Moreover, this annotation openness will have implications in our evaluation methodology, as we can consider one or multiple annotations together, depending on the evaluation objective.

[91]Computer code to analyse, parse and evaluate audio files according to this convention is referenced in Appendix C.

As noted above, this annotation framework crystallised while studying the annotations and analyses described in the following blocks. As such, it has been used to convert and unify the annotations and comments left by our collaborators into a unified labelling system, computer-readable and humanly comprehensible.

5.2 New Music Collections

5.2.1 The Beatport Dataset

As observed in the opening paragraphs of this chapter, the initial GS^{K} dataset presented some methodological limitations, especially regarding the distribution of items across genres and modalities, with 85% of its total entries labelled as minor. Although this ratio seems to adjust to real music distributions, as it is also suggested by the KeyFinder dataset (KFD) and previous experiments by Gómez (2006a, p. 131), we wanted to obtain a more even distribution of major and minor modalities in order to derive various tonality profiles from real music. We decided that Beatport was a good candidate upon which develop a new annotating strategy based on three related facts: (a) our previous experience collecting the GS^{K} data had shown us how easy was to obtain two-minute audio excerpts from virtually the whole Beatport database, adding to thousands of tracks (b) with valuable metadata regarding artist names, titles and remixers. Furthermore, the Beatport's catalogue seemed updated, (c) containing audio tracks in circulation at the time of the download (Fall 2015), and representing popular EDM subgenres at that moment. Therefore, we embarked in a gathering process of 1,486 additional audio tracks with accompanying metadata, including artists, title, label, key, BPM, and remix version, which where subsequently annotated by an external collaborator, and subsequently revised by the author of this document.[92]

The approach to balance the new collection was based in Beatport's own genre and key tags. With a python script, we downloaded random entries mostly labelled as 'major' by Beatport, while keeping a balanced distribution across genres and tonic notes. While we gave credit to the genre labelling, in the annotation process we disregarded Beatport's key labels, substituting them with manual annotations. Despite our efforts to balance the new collection, we were left with only 28% of the total tracks in major, 63% in minor, and a smaller group of excerpts labelled either with an hyphen (–), or other ambivalent annotations, normally consisting of multiple tags.

[92] The credit for the annotation process goes to Eduard Mas Marín.

As an additional asset of this new batch of annotations, we asked our external collaborator to assess his ‘labelling confidence’ for each track, what could be seen as an indicator of the variability in tonal complexity in EDM per se, and for which we predefined a three-level scale, expressed in the following terms:

(2) *Confident.* The annotator thinks that other expert analysts would likely provide the same label.

(1) *Ambivalent.* Although confident about his labelling, the annotator acknowledges that the track is highly subjective and could be interpreted differently.

(0) *Insecure.* The annotator has difficulties to make a decision about the track. However, in the annotation process of this collection, we asked our collaborator to produce a label in any case, using the ‘unknown’ keyword when it was not possible to make a clearer judgement.

Furthermore, we encouraged our annotator to write down his impressions and/or additional descriptions regarding any particular track. Some of such additional descriptions express doubts about the annotation decision (“tonic centre is clear, but difficult to establish a particular modality”) or detail a modality beyond the traditional major and minor labels (“phrygian”, “mixolydian”). A few other comments indicate some degree of modal ambiguity (“major and minor modes coexist in the track”, “50% G major - 50% E minor feel”) or confirm key changes and their approximate timing (“the key changes in minute 1:20”).

By looking at the comments and confidence levels simultaneously, we managed to write down additional —and occasional— modal information (for example, the presence of phrygian and mixolydian). However, more interestingly, we were able to isolate challenging tracks, attempting to provide explanations regarding their difficulty. Across the whole dataset, we found four main types of divergence with regard to the binary major/minor classification, which can be broadly grouped into (a) *bimodally* ambivalent tracks (with more than one tonic candidate), (b) fragments with major/minor modal ambiguity (*majmin*), (c) excerpts not suggesting any particular key (*no-key*), and (d) entries conveying a clear tonal centre without a specific major or minor modality (*other*). Table 5.2 presents the distribution of the items in the dataset according to these broad descriptors, grouped by their confidence levels.

Additionally, the distribution of the dataset items by their tonic notes is shown in Figure 5.2, where it can be seen that despite the efforts to obtain a modally balanced collection, over 60% of the total still represents minor tracks. On the other hand, although there are visible peaks on C and F, the collection across multiple tonics

	manual annotation				
label	*confident*	*ambivalent*	*insecure*	SUM	*Beatport*
major	366	33	3	**402**	(1,447)
minor	783	108	12	**903**	(39)
majmin	7	33	0	**40**	
bimodal	21	27	1	**49**	
other	5	6	6	**17**	
no-key	0	1	72	**73**	
unknown	0	0	2	**2**	
SUM	1,182	208	96	**1,486**	(1,486)

TABLE 5.2: Distribution of the Beatport dataset across confidences and additional modal labels. The initial Beatport modal distribution is also shown.

is relatively balanced. The 'no-key' entry is indicated as a new bar, concentrating around 5% of the total number of tracks. The ambiguous label represents mainly major/minor ambiguity, and items with a bi-tonical quality are not included in the figure, to preserve the correct ratio of items.

In the preparation of this manuscript, we rewrote the web scraping code almost from scratch, in order to facilitate the acquisition of audio directly from the Beatport website, together with all the available metadata, including but not limited to artists, title, remix version, label and key.During the revision process, we realised that Beatport refurbished its website in 2016, as a consequence of structural changes in the company. One noticeable difference is a slightly new taxonomy of the available genres. Tracks previously labeled as 'pop rock' have completely disappeared, whereas other labels, such as 'chill out' and 'electronica', have gone into a new group of 'electronica / downtempo'. Besides, there are new labels such as 'big room' (including a portion of tracks previously labelled as 'progressive house') or 'trap / future house'. Another potentially interesting change was the incorporation of Zplane's *tonart* algorithm for the automatic key tagging.[93] However, our second download in July 2017 confirmed that the key labels remained identical for the available tracks. Unfortunately, in these second download, we found that from the 1,486 original tracks in the BP dataset, 103 were already lacking the artist info page with the associated metadata, although all the audio snippets were still accessible using the original track's id number.

The new download operation redistributed our audio samples across genres. The effect of this redistribution is noticeable in Figure 5.3, where most genres are represented with an average of 94 tracks, except for the merged tracks in 'electronica /

[93] According to Zplane's website (http://licensing.zplane.de/technology#tonart).

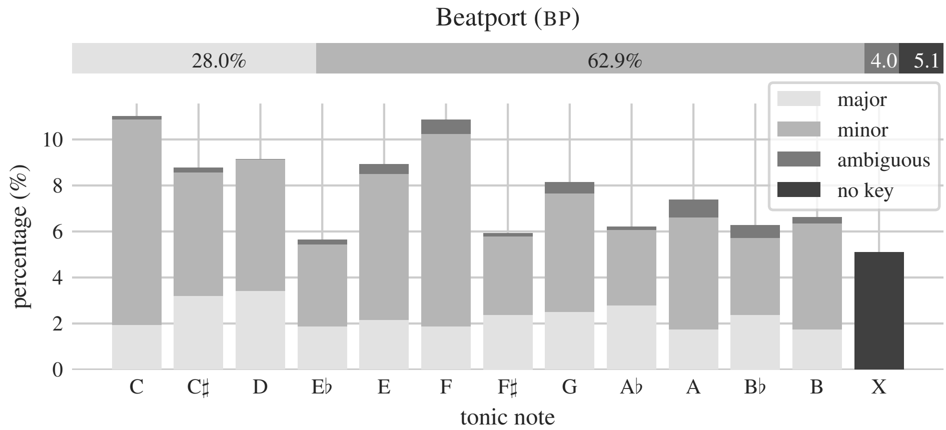


FIGURE 5.2: Distribution of modalities across tonics in the Beatport dataset. Bi-tonical excerpts are not shown in the plot to preserve the ratio of items along the chroma axis.

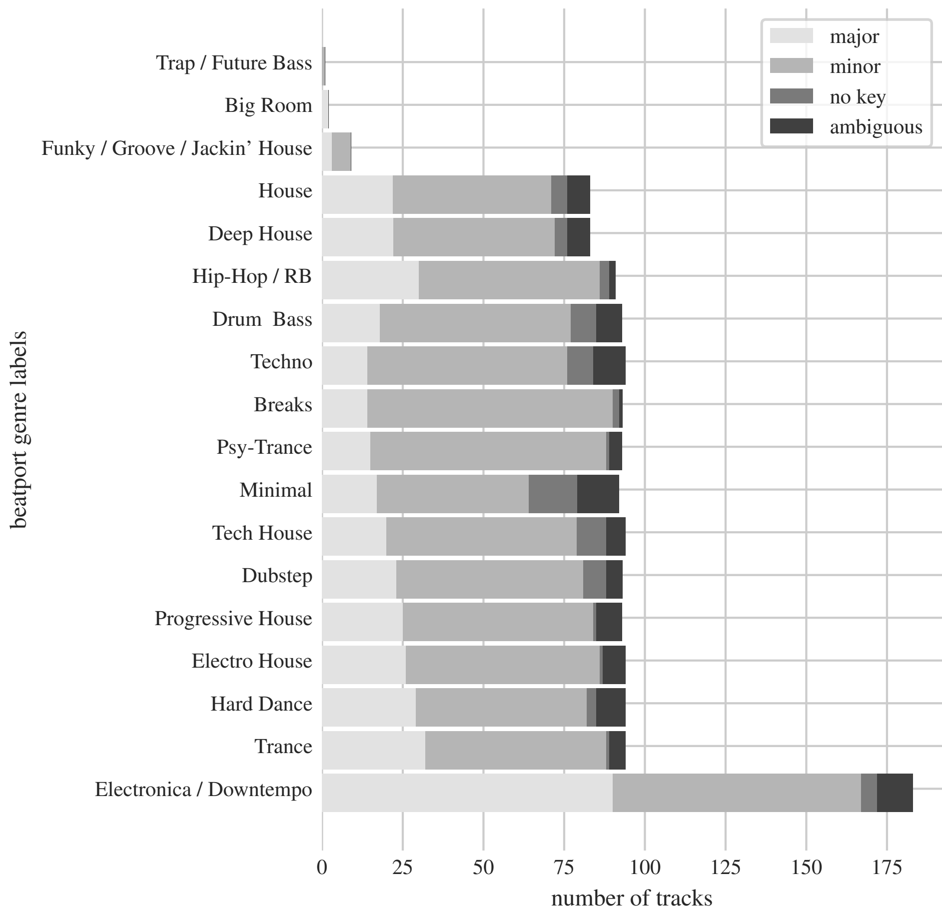


FIGURE 5.3: Distribution of tracks by genre and mode in the Beatport dataset. The 'ambiguous' label represents tonical tracks not included within the major or minor modes.

downtempo', doubling the items compared to all other subgenres. This new label seems to contain items that could be considered as borderline-EDM, many of them closer to other popular music styles. This might be indicated by the prominence of major tracks, which is clearly less common in all other subgenres. Last, the three entries at the top of Figure 5.3 represent new styles, stemming from containers such as 'house' and 'hip hop'.

5.2.2 The GiantSteps+ Dataset

As advanced in Section 4.1.3, in the process of elaborating this manuscript, we performed a thorough revision of the complete GS^{K} collection, adding new modal and pitch-class tags, and making a few changes in the data as it is published, including the correction of 63 global-key labels. We also removed three exact audio duplicates under different entries[94,95] and discarded two additional tracks that are clearly non-EDM styles.[96] To compensate for these missing items, we incorporated a new entry,[97] adding to a total of 600 tracks. Just like with the BP dataset, we prepared new computer code to download the audio files directly from Beatport. Compared to our initial download in May 2014, our latest access to the site in July 2017 revealed that 64 tracks from the original GS^{K} dataset were no longer available, although all audio clips remain accessible. Therefore, in the new data collection, 540 entries contain the complete metadata offered by Beatport, 42 items have partial metadata (kept in the ID3 tags of the files) and 18 tracks are kept with the genre label only.[98].

If with the Beatport collection we wanted to obtain a larger and balanced collection across genres and keys, the new GiantSteps+ dataset (GS^{+}) represents an effort to analyse, with more degree of detail, the audio tracks already present in the original GS^{K} collection. In particular, we were interested in obtaining additional modal information, beyond the binary labelling, which in turn, could serve to contrast the previously extracted labels with an expert's judgement. For this endeavour, we chose a subsample of 500 tracks, based on two criteria. First, we filtered out the 60 items for which we did not have comprehensive metadata, as explained in the previous paragraph. From the 540 remaining tracks, we discarded genres containing less than 10 items, what, in turn, removed a few tracks falling out of the scope of what is

[94] Beatport assigns a unique number identifying each single track in their database. In the remainder of this document, we sometimes make reference to Beatport tracks just by referencing their unique identifier.

[95] The duplicate tracks are 4320199 = 1922470, 5085261 = 2666332 and 5740146 = 1986370.

[96] 3284384 and 5015793.

[97] 140603.

[98] ID3 tags is the standard metadata format for MP3 files. Common fields as track, artist or album names, are normally embedded into the actual sound file using this standard (http://id3.org)

typically regarded as EDM, including genres such as “reggae / dancehall / dub”, “funk / soul / disco”, “hip hop / R&B”, and “DJ tools”, leaving us with 500 tracks distributed across 14 different EDM genres.

In the revision process, we implemented a different analysis strategy, writing down the tonic pitch and the detailed pitch-class set independently for each track (e.g. {C:047↊}).[99] If a given pitch structure corresponded to a well-known scale pattern, this additional label could be added in a separate field (for example, {A:0357↊} is typically referred to as A minor pentatonic). However, this operation was optional, since we wanted to be able to extract modal labels in a later stage, directly from the pitch-class annotations with a computer program. In case key changes occurred, these had to be written down too, together with a time mark. Besides, just as with the Beatport dataset, the annotator was asked to write down his comments and impressions on individual tracks, as well as to measure his degree of labelling confidence, according to the threefold-scale described in the previous block.

One of the goals of this annotation strategy, based on pitch-class set description, was to be able to parse the results in different ways, allowing us to study the data without being overly conclusive beforehand. In this way, we could study the number of tracks containing a tonic minor triad, or a leading tone degree ($\natural\hat{7}$), independently from pre-established modal labels. Additionally, we could infer the major/minor ambiguity by measuring the presence of both thirds on each track. As a means to create the basic modal descriptors for this new collection (major, minor, majmin, other or no-key), we looked for simple tonal indicators in the annotated pitch-class sets. Our parsing methodology is summarised in the following steps:

1. We converted the annotations from absolute format (e.g {C:c,d♭,e♭}) into separate tonic and ordered pc-set fields (C and {013}). This stage allowed us to study the variability of scales and pitch cardinality (i.e. number of pitch classes) in the corpus. Atonical passages were annotated with an ‘X’ character instead of the tonic, followed by the pc-set in normal order (see Section 2.1.6).

2. We looked for characteristic subsets in the pc-sets, in order to group the annotations into the five broad modal categories defined (major, minor, majmin, other and no-key). This step was achieved by looking for the major ({047}) or minor triad ({037}) sets, or a combination of both ({0347}, majmin). In principle, we assumed as ‘no-key’ all tracks without an annotated tonic, and gave the provisional label ‘other’ to the remaining pc-sets containing a tonal centre without tonic major or minor triads (e.g. tracks with single tones or pc-sets with diminished chords).

[99]The credit for the raw analyses of these 500 tracks goes to Daniel G. Camhi.

3. Additionally, we looked for more specific modal definitions, by combining the previous measures with other tonal indicators, such as leading tones or phrygian lowered seconds, appending new labels to the basic modal descriptor, according to the vocabulary defined in Section 5.1. These new tags were not mutually exclusive, so a single entry could contain a number of these if the pitch-class set matched different parsing rules.

4. In a later stage, we checked manually the labels and assessed the comments left at the time of the analysis, eventually assigning a pre-final key label to each track based on these.

5. Last, we compared the obtained annotations with the original GS^{K} tags, in order to look for potential inconsistencies between them. We knew beforehand that annotations containing 'other', and 'no-key' labels would create a conflict with previous labels, so we carried personally another listening test on these ambiguous tracks, making a final decision based on our personal criterion (whether to keep the new labelling or maintain the previous binary tag). After scrutiny of these difficult tracks, we checked all other divergences between the two annotation batches, correcting the respective annotations in one or other set after aural inspection of the conflicting tracks. In this new annotation batch, we took an ambiguity-friendly approach, using comments, estimations and listening assessments to annotate possibly ambiguous tracks as explained before (e.g. "C minor | C major"). However, the corrections to the GS^{K} data contain one single annotation, in order to preserve the original labelling format.

As advanced above, with these operations we managed to correct 63 labels in the original GS^{K} dataset, most of them consisting in relative, parallel or neighbouring errors. Regarding the new provisional tags, we re-labelled 106 tracks either with corrections or additional labels, according to our analysis. The basic modal ratio of the revised GS^{+} dataset is summarised in Table 5.3, arranged by confidence level. The distribution of items according to their genre labels and modalities is shown in Figure 5.4, where it can be seen how this collection is clearly biased toward house genres. Despite this tendency, it is interesting to note that ambiguous tracks span through all the represented subgenres, what might suggest that modal ambiguity could indeed be seen as characteristic of the meta-genre as a whole. Additionally, Figure 5.5 shows the distribution of items across tonic notes. After our revision, the dataset was left with only three atonical tracks, although it is worth noting that the presence of purely major tracks ($\approx 10\%$) is comparable to the share of modally ambiguous entries, what likely reflects more faithfully the complexity of tonal practises in EDM, compared to the original GS^{K} key distribution.

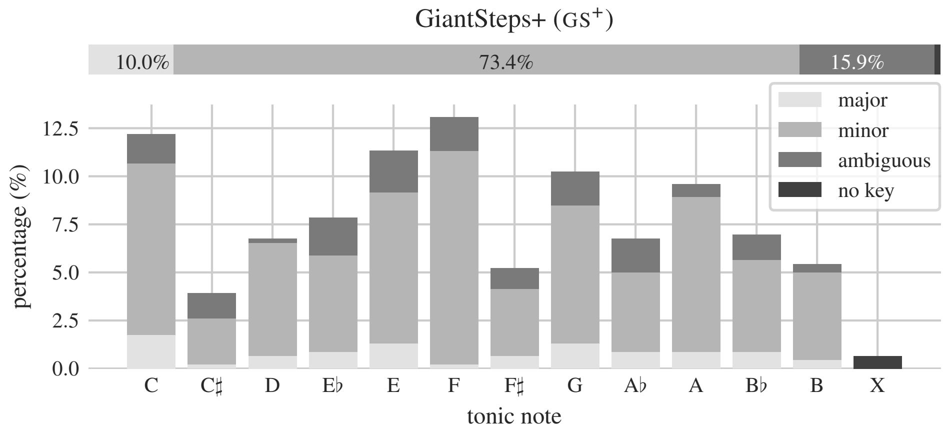


FIGURE 5.4: Distribution of tracks by genre and mode in the GS+ dataset. The 'ambiguous' label represents tonical tracks not included within the major or minor modes.

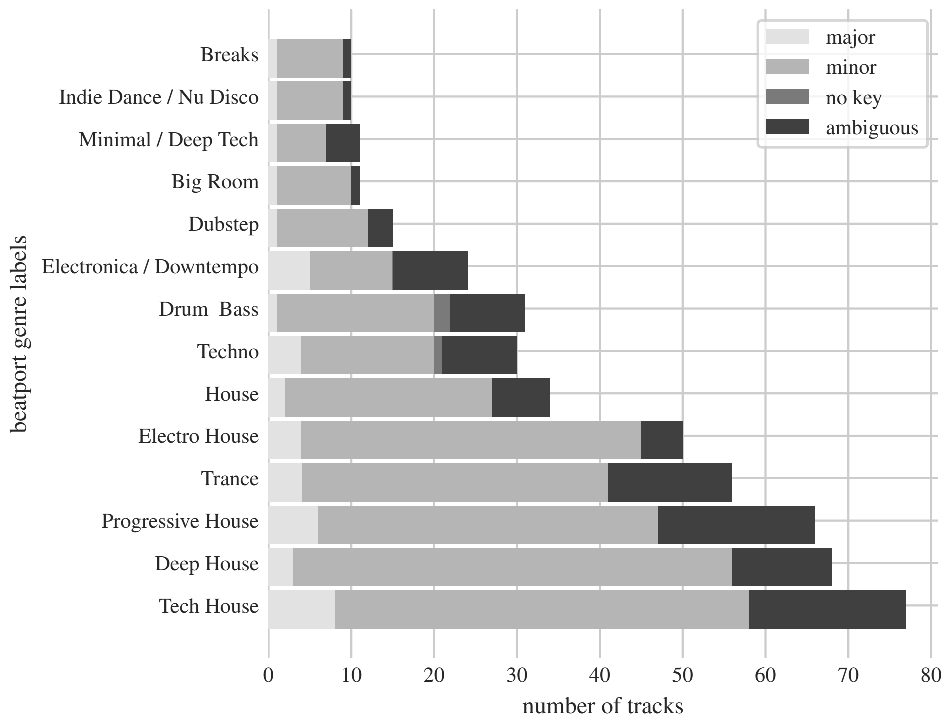


FIGURE 5.5: Distribution of modalities across tonics in the GS+ dataset. Bi-tonical excerpts are not included in the graphs to preserve the ratio of items. Ambiguous tracks refer to tracks with modal ambiguity.

	manual annotation				
label	*confident*	*ambivalent*	*insecure*	SUM	*Beatport*
major	40	4	2	**46**	(245)
minor	293	41	3	**337**	(245)
majmin	37	7	3	**47**	
bimodal	38	3	0	**41**	
other	18	7	1	**26**	
no-key	2	1	0	**3**	(10)
unknown	0	0	0	**0**	
SUM	428	63	9	500	(500)

TABLE 5.3: Distribution of tracks by confidence in the GS+ dataset.

As an additional experiment, during the revision and analysis process of the GiantSteps+ data, we asked our expert to fill, for each track, a simple checklist containing 17 potentially characteristic identifiers of various EDM subgenres. These included tonal indicators such as pedal tones, chord sequences or riffs, but also other textural marks, such as the presence of lead melodies, vocals, glissandi, incidental effects or spoken voices. The global results of this simple questionnaire are shown in Figure 5.6. All the tracks in the dataset contain pitch, and almost all, drums. While the second trace is certainly characteristic of EDM, the invariable presence of pitch is surely due to the origin of the data, mostly coming from the Beatport user forum to correct key labels and other online key estimation evaluation pools. Over 40% of all the entries have a pitch-only section, likely belonging to a break or build-up structure. In contrast, only a few tracks have drum breaks. 60% of the fragments present a lead melody, which in half of the instances is apparently vocal ($\approx 30\%$). Typical tonal indicators, such as chord sequences, riffs, pedal tones or arpeggios are individually situated under the 40%, with an overall presence comparable to other 'difficult' tonal effects, such as glissandi and incidental recordings. Key changes occur in around 5% of the tracks.

The lower part of Figure 5.6 arranges the same information broken down by subgenre. All rows are normalised to show potential characteristics across various subgenres. The figure suggests that —although timidly— some of these textural and tonal descriptors might help in differentiating subgenres, whilst others seem to characterise EDM as a meta-label. For example, drum breaks or spoken voices do not seem characteristics of the GS+ collection. Similarly, key changes are rare in the whole corpus. Classic tonal indicators, such as chord sequences and riffs are present in trance and house variants, whereas techno and minimal seem to favour static tonal structures,

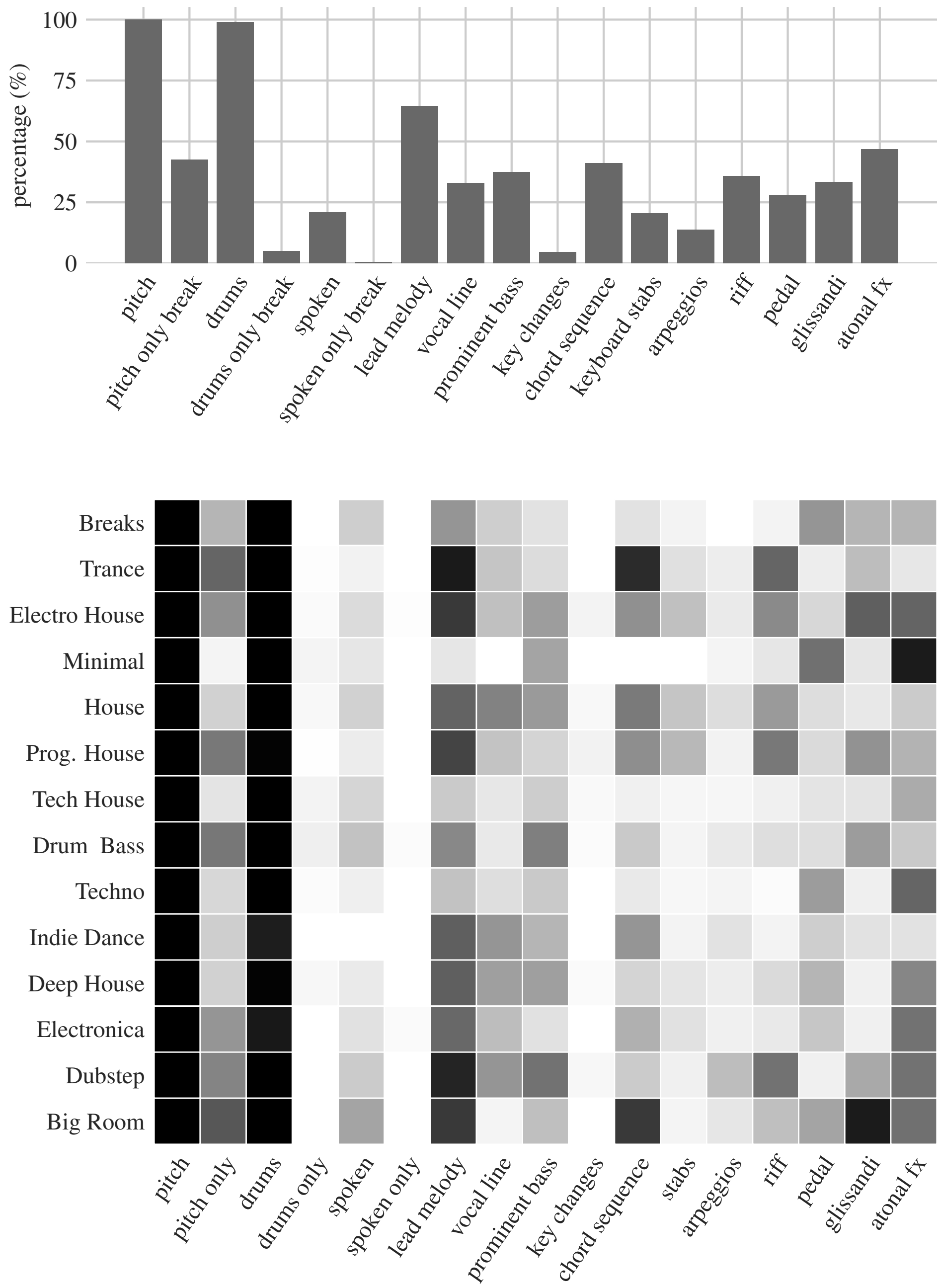


FIGURE 5.6: Tonal and textural features in the GiantSteps+ dataset. The matrix (below) represents the features normalised per subgenre.

such as pedal tones. Dubstep also shows some preference for arpeggios —which do not seem all that characteristic of other subgenres— and prominent basslines, a common feature in drum‘n’bass, deep house and minimal. Regarding other tonal effects, glissandi seem idiomatic of big room and electro house. Other atonal effects clearly define the sonic world of minimal, electro house and techno, although they moderately appear across many other styles.

5.3 Generalising Tonal Practises in EDM

In this section, we elaborate on the analysis of the two datasets described, in an attempt to distil some recurrent tonal aspects across various EDM subgenres. Since our analyses have focused above all on key characterisation, what we discuss in this section could be regarded as ‘timeless' observations —still images of fragments of music— defined by particular scales and vertical configurations, rather than by their timely succession.

As already advanced in the Introduction, our strategy bears resemblance with what Tagg has referred to as the ‘extended present', roughly corresponding to the duration of a hypermetrical loop (Tagg, 2012, pp. 272–273). However, if we think of the loop as an endless repetition or variation of the same motif, the extended present, could somehow be regarded as an ‘intended infinity', not in terms of texture or timbre —which are the principal drivers of musical flow in EDM— but in terms of a constant tonal ground.

Therefore, we exclude from our digression larger temporal scopes, such as the sequential arrangement of tracks or the DJ-set as a complete musical structure, where a proper sense of key evolution could emerge in the succession of different tracks, according or not to the regular customs of harmonic mixing. Our study of short-term configurations is justified by two main reasons. First, the hypermetrical loop arrangement of EDM tracks appear as the optimal container to study essential tonal configurations, upon which further enquiry could be conducted in the future. Moreover, computational tonal analysis is typically performed over short-time windows, and this restriction seemed favourable to our interests in assessing computational models of key estimation in EDM.

Yet another important motivation for this self-imposed limitation comes from our personal interest in studying the tonal implications of EDM's compositional and mixing configurations, from sequential mixing, to overlaying records and synthesisers, culminating in the Digital Audio Workstation as the main music production tool nowadays, as suggested by Figure 5.7. Although the practise of harmonic mixing is mostly concerned with the emotional effect of the sequential

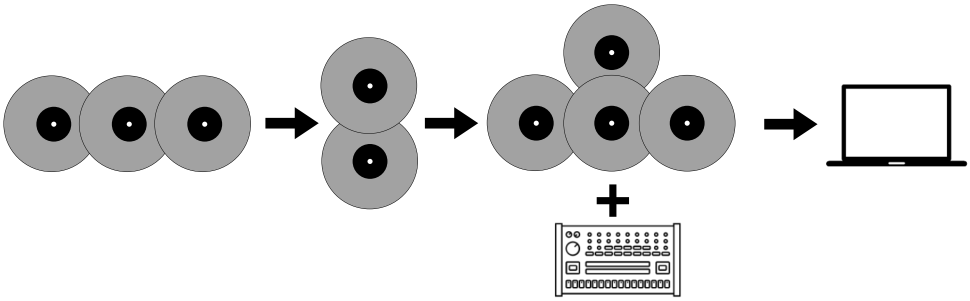


FIGURE 5.7: Typical mixing configurations in EDM. The arrows vaguely suggest a chronological evolution, from sequential mixing to overlaying records and synthesisers, culminating in the computer and the Digital Audio Workstation as the main music production tool nowadays. We are particularly interested in the tonal effects produced by overlaying records and other sonic sources.

arrangement of keys —conceptually closer to modulation as a maker of narrative,— we are more attracted towards the ambiguities presented by the simultaneous overlay of records, combined or not with synthesisers and drum-machines, and in general, in compositional approaches revolving around mixing and multi-tracking.

5.3.1 Key Changes

We would like to start our investigation trying to confirm the general assumption that EDM typically lacks of the tonal directionality present in euroclassical music and the alternating structuring found in popular musical styles. Although we had low expectations regarding modulation processes in EDM, analyses of both corpora confirmed —at least in our two-minute excepts— that key changes do not seem all that common in this music. The two collections described in this chapter add to a total of 70 tracks with structural key changes (47 in BP and 23 in GS$^+$), totalling to $\approx 3.5\%$ of the 1,986 files analysed ($1,486 + 500$).

As Figure 5.8 illustrates, most of these key changes occur between nearby regions, mainly relative, neighbour and parallel keys. Moreover, key changes do not seem to proceed gradually, by pivoting or creating a tonally ambiguous time period before confirmation, except perhaps, for tracks influenced by disco and other song traditions, which are more susceptible to present 'prepared' key changes, with cadential implications or pivotal chords, such as in Roisin Murphy's "Golden Era Disco Mix" by David Morales [3443052, house], where the global dynamics anticipate a modulation process achieved via a pivotal chord.

As a general rule, however, if present, key changes tend to occur suddenly at the start of a new hypermeter or after an atonal (unpitched) transition, typically accompanied by drastic changes in instrumentation, texture and mood. For example, in "Objects

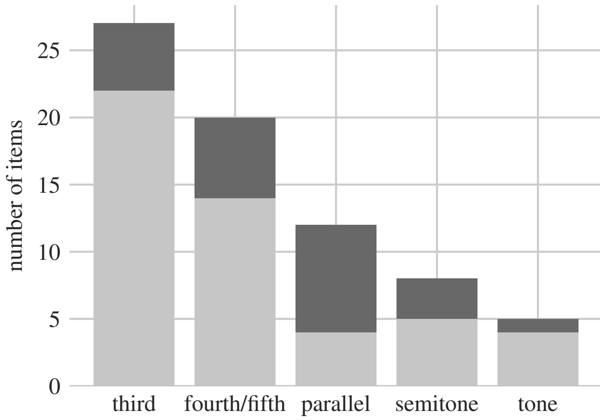


FIGURE 5.8: Key changes in the BP and GS+ datasets. Most changes are produced between relative and/or neighbouring keys.

and Purpose" [814505, tech house], Bronnt Industries Kapital create a temporary suspension of the main theme in A major with a sudden interruption on a single tone F, framed by two atonal instants, implying a total change of texture lasting for around 30 seconds. Similarly, in Skrillex's "All Is Fair In Love and Brostep" [5264038, dubstep], the key change from E minor to E♭ minor happens abruptly after a short spoken-voice interlude of less than two seconds, remaining in the new key of E♭ minor for the rest of the fragment. In tracks like "The Happy Pill" by Uzie [3526370, electro house], the key change from D minor to B♭ major is produced by the sudden disappearance of an overly present bassline playing a repetitive phrygian motive d→e♭→d, leaving the harmonic-filler alone, playing arpeggios on a sustained B♭maj chord. In this example, the confirmation of the new key is not achieved via a cadence, or a particularly characteristic pattern, but simply by accumulation of time in the new tonal situation. After a few hypermetrical repetitions, the listener seems to forget the previous bassline in D minor phrygian, accepting the region of B♭ major as the new tonal centre.

5.3.2 Common Diatonic Sets

In our analyses of the 500 fragments comprising the GS+, we found 166 unique pitch-class sets, after reduction to their prime form.[100] From these unique sets, the ten most common are listed in Table 5.4. The predominance of aeolian-related modes is clear over all other diatonic modes, followed in a much smaller number by

[100] As explained in Section 2.1.6, a *prime form* indicates a pitch-class set in normal order and transposed so that its first element is pitch-class zero.

pc-set	*tracks*	*(%)*	*tonic triad*	*closest mode(s)*
{02357 8↊}	122	24.5	minor	aeolian
{02357↊}	29	5.8	minor	aeolian and/or dorian
{023579↊}	22	4.4	minor	dorian
{03578↊}	18	3.6	minor	aeolian
{024579↋}	18	3.6	major	ionian
{0123578↊}	13	2.6	minor	aeolian-phrygian
{023578↊↋}	13	2.6	minor	aeolian-harmonic
{0357↊}	10	2	minor	minor pentatonic
{013578↊}	9	1.8	minor	phrygian
{023578}	7	1.4	minor	aeolian

TABLE 5.4: The ten most frequent pitch-class sets in the GS^{+} dataset.

dorian and ionian, and a few other variants of minor modes, aligning with the raise of minor modality in popular music after the 1960's suggested by Schellenberg & Von Scheve (2012), and conforming to the statistical distributions reported in Section 4.1.3. However, we should not forget that the GS^{K} data was already clearly biased towards minor modalities. In any case, in this corpus, our analysis confirm the predominance of the aeolian over other minor types, with an almost total absence of the minor harmonic scale, so characteristic of euroclassical praxis.

Figure 5.9 shows the pitch cardinality (i.e. the number of total pitch classes) distribution in the tracks of the GS^{+} collection. It can be easily seen how heptatonic sets —including the diatonic modes— are by far the most frequent (209 items), followed by hexatonic and octatonic collections. As shown in Table 5.4, 122 of such heptatonic sets found correspond to the aeolian scale, although other diatonic variants are as well represented, including dorian (22), ionian (18), phrygian (9) a mixolydian (5), adding to around 35% of the dataset. Some of these modal qualities are sometimes conveyed with reduced pitch-class sets of three-to-six elements, by grouping the elements of the tonic triad with other characteristic modal degrees. Looking at smaller pc-sets, we found 45 entries containing the elementary phrygian set {0137}, and a few other entries with phrygian-major qualities, such as Mark Broom's "M28" [3339291, techno], Dubfire's "I Feel Speed" [435443, progressive house] or Excision's "Headbanga" [3402886, dubstep]. Furthermore, the presence of chromatic sets seems relatively frequent, with 12 tracks containing the highly chromatic set {012345}. Other 41 elements present the chromatic minor-third cluster {0123}, apparently characteristic of an aeolian mode with phrygian cadential resolutions, as in Mathew Jonson's "Learning To Fly" [1964905, techno] or Tony B's "Je T'aime" [3995054, electronica].

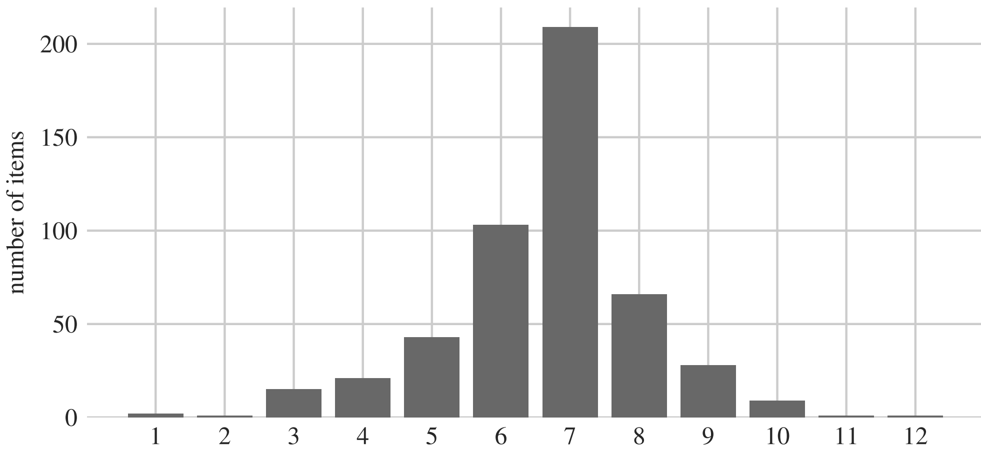


FIGURE 5.9: Cardinality distribution in the GS+ dataset, where it can be seen that most fragments in the corpus clearly contain heptatonic pitch-class sets (i.e. diatonic modes).

5.3.3 Pitch Sparsity

One of the most interesting properties of reduced pitch-class sets is the varying degree by which they can suggest particular tonical and modal contexts, typically presenting openness to multiple interpretations. For example, the only pitched contents in The Chap's "Woop Woop" [298989, minimal] consist in a repeating bassline with pitch-class set {69Ɛ} under a track otherwise populated with atonal effects (Figure 5.10, left). In this track, the inferred modality must be decided upon very restricted materials, such as the minor third f♯–a. Furthermore, the rhythmical and metrical balance between both tones, creates a perfectly bimodal track (F♯ minor | A major), given that other tonally disambiguating elements are virtually absent. Similarly, in Marco Carola's "Play it Loud!" [1681085, tech house] the pitched materials are almost solely percussion over a repetitive bass-layer that presents a heavily detuned and reduced to the major third pc-set {6Ʈ}. However, in this particular example, the tonal centre is more easily identifiable as G♭ major, given the metrical prominence of g♭ and the fact that both notes are interpretable as part of a G♭maj triad.

Besides the regular major and minor labels, in Section 5.1 we introduced two additional tags to characterise entries falling out of this binary vocabulary ('other' and 'no-key'). Although 'other' is quite an unprecise word, we chose this tag to denote tonical tracks that are neither major nor minor. These tracks are typically divided into two general groups, broadly consisting of (a) ambivalent tracks presenting reduced pitch-class sets without clear indicators favouring any major or minor scalar configuration, in which the more extreme case is represented by 'monotonic excerpts;

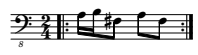
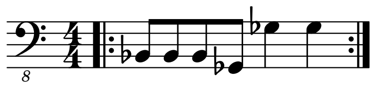



FIGURE 5.10: Bassline transcriptions of The Chap's "Woop Woop" [298989, minimal] (left) and Marco Carola's "Play It Loud!" [1681085, tech house] (right). These reduced pitch-class sets clearly present some modal hints, despite their limited pitch content.

and (b) an explicit tendency composed of tracks with a tonic diminished triad or a pseudo-locrian scale, with mostly dissonant structures. Similarly, the 'no-key' label represents two other paths, grouped into (c) clearly *atonal* –unpitched— tracks, typically consisting of just percussive events; and (d) atonical excerpts, presenting pitched materials without a particular sense of tonically, as epitomised by serial music.

Amodal, Atonal, and Other Fragments

The most extreme case of pitch sparsity can be traced in tracks where the only pitched sound is the tone of a percussive instrument, as exemplified by Seba's "Predator" [5078640, drum'n'bass], where a conga tuned in e♭ seems to provide the most important tonal cue to a track that otherwise contains vague reminiscences of a chromatically ornamented E♭ minor. Other similar examples include Arjun Vagale's "Drohnen" [3003894, techno] or Glitter's "Dance Floor" [4840289, minimal] in C other monotonic and E other monotonic, respectively (according to the nomenclature described at the beginning of this chapter). We could use the term *amodal* to refer to these type of tracks, for they simultaneously represent the greatest unambiguity regarding their tonal centre and a highly ambivalent modality, since no specific scale can be inferred from a single tone. As explained above, we include monotonic tracks in the more general container of 'other', which agglutinates tracks with a clear tonic but an open modality, thus being quite versatile for harmonic mixing purposes. Besides monotonic tracks, the amodal class typically comprises items with up to three pitch-classes, excluding intervals characterising major or minor modes.

François Manzo's "Decadence" [4372159, minimal] at first sounds like an atonal track, presenting only percussive elements, and a low distorted voice in the background. However, after repeated listening —the looping mechanism— one becomes aware that two percussive instruments are indeed tuned, producing a slightly detuned perfect fifth all along the fragment, that ends up conveying a quite clear tonal centre. Similarly, Fallhead's "Field & Corridor" [923844, minimal] is essentially atonal and mostly percussive, although a soft sense of tonicality around a♭ is produced by the occasional appearance of short events tuned to this pitch across different octaves. In "Fragment" by DatA [3400782, drum'n'bass], the snare drum seems to be tuned to b♭, although this is literally all the pitch material in this excerpts until the appearance of a tone cluster with sampled string instruments, towards the end of the fragment.

Most listeners would consider these examples simply as atonal (recall that we use this term literally, in the acceptation proposed by Tagg (2014), to describe music made without pitch). However, at the same time, there are subtle indicators of tonicality in the form of tuned percussive events, that due to the highly repetitive nature of these musics, end up constituting a real tonal centre to these musics. On the other hand, in our analyses we did not find characteristic atonical excerpts, and even tracks containing pc-sets with a very high cardinality (10 to 12 pitch-classes) still provide a sense of tonal centre (e.g. Gaiser's "Some Slip" [2081732, minimal] or "Beholder" by DJ Hidden [2725289, drum'n'bass]).

Semitone and Tritone Ambivalence

As mentioned above, the 'other' label also includes samples centred around the diminished chord and/or the locrian mode, whose main characteristic interval is the tritone, which other authors have previously associated to techno (see Section 2.4 for early references by Tagg (1994) and Spicer (2004). Considered individually, the tritone has the special quality of dividing the octave into two equal parts, creating the only inversionally identical interval in an octave. This places this interval in the position of being the most 'neutral' interval. Moreover, the fact that the tritone it is neither physically —harmonically— related to other musical intervals, nor it feels melodically 'natural', situates this interval as residual in most modal practises. However, it is this interval's neutrality what can be used to create perfectly bimodal tracks (in the acceptation explained in Section 2.3.4, and to which we return shortly). For example, in "Goatherd" by The Cow [6235742, breaks] or "Rave On" by Electric Rescue [61578, techno], the main pitch material consist in basslines with tritone alternation, presenting just two pitch classes {06}, and making absolutely unnecessary to favour one or other tonic note.

On the other hand, the semitone seems to carry much narrative power, as it likely represents the most expressive melodic interval —the closest note to another note. This property has been exploited in euroclassical music, situating the leading tone ($\natural\hat{7}$) at the forefront of euroclassical tonality. However, the power of the descending leading tone (summarised in the phrygian $\flat\hat{2}$) has not been exploited in euroclassical music and other popular music styles, although it is found as a melodic tension in music from the Renaissance and traditions such as flamenco.

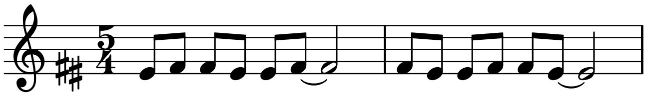


FIGURE 5.11: Semitonal ambiguity in Ligeti's *Musica Ricercata II*. This composition, presenting just two notes one semitone apart, represents the quintessential play of semitonal ambiguity between upper ($\natural\hat{7}\rightarrow\hat{1}$) and lower leading tones ($\flat\hat{2}\rightarrow\hat{1}$), suggesting one or another interpretation uniquely based on the metrical organisation of short and long events.

A paradigmatic example to illustrate the double ambiguity of the semitone is provided by the second number of Ligeti's *Musica Ricercata*, shown in Figure 5.11, where the semitonal play between the two notes seems to alternate the sense of tonicality between the two only notes of this excerpt, interpreted as upper ($\natural\hat{7}\rightarrow\hat{1}$) and lower ($\flat\hat{2}\rightarrow\hat{1}$) leading tones relations.[101] This type of semitonal double-tonic is frequently heard in dubstep and drum'n'bass, where sometimes the only pitched elements are a semitonal movement in the bass.

Both tritone and semitone intervals are expressed in the diatonic locrian mode, although in EDM they tend to appear in sparser configurations. Take, for example, the case of "Move it 2 The Drum" by Ambush [4311630, tech house], where the {016} set ($\hat{1}$, $\flat\hat{2}$, $\flat\hat{5}$) is used throughout. This particular pc-set is sometimes used in drum'n'bass and dubstep (e.g. F3tch's "Fuck Your Mum" [1787061]), either alone or inserted in a larger pitch contexts, and has been referred to as the 'viennese trichord', to express Webern's preference for pitch-class sets clearly deviant from consonant harmony. However, in EDM, semitones and tritones are typically used tonically. Other examples of pieces constructed around the tritone interval include "Rave On" by Electric Rescue [61578, techno], Kaiza's "Kaneda VIP" [5536316, drum'n'bass] or "Goatherd" by The Cow [6235742, breaks]. Among tracks explicitly presenting a locrian mode are Louie Fresco's "Owl Night" [3298819, deep house] and Manel Díaz's "Dopamine" [5419394, minimal].

Moreover, 86 items from the GS^{+} dataset contain the lowered leading tone relationship ($\flat\hat{2}\rightarrow\hat{1}$), especially in genres such as tech house and techno. Tritone relationships appear in 53 entries, integrated within larger pc-sets, and mostly clustered in drum'n'bass and tech house subgenres. Regarding the 'viennese trichord' mentioned above ({016}), it appears as a subset in another 23 excerpts, showing that typically 'dissonant' patterns are common in some styles of EDM, especially tech house and drum'n'bass.

[101] This piece of music has been made popular mostly due to its inclusion in Stanley Kubrick's film, *Eyes Wide Shut*.

5.3.4 Tonical Ambivalence and Modal Ambiguity

In previous sections of this chapter, we have repeatedly referred to the tonal ambiguity found in our recently analysed datasets. Although the types and means to present such ambiguity are multiple, in our analysis we found two main tendencies, grouped broadly into *tonical ambivalence* and *modal ambiguity*. According to the Oxford Dictionary of English, the term 'ambiguity' carries a positive valence, implying the "quality of openness to more than one interpretation" (Amb, 2010a). On the other hand, the notion of 'ambivalence' typically raises a somewhat more negative connotation, describing the "state of having mixed feelings or contradictory ideas about something" (Amb, 2010b). Throughout this dissertation, we have tried to adjust these two notions to refer to distinctive musical situations, respecting their original connotations as much as possible. We typically consider tonally ambiguous excerpts those providing an 'excess of information', by which multiple interpretations are possible and acceptable. In particular, we use the notion of 'modal ambiguity' to refer to excerpts with a clear tonal centre, but suggesting multiple modalities, typically merging major and minor scales, as explained in Section 2.3.2. Complementarily, we prefer the term 'ambivalence' to describe scenarios with sparser tonal information, presenting unclear moments that could be potentially read under various lights, such as the monotonic and 'other' fragments described in the previous section. Tonical ambivalence mostly comprises practises that can be characterised as *bimodal* or *polyscalar* (in the restricted meanings introduced in Section 2.3.4), conveying the perception of more than one tonal centre in the same fragment. According to the figures in Tables 5.2 and 5.3, the Beatport and GiantSteps+ datasets comprise 87 tracks presenting major/minor modal ambiguity, plus another 90 items labelled as bimodal, therefore ambivalent regarding their tonal centre.

Modal Ambiguity

As we have just noted, we refer to modal ambiguity to denote tracks with a clear and single tonic, but an ambiguous modal structure. This type of ambiguity resembles some of the practises present in rock modality, as described in Section 2.3. For example, (a) a melodic minor bassline could be harmonised with major chords. A similar effect could be produced by (b) extremely saturated timbres from synthesisers, presenting rich harmonic series that might show a clear major chord in the chromagram, without actually playing the chord components. This last scenario is assimilable to Lilja's claims about the modal ambiguity introduced by metal power chords (Lilja, 2009), and can be heard in tracks such as "La Girls" by David Tort [298024, house], where, despite the lack of a third in the pitch-class set, the timbre of the synthesiser boosts the major third over the fundamental tone.

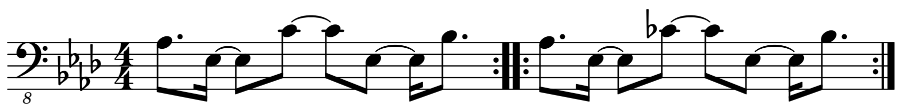


FIGURE 5.12: Example of major/minor ambiguity in Ettica's "Forcefields", where each successive hypermeasure presents arpeggios alternating over the major and minor triads.

Yet another possibility consists in the usage of pitch-class sets containing both major and minor thirds, normally sequentially arranged in melodic lines and/or pitch aggregates as exemplified by the excerpt from Ettica's "Forcefields" [4955940, tech house] shown in Figure 5.12, where major and minor triad arpeggios alternate over the tonic A♭.

Tonical Ambivalence

In contrast, tonical ambivalence typically comes into play in modulation processes, where a sense of tonal centre is gradually —or not— substituted by a different tonic. As shown in Section 5.3.1, key changes are not very characteristic of EDM. However, we found a number of bimodal tracks in our corpora that convey comparable degrees of tonical ambivalence. A bimodal excerpt is different from a key change in that it does not provide a directional or sequential movement from one key to another, but it easily allows a non-conflictive multiple interpretation of an excerpt as having two different tonal centres —a visual analogy to this phenomenon could be the figure-ground grouping in Gestalt psychology (Fineman, 1996). As explained in Section 2.3.4, bimodality is often produced sequentially by equivalent forces exerted by relative or neighbouring keys, although other likely scenarios include loops with an oscillating movement between two notes or chords.

For example, "Fake Emotion" by Modeselektor [65102, electronica], presents a single Cmaj6/Amin7 chord throughout the fragment, while the bassline alternates between a and c with equivalent metrical weight. Furthermore, the vocal melody in this example consists in a repetition of the same four notes in the same order (c→g→e→a), so that the complete pc-set of all pitch layers adds to only four notes {0479}, which could easily be interpreted as a subset of either C major pentatonic or A minor pentatonic. A similar case is presented in the two remixes of Slam's "Alien Radio" [techno], which essentially consist of a single bassline with relative tonics d/f. In one of the versions, a remix by Tony Thomas [297059], the harmonic-filler plays a constant Fmaj chord, with occasional appearances of the blue note a♭, whereas in Darren Emerson's mix [297065], the pitch content of the track is essentially limited to the bassline. An example of a bimodal chord shuttle is provided by Speh's "Reaching You" [3116337, drum'n'bass]. Here, the tonal material is reduced to an oscillating

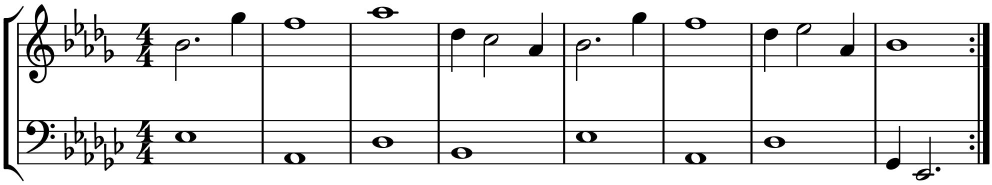


FIGURE 5.13: Main bassline and lead melody of Schatrax and Silicone Soul's "Mispent Years (Silicone Soul Darkroom Dub)" . They have been rhythmically simplified for readability. Both layers could be interpreted as being in E♭ minor and B♭ minor, respectively (note the different key signatures).

shuttle (Amin7↔Dmin7), and although the hypermetrical cycle starts on Dmin, the few pitched interventions in the vocal layer present an a-g-a movement that counterbalances the modality towards Amin. Furthermore, at least once in the except, the beat-layer is removed on Dmin and reentered in Amin, creating a temporary sensation of hypermetrical shift towards A minor.

There is another possible bimodal configuration, vertical or polyscalar, by which different tonal layers seem to present different scales, typically complementary (consonant), but conveying two relatively clear tonics. Truth's "Antent" [4805277, dubstep] typifies this vertical bimodality, with a clear G♯ aeolian sequence as harmonic support (i→♭VII→♭VI) over which a melody seems to tonicize a b with a B pentatonic scale. The two elements, heard independently, would be clearly perceived as conveying two different tonics. However, given that they operate simultaneously —and that they are relative keys— the ambivalence is guaranteed. A similar layering happens in Schatrax's "Mispent Years" [191347, house], shown in Figure 5.13, where the bass layer presents an orthodox sequence in E♭ minor, whereas the melodic line would likely be heard in B♭ minor, if detached from the other textural layers. As one last example, Tony Traxx's "Her Shoes" [5905170, deep house] represents a complex example of both types of bimodality operating at the same time (it is actually a polytonal track). The keyboard part plays a metrically balanced chord shuttle Emaj7↔C♯maj7 over a bassline centred on g♯. Above this, a melody outlining a D♯ aeolian, counterbalances the tonal weight, and creates the impression that the tonic is actually d♯. However, the insistence of the bassline and the chord sequence also make plausible an interpretation of either c♯ or g♯ as tonic notes.

***

In this chapter, we have presented our study of tonal practises in EDM, based on the analysis of two new corpora, adding to nearly 2,000 fragments grouped across different subgenres. We have introduced a novel annotation protocol, that attempts to indicate passages with tonical ambivalence and modal ambiguity, accordingly annotating our corpora with these new labels. Furthermore, we have attempted a general description of tonal practises in EDM, presenting typical modal distributions in this meta-genre, as well as discussing other characteristic tonal effects, mostly originating in the use of sparser pitch collections. In the following chapter, we discuss our methods for computational key estimation, which should be understood as an effort to incorporate some of the labels introduced along this chapter within the classification vocabulary of algorithms.

Chapter 6

Automatic Key Estimation in EDM

"To an ever greater degree
the work of art reproduced
becomes the work of art
designed for reproducibility."
Walter Benjamin

In this chapter, we finally describe the approaches to automatic key determination in EDM that were developed in the course of our research. As such, most parts of this chapter are taken from two existing publications (Faraldo et al., 2016a, 2017), although we offer additional and complementary supporting material, including more detail of analysis and evaluation.

As we have seen in Chapter 5, EDM presents several tonal practices clearly differentiated from other musical styles, such as the generalised absence of modulation, the lesser importance of chords and harmony —except for genres with roots in song traditions— and a tendency to pitch sparsity, manifested in reduced pitch-class sets with with less than seven elements. The current chapter represents an attempt to develop key estimation algorithms that take into account some of these tonal idiosyncrasies, widening the classification vocabulary beyond the common binary output, with the intention of bringing forward creative applications in the domain of applied MIR, and as a means of obtaining music-theoretical insights. As reported in previous chapters, the totality of existing methods addressing key estimation in EDM remain within the euroclassical modal division into major and minor tracks, something that has proven unnecessarily constraining for most popular music styles (e.g. Temperley & De Clercq, 2013), and certainly for a good amount of electronic dance music, as we have tried to underline at several points in the text.

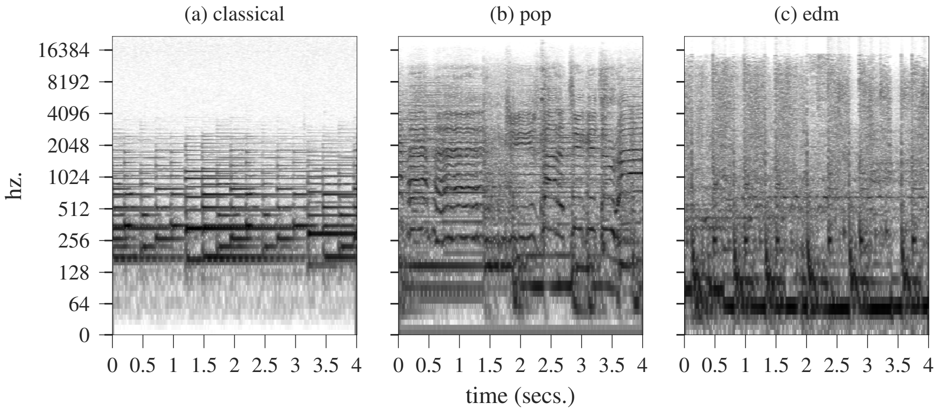


FIGURE 6.1: Log-spectrogram of four-second excerpts of musics from different styles: (a) Bach's "Prelude 1 in C BWV 846" from *The Well-Tempered Clavier*, rendered by Glenn Gould; (b) The Beatles "Ticket to Ride"; and (c) DJ Hidden's "The Narrators" remixed by Eye-D, an example of drum'n'bass excerpted from the GiantSteps dataset.

We start our presentation with a few considerations about timbre in EDM, before describing the various stages of our proposed approaches. Our explanation is organised in a linear fashion, discussing the methods we have developed within a chronological narrative. We start describing our variations on HPCP calculation and other low-level features. Then, we continue presenting our statistical profiles, along other processing stages introduced at several stages of the processing pipeline. We conclude the chapter with a final discussion of our methods in comparison with existing state of the art algorithms.

6.1 Timbral Considerations of EDM

As we have seen, two of the sonically distinctive features of EDM are its all-electronic sound and the central role played by percussion, over which other pitched materials and sound effects might be layered. These characteristics are reflected in the spectral representation of EDM signals, where saturated synthesisers often turn into complex spectral envelopes —quite different from those originating in acoustic sources— and the ubiquity of percussive sounds populate the spectrum with fast transients. These rich spectra present important challenges to the extraction of pitch and tonal information from audio, which should not be ignored in the design of algorithms for EDM.

Figure 6.1 shows log-frequency spectrograms of three musical excerpts belonging to musics from different eras. All three spectrograms span a duration of four seconds and were taken sixty seconds into the track, in order to avoid possible sparsity at the

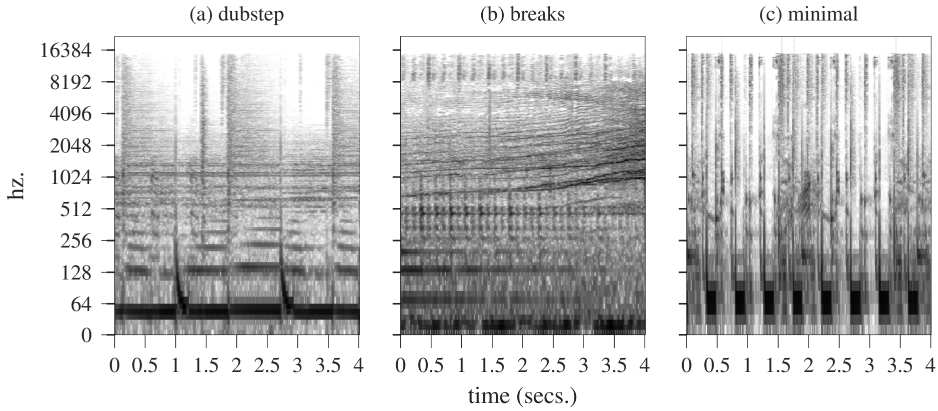


FIGURE 6.2: Log-spectrogram of four-second excerpts of three EDM subgenres —dubstep, breaks and minimal— (a) GroYen's "Quietude", (b) "The Guitar" by Romanto and Out of the Drum, and (c) Field and Corridor's "Fallhead". These three examples are taken from the Beatport dataset.

beginning of the recordings. Figure 6.1a shows an excerpt of the opening Prelude from Bach's *The Well-Tempered Clavier* cycle. This spectrum is relatively narrow, since it is a recording of a solo polyphonic instrument —in this case, Glenn Gould's piano— showing clear horizontal bursts of energy, corresponding to single musical tones and their harmonics. In contrast, Figure 6.1b presents the richer spectrum of "Ticket to Ride", a pop song by The Beatles's from their album *Help!* A typical pop-rock sound can be inferred from it: the presence of a drum-kit creates spikes that extend quickly and vertically through the spectrum; the lower end is populated by the presence of a bass drum and a bass guitar; there are constant horizontal lines representing chord strumming by rhythmic guitars; and serpentine lines indicate the presence of vocals, with less stable pitch and expressive oscillations. Therefore, although presenting a considerably richer spectrum than Bach's keyboard music, we can infer from the spectrogram that pitch is still a prominent aspect of this music. By looking at Figure 6.1c, however, we do not obtain such a clear impression regarding the tonal aspects of the excerpt. In this drum'n'bass example, from Eye-D's remix of "The Narrators", the whole spectrum lacks thiner horizontal lines suggesting the presence of pitch. Furthermore, most of the energy concentrates in the lower end, possibly mixing tonally relevant sounds with other expressive effects.

Additionally, Figure 6.2 shows three log-spectrograms from other EDM subgenres, namely (a) dubstep, (b) breaks, and (c) minimal techno. All three spectra present a concentration of energy in the lower end, with sudden changes in the spectral distribution. The dubstep spectrogram represents a reverberated snare drum over tuned percussive sounds, whereas the excerpt of Romanto's "The Guitar" captures an ascending general glissando as part of a build-up section towards the 'drop'. The example of minimal consists in a essentially percussive fragment, with extremely short and high-pitched events.

The items in Figure 6.2 represent extreme cases when it comes to hinting underlying tonal structures. Generally, we can assume that other EDM genres —especially house and its variants— are more likely to present typical distributions of spectral energy, just as much as they recombine musical elements from previous traditions. However, either as a matter of sparsity (as in minimal techno), timbral saturation (as in progressive house and trance), or excess of activity in the lower end (such as in drum'n'bass or dubstep), EDM tends to present some of the following distinctive spectral traces:

- Local spectra tend to be flatter, due to the ubiquity of percussive sounds, potentially masking regions with meaningful tonal content.
- Similarly, tonal motion often concentrates on the lower register, where spectral calculations normally offer less resolution.
- Some types of EDM are characterised by tonal effects such as *glissandi* or pitched percussive elements, that can be difficult to identify as quantised and/or stable pitch units.
- Extreme timbral saturation plays a role in some sub-genres, creating spectral envelopes that might bear little resemblance with the 'natural' envelopes of acoustic instruments.
- Furthermore, pitch is no-longer a primary constituent of this music. Some styles such as techno, minimal or drum'n'bass could present very sparse or no pitch materials at all.

6.2 An Evaluation Method Receptive to Tonal Ambiguity

Prior to the presentation and discussion of the various stages of our methods and their corresponding evaluation, we would like to introduce the evaluation strategy applied in the following sections, which involves a small modification of the MIREX evaluation method used in the preliminary evaluation in Section 4.3. Our intention with this step, is to incorporate some open-ended descriptions regarding bimodal excepts and major/minor ambiguity, as described in the previous chapter.

With this laxer evaluation method, we consider as correct any estimation included within the range of bimodal (e.g. C major | A minor) or modally ambiguous annotations (e.g. C major | C minor). We expect that this approach will increase the performance for all algorithms in all scenarios, supporting the importance of ambiguity and ambivalence in EDM. On the other hand, for the purpose of comparison amongst the various solutions, this laxer method should not introduce significant differences. In any case, we consider that this assessing methodology remains closer to the musical and perceptual reality of tonal ambiguity, becoming tolerant to different although valid readings, rather than imposing a disambiguation where this might not exist in the musical reality. Besides the common 'major' and 'minor' labels, we incorporate the tag 'other' to account for passages where a major or minor modality can not be directly inferred from listening, and a 'no-key' classifier, describing both atonical and/or unpitched excerpts. Apart from recognising these new labels and assessing ambiguity positively, this method follows the MIREX weighting convention discussed in Section 4.2.1, producing an overall global score according to the same weights. We could have weighted differently the various types of errors by giving more importance to parallel and relative keys than to neighbour relationships, since the latter originate in harmonic/chordal tonality, whereas the scalar configuration of EDM (with aeolian and mixolydian modal variants) seems to favour relative and parallel keys as closer than neighbours. However, we apply the weighting system as presented in Table 4.2 in order to compare the performance of our multi-modal labels with previous methodologies. After all —and despite the concrete figures obtained— the relevance of assessing types of familiar errors lays in understanding where the tonal confusion of algorithms —and our own tonal perception— might reside.

Nevertheless, the power of this evaluation method necessarily relies in annotations providing a greater detail of verbosity, a requirement that is only met by the BP and GS+ datasets. For all other test collections, the more restricted MIREX scoring system should produce exactly the same results. Similarly, assessing 'no-key' and 'other' tracks with systems only capable of a binary vocabulary would carry no additional advantage or information. In these cases, we exclude from the evaluation the items labelled with these additional tags.

6.3 The Basic System: EDMA and EDMM

In this section we describe our first approach to key estimation in EDM. Since the academic milieu of this research has been provided by the Music Technology Group at Universitat Pompeu Fabra, it felt natural to ground our investigations in previous developments within the Group. In this sense, MTG's audio analysis framework *Essentia*[102] (Bogdanov et al., 2013), seemed an optimal starting point to test readily available technology, providing a number of remarkable advantages:

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

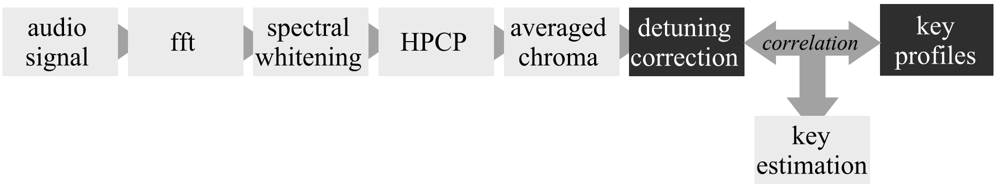


FIGURE 6.3: Processing pipeline of our baseline algorithm, taking the approach by Gómez (2006b) as a reference model. We have darkened the processing stages to which we have mostly contributed.

- As a solid and actively maintained library, Essentia contains an increasing state of the art methods for a variety of MIR endeavours, including tonality related tasks.
- It is implemented in C++, being fast and efficient, and its particular data-flow design makes it ideal for prototyping.
- Furthermore, Essentia provides a Python interface, a language that is becoming the standard in scientific computing.
- The HPCP and key detection methods developed by Gómez (2006b), and described in Section 3.3, were already implemented in the framework.

6.3.1 General Description

Figure 6.3 provides an overview of our baseline system, developed in its entirety in Essentia. A detailed description of this method is provided in Faraldo et al. (2016a). This approach should be seen as an elaboration of the method by Gómez (2006b), to which we added specific key profiles, obtained from a subset of the KeyFinder dataset (KDF), presented in Section 4.1.3. Besides, we inserted a detuning detection function, which is a simplification of the one proposed by Harte et al. (2006), and that proved highly successful in our experiments. In order to obtain a global-key estimate, we aggregate all the chromagrams from the analysed windows, and correlate this global descriptor with chromatically rotated key profiles, as proposed by Gómez (2006b).

6.3.2 DSP and HPCP Configuration

Before introducing our newly created profiles in the next subsection, in this block we present a preliminary consideration of some low-level decisions, regarding the spectral analysis and HPCP calculations, at which we mostly arrived by a mixture of theoretical assumptions and heuristic experimentation.

	parameter	*defaults*	*chosen value*
spectral	window size (pt)	4,096	=
	hop size (pt)	2,048	16,384
	window shape	blackman-harris	hann
peak picking	peak threshold	0.00001	0.0001
	max. peaks	10,000	60
	frequency range (Hz)	40–5,000	25–3,500
HPCP	split frequency range	✓	✗
	split frequency (Hz)	500	n/a
	contribute first n harmonics	8	4
	reference freq. (Hz)	440	=
	size (bins)	36	=
	weighting	squared cosine	cosine
	weight size (semitones)	1.3	1.0
	normalisation	unit norm	=
	non-linear transform	✗	=
key	polyphony	✓	✗
	three-chords	✓	✗
	n. harmonics	4	n/a
	slope	0.6	n/a

TABLE 6.1: Basic configuration of our key estimation algorithm as described in Faraldo et al. (2016a), compared to the default settings as described in Gómez (2006b) and/or implemented in Essentia.

As we have already declared, our method relies entirely on the HPCP implementation available in Essentia. However, we performed substantial modifications in the default calculation parameters, as shown in Table 6.1. For example, we used a hop size four times the analysis window. We realised that this increment significantly speeded up the calculation time, providing slightly better results than with a hop size equivalent to the window size. Our explanation to this behaviour might be justified by the fact that a hop size of 16,384 at 44,100 Hz represents 371.5 ms of audio, what is equivalent to a single beat at a tempo of 161.5 BPM. This implies that even in faster EDM subgenres, such as drum‘n’bass (with tempos between 150-170 BPM), the algorithm analyses 3 or 4 frames per bar, what seems to be sufficient for key determination, assuming that chord sequences are not a prominent feature of most EDM genres, and that tonal changes tend to occur over longer periods of time. Regarding the peak-picking configuration, we reduced considerably the amount of peaks, upon the assumption that EDM tracks would contain more peaks than other musical styles. Furthermore, we changed the frequency range of our analysis, to accomodate low

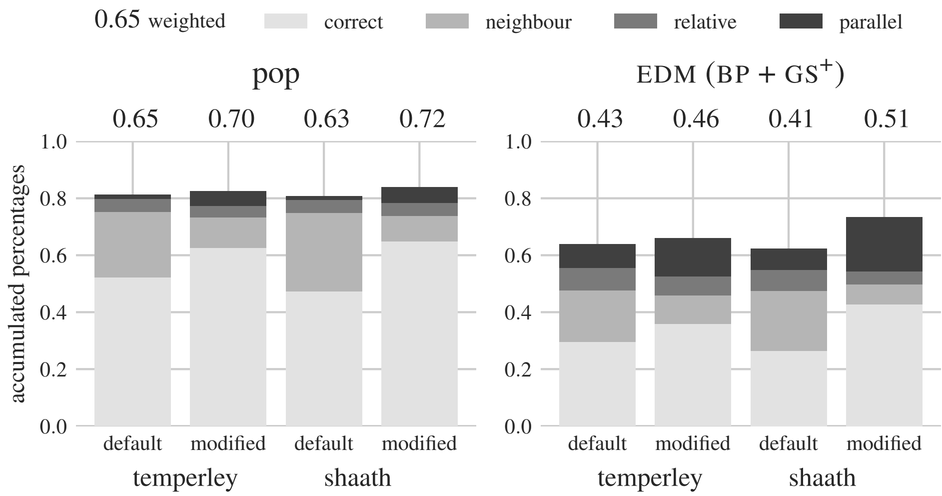


FIGURE 6.4: Effect of DSP and HPCP parameters with the profiles by Temperley (1999) and Sha'ath (2011) on combined datasets of pop and EDM. Throughout this chapter, all the bars in the evaluation figures show the accumulated percentages of commonly accepted error types (neighbours, relative and parallel) besides the percentage of correctly classified items. Above each bar we present an overall weighted score, calculated according to the weights in Table 4.2.

frequencies down to 25 Hz —to include tuned bass-drum sounds in our analysis— and we cut the higher end at $3,500$ Hz, in order to get rid of an excessive presence of harmonic components beyond that frequency. In this regard, we also reduced the contribution of the peaks in the signal to the various HPCP bins from eight to four harmonics.

Regarding the key profile adaptation, we remind the reader that Gómez (2006b) arrived at her 'polyphonic' profiles by adding chord-component contributions and harmonic weights to originally symbolic models, as reflected in the lower part of Table 6.1. In contrast, we reject a profile redistribution based on chordal polyphony, on the assumption that a good deal of EDM is not based on chord structures and could be regarded as essentially melodic. Furthermore, our key profiles —discussed in the next subsection— were extracted statistically directly from the audio signal, so this adaptation stage seemed unsuitable in our design.

Throughout this chapter, all figures with evaluation results are presented as bar-charts with accumulated percentages of commonly accepted error types (neighbours, relative and parallel) together with the percentage of correctly classified items. Above each bar we present an overall weighted score, calculated according to the values in Table 4.2, according to the MIREX convention. Besides, all the experiments carried throughout this chapter assume monaural audio files at a sampling frequency of 44,100 Hz. Figure 6.4 shows the effect of the DSP and HPCP modifications on combined corpora of popular music and EDM. The 'pop' label comprises the three popular music datasets described in Section 4.1 (BTL, BB, and RS), whereas in the remainder of this section, the EDM evaluations are conducted with the merged GS^{+} and BP datasets, excluding KFD to avoid likely overfitting effects, given that our new key profiles were derived from a subset of this corpus, as it will be detailed shortly. In this evaluation, we use the profiles by Temperley (1999), for they are generally regarded as highly effective profiles (although they are biased towards euroclassical music, as we have seen). Additionally, we conduct the same evaluation with the profiles proposed by Sha'ath (2011), which consist in heuristic modifications of the probe tone results by Krumhansl & Kessler (see Chapter 3 for details and figures). As it can be inferred from Figure 6.4, our low-level modifications have an impact in all the scenarios evaluated. In the the pop music dataset, the improvement mostly implies a shift from neighbour errors towards correctly classified instances, obtaining a better weighted score with both key profiles. Regarding the EDM dataset, the situation is comparable, although the impact of the modification of default low-level parameters is more extreme with Sha'ath's profiles. Both profiles offer a worsened performance on EDM, when compared to pop music dataset, although this seems to be the regular scenario, as discussed in Section 4.3.

6.3.3 EDMA and EDMM

As explained in Chapter 3, one of the most important ingredients of a template-based system is the particular set of tonal hierarchies represented by the key profiles. In order to improve the performance of the baseline key estimation system presented in the previous block, we extracted new major and minor profiles from a collection of audio files and annotations gathered from the internet. Our main resource was Sha'ath's KeyFinder dataset (KFD), from which we excluded entries belonging to other popular musical styles, comprising around 20% of his annotations. We completed our training set with other online sources described in Section 4.1.3, totalling to 925 full-length EDM tracks. With this collection, we performed two subsequent operations in order to obtain a new set of profiles, summarised in the following bullets:

1. First, we extracted major and minor profiles, as the median vector of the averaged chromagrams of the complete training set. Throughout this work we refer to these profiles as EDMA. The resulting vectors are shown in Figure 6.5, where it is perhaps worth noting the higher presence of the subtonic ($\flat\hat{7}$) in both modalities, indicating a prominent presence of mixolydian and aeolian, over the euroclassical ionian and minor harmonic distributions.

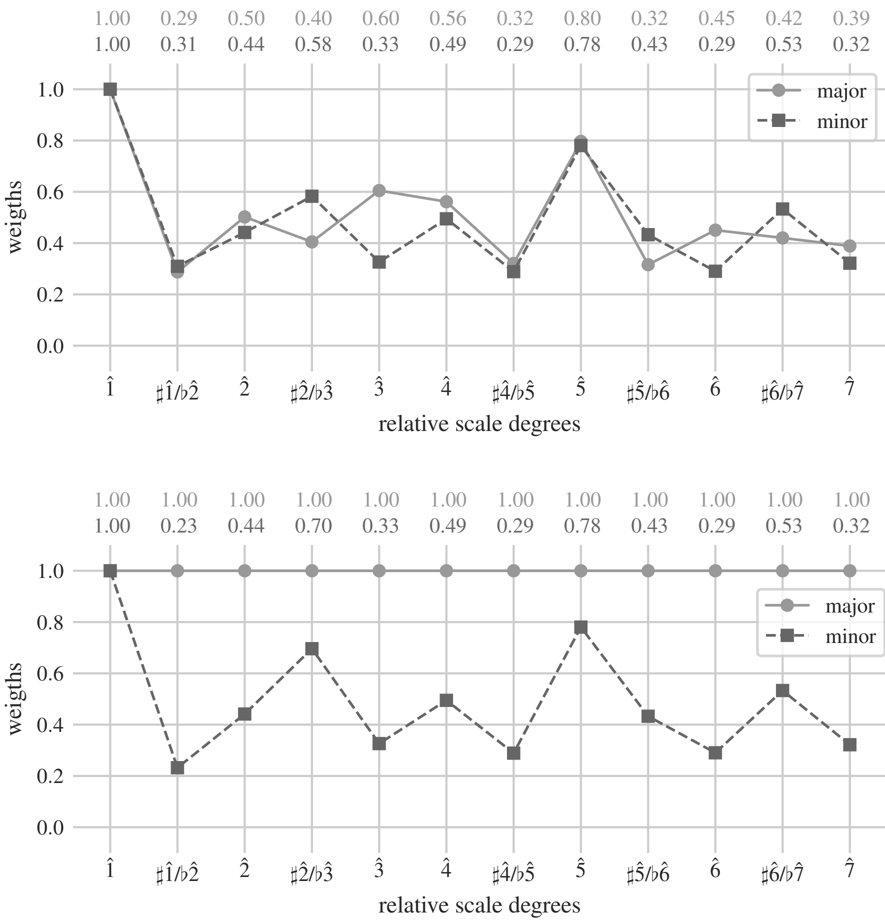


FIGURE 6.5: Key profiles derived from statistical analysis of the KFD dataset: EDMA (top) and EDMM, (bottom), obtained after further manual adjustment.

2. After the EDMA profile extraction, we performed some heuristic adjustments in the minor profile, slightly raising the weight of the minor third ($\flat\hat{3}$) and lowering the $\flat\hat{2}$. More radically, we flattened completely the major profile, forcing all estimations into minor, based on the lower proportion of major tracks in EDM corpora, as we have shown in previous chapters. These manually modified profiles are shown in Figure 6.5 (bottom).

Figure 6.6 presents the estimation results of our method with the new EDMA and EDMM profiles, using the HPCP settings presented in Table 6.1. The most noticeable effect is the drop in correctly classified instances with the EDMM in popular music. This is a natural effect of the flat major profile, as indicated by the large percentage of parallel errors, given the larger proportion of major tracks in the popular music datasets. On the other hand, the effect of our modified HPCP calculation has a direct

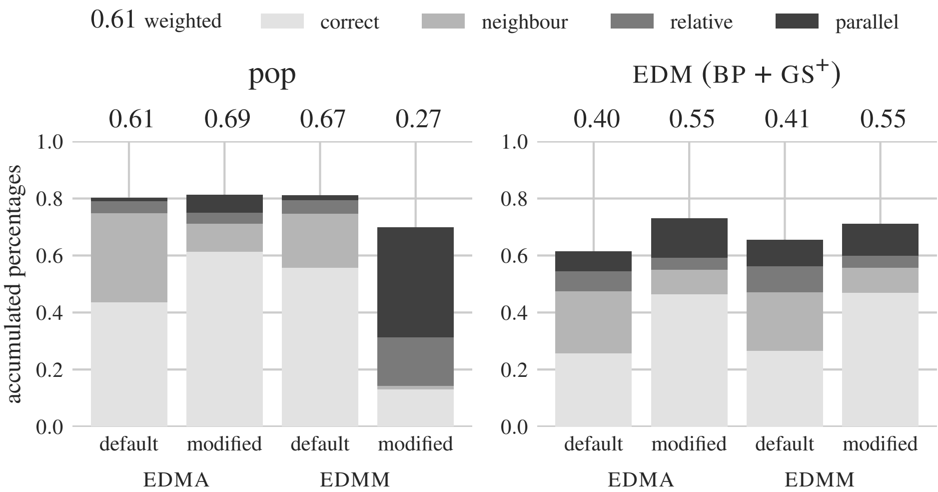


FIGURE 6.6: Effect of low-level parameters with the EDMA and EDMM profiles on the combined datasets of pop (left) and electronic dance music (right).

effect in the performance of both profiles in electronic dance music, reflected in an increment of 0.15 points for both profiles, obtaining a timid improvement over the profiles by Temperley and Sha'ath shown in Figure 6.4.

6.3.4 Spectral Whitening

Besides the new profile creation just described, we inserted a spectral whitening stage, in order to 'flatten' the spectrum according to its spectral envelope, increasing the weights of the predominant peaks. Spectral whitening is a technique widely used in both digital image- and audio-processing to obtain a spectrum with more uniformly distributed energy, normally enhancing the contribution of higher frequencies. With this pre-processing step, we intended to remove the potentially distorting effect of equalisers, making that all the pitches across the selected range contribute equally in the HPCP calculation. This technique has been previously used by Gómez (2006a), and other authors have proposed similar solutions (Mauch & Dixon, 2010a; Müller & Ewert, 2010), as noted in Section 3.3. For our convenience, a spectral whitening function based on a method by Röbel & Rodet (2005) had been previously implemented in Essentia, taking full advantage of it.

Figure 6.7 shows the effect of applying a spectral whitening function prior to the HPCP calculation. The left column shows 36-bin raw chromagrams, whereas the right side illustrates the equivalent HPCP after a spectral whitening function. The audio content corresponds to the first four seconds of "Far from the Tree", by Bob Moses [5152629, deep house] (top) and Rektchordz's "No Dice" [842552, breaks] (bottom).

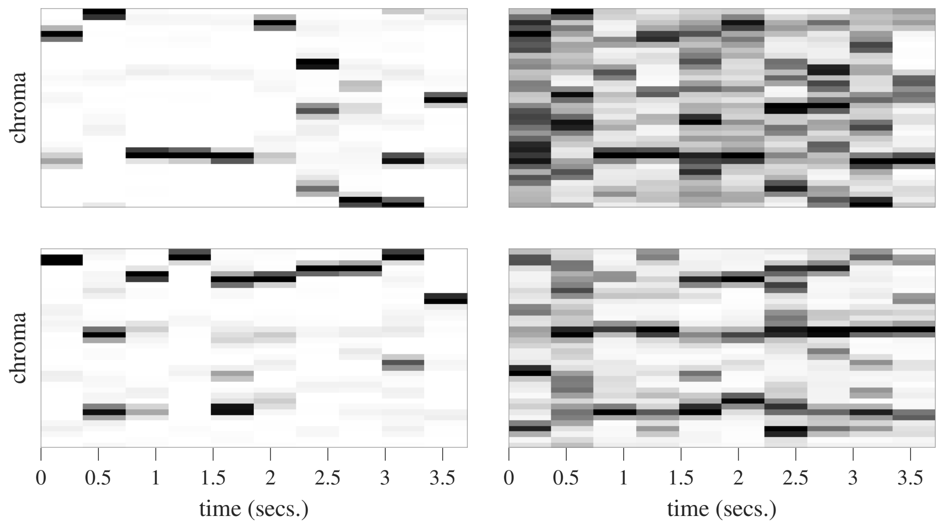


FIGURE 6.7: Effect of spectral whitening on HPCP calculation, over two excerpts of EDM tracks. The left column shows 36-bin raw chromagrams, whereas the right side illustrates the effect of spectral whitening on the same audio fragments.

6.3.5 Detuning Detection

As a last processing stage in our method, we inserted a simple function to detect tunings perceptually deviant from the standard pitch reference $\mathsf{a}_4 = 440$ Hz. Other authors have applied a tuning detection algorithm to detect the reference frequency of the analysed object (e.g. Peeters, 2006b; Dressler & Streich, 2007). However, we wanted to keep a fixed tuning reference —considered the standard in many musical contexts— and indicate lower and higher deviations from it (according to our annotation strategy described in Section 5.1), instead of dealing with potentially different references, since we considered this a more useful strategy for practical applications (e.g. harmonic mixing).

Our approach to detuning detection is a simplification of the method by Harte (2010), explained in Section 3.3.4, relying on an HPCP resolution of 3 bins per semitone. This allows to make corrections in the alignment of the main pitch-classes by rotating the chromagram $\pm 1/3$ semitone. Our system finds the highest peak in the averaged chromagram and shifts the spectrum ± 1 bin, depending on this unique position. This calculation is performed once per audio analysis, after the aggregation of all the chroma vectors. The motivation behind such simple approach is grounded upon the fact that all the key profiles discussed in Chapter 3 consistently present maxima in the tonic ($\hat{1}$) and/or the fifth degrees ($\hat{5}$), the two most prominent scale degrees independently from any modal configuration. Unlike other intervals, which show larger deviations from the harmonic spectrum, the equally-tempered fifth deviates in just 2 cents from the perfect fifth from the harmonic series —as present in just intonation or pythagorean tuning. Therefore, we have reasonable confidence to assume that the maximum peak of the averaged chromagram will normally represent either one or another. For the same reason, shifting the HPCP on a frame basis produces less satisfactory results, since it is after accumulation along time that energy concentrates more clearly in these bins.

Our algorithm labels detuned tracks with a caret (^) or an underscore (_) accompanying the tonic chroma (e.g. C_, G♯^). Although for evaluation purposes we disregard this additional information, it could be definitely useful in practical applications of key finding for harmonic mixing endeavours, where a difference of half a semitone would be disruptive enough, at least in terms of vertical mixing.

6.3.6 Evaluation of Spectral Whitening and Detuning Detection

Figure 6.8 shows the influence of the additional processing stages just described. The left bar of each barplot presents the estimation results of EDMA without any of these steps (just like in Figure 6.6). The other bars show the individual influence of the spectral whitening stage and the detuning detection function, as well as their operation in combination. The positive effect of spectral whitening is noticeable in all instances, with the exception of the pop music dataset assessed with the EDMM profiles. The detuning detection algorithm provides a timid improvement in the corpus of EDM, but its operation is best observed in the output of EDMA on popular music, with an increment of about 0.07 points over the simple version. According to Harte (2010), The Beatles recorded their *Help!* and *Please Please Me* albums almost entirely in a lower reference tuning, what could be behind this noticeable positive effect. Contrarily, the presence of excerpts with deviant tuning references in the EDM datasets is small, although the assumption of likely detuned fragments is not alien to EDM. At least, it must be certainly common during live DJ sets, as a consequence of tempo-matching operations, facilitated by the $\pm 8\%$ pitch/tempo control in professional vinyl-record players, which can produce pitch shifting effect of ± 1.5 semitones. In all scenarios, the combination of both processing steps provides yet another small improvement, roughly adding up the contributions of each separate stage, given that each processing function addresses clearly different problems.

A version of this method was submitted to the MIREX competition in 2016, obtaining the best score in the GiantSteps (GS^{K}) dataset, as shown in Table 4.5. However, the results on the EDM collection discussed in this section differ slightly from the ones reported in previous publications (Faraldo et al., 2016a,b), although, there is nothing

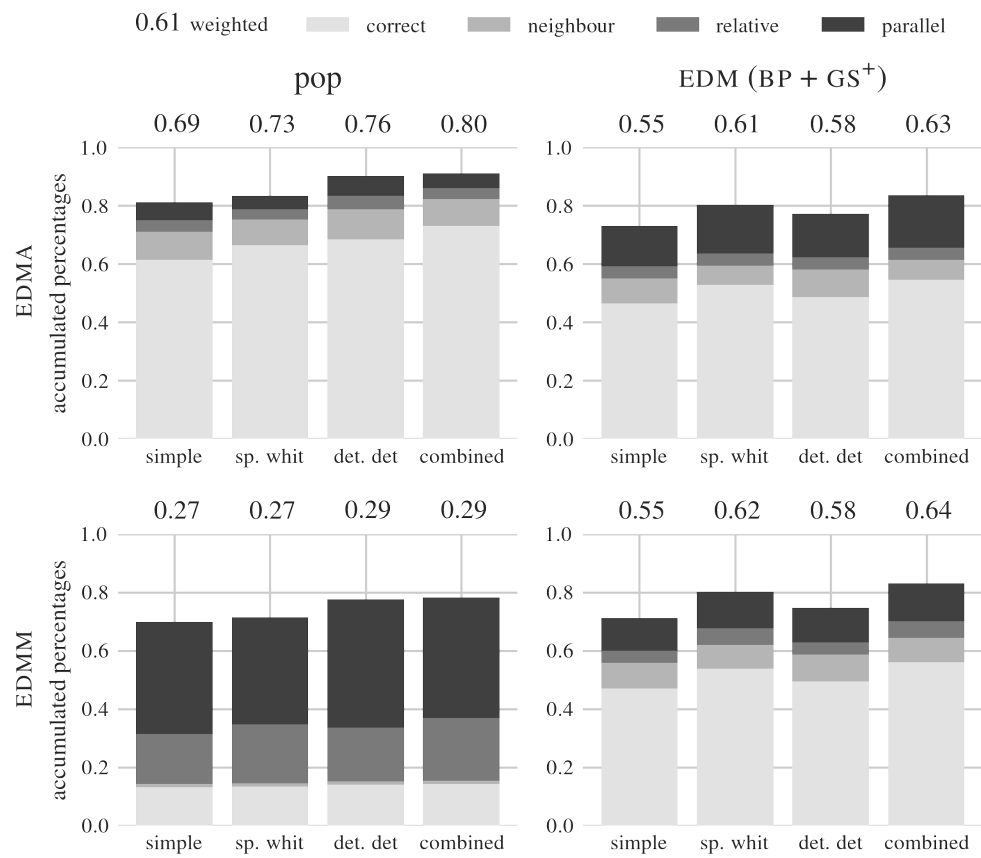


FIGURE 6.8: Evaluation of the effects of spectral whitening and detuning detection in the EDMA and EDMM profiles. The first bar of each plot shows the results of our method without any of these steps, as presented in the previous figure. The other bars show the effect of the two processing stages, plus their combination.

worrying about this divergence. Quite the opposite, the different results reflect the effect of some of the operations performed so far. First, the merging of BP and GS^{+} test collections neutralises the strong bias towards minor modalities present in the GiantSteps and KeyFinder datasets, used for evaluation in our previous paper, what naturally counteracts the positive effect that EDMM has in these other corpora. Moreover, the results between EDMA and EDMM are almost identical thanks to one of the decisions involved in our evaluation methodology, explained in Section 6.2. Since our evaluation method accepts double labels indicating modal ambiguity (F major | F minor), what we observe in these results is not a worsened performance of the EDMM model, but the positive judgement of tracks that are ambiguously annotated as major *and* minor. And although our key estimation method does not provide means to point at such modally ambiguous tracks, the evaluation procedure compensates for that,

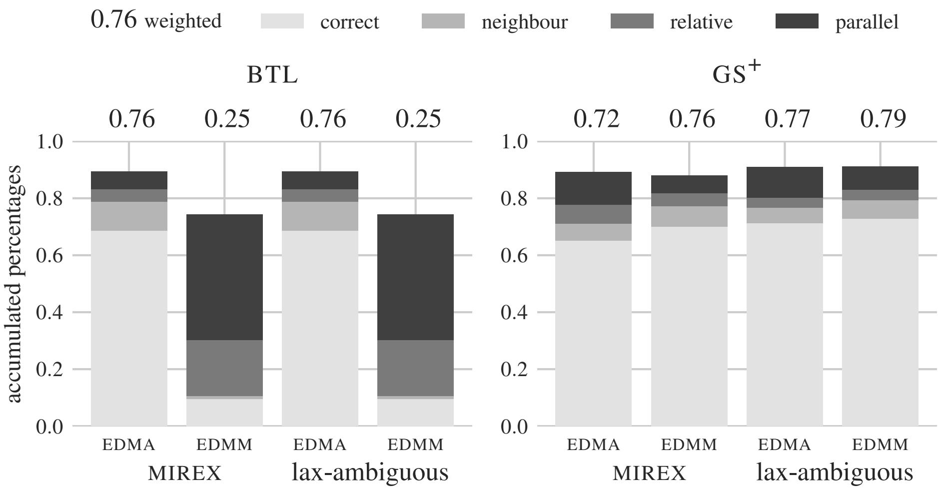


FIGURE 6.9: Evaluation of our method on BTL and GS$^+$ datasets, replicating the experiment presented in Faraldo et al. (2016a). Both datasets are evaluated according to the MIREX convention, along our laxer evaluation method ('lax-ambiguous'), which evaluates positively modally ambiguous tracks.

since our merged EDM corpora contain a relevant number of modally mixed tracks. On the other hand, the overall performance of our method is visibly lowered, given the much greater tonal complexity of the Beatport dataset.

We would like to close this section reproducing the experiment in Faraldo et al. (2016a), as a means to measure the influence of the recently proposed evaluation method and assess the effect of the additional annotations and corrections in the GS$^+$ dataset reported in Section 5.2.2.[103] Figure 6.9 shows the results of our improved method (with spectral whitening and detuning detection) with the EDMA and EDMM profiles on the Beatles dataset (left) and the revised GiantSteps collection (GS$^+$). We present evaluation results with the plain MIREX weighting system, as well as with the laxer evaluation method, labelled 'lax-ambiguous' in the figure ('other' and 'no-key' tracks are excluded from this evaluation, since the algorithm does not produce such labels). As expected, the performance of EDMM is poor in pop music, no matter the evaluation method, although the results for the EDM corpora improve, achieving a MIREX weighted score of 0.76 points. The MIREX 5% increment in the GS$^+$ dataset

[103]The original results published in Faraldo et al. (2016a) are summarised in the following table:

set	*profile*	correct	*fifth*	*relative*	*parallel*	*other*	weighted
BTL	EDMA	*.670*	.123	.050	.067	.089	*.760*
	EDMM	*.101*	.067	.195	.480	.156	*.288*
GS^K	EDMA	*.581*	.101	.066	.106	.146	*.673*
	EDMM	*.642*	.108	.033	.068	.149	*.720*

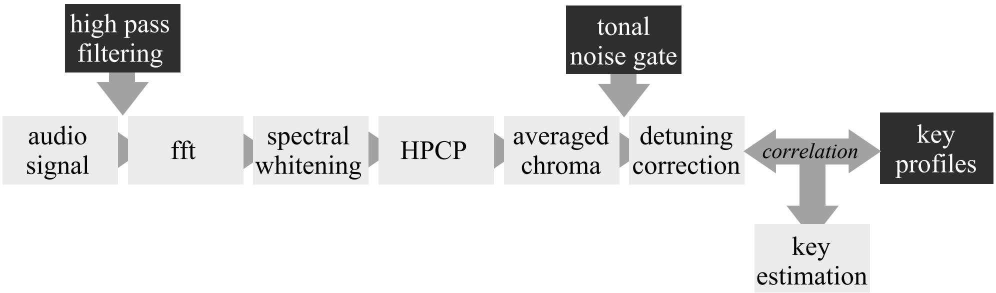


FIGURE 6.10: Processing pipeline of our key-finding algorithm with two additional steps. Besides the creation of new key profiles, we inserted a high-pass filter to attenuate low-frequency percussive components, and a gating function to discard tonal noise (darkened).

—compared to our published paper— is attributed to the revision of the annotations, totalling to 600 tracks with 63 corrected items and 88 ambiguous labels. Regarding the evaluation with our ambiguity-friendly evaluation method, as expected, the results with the Beatles dataset do not present any variability, since the data is annotated unambiguously as a single key. In contrast, the evaluation of the GS$^+$ improves ≈ 0.05 points with both profiles, due to the positive consideration of modally ambiguous tracks.

6.4 A Method Addressing Difficult Tracks

In a second publication (Faraldo et al., 2017), we wanted to address some of the shortcomings of our basic approach. In particular, we intended to solve the bias towards minor modalities introduced by the EDMM profiles, and obtain additional insights regarding ambiguously modal tracks. With this goal, we altered slightly the basic processing pipeline outlined in Figure 6.3, inserting a high-pass filter prior to the time-to-frequency conversion, and a chromagram gating function, in order to obtain profiles without tonal noise in modally irrelevant degrees. These new processing steps are shown in Figure 6.10. Additionally, we created new tonality profiles based on the Beatport dataset, which provided a relatively balanced distribution across major and minor keys. Furthermore, we attempted to obtain additional modal profiles, reflecting distributions with major/minor ambivalence and other ‘amodal’ configurations, with a clear tonic but without a explicit modal definition.

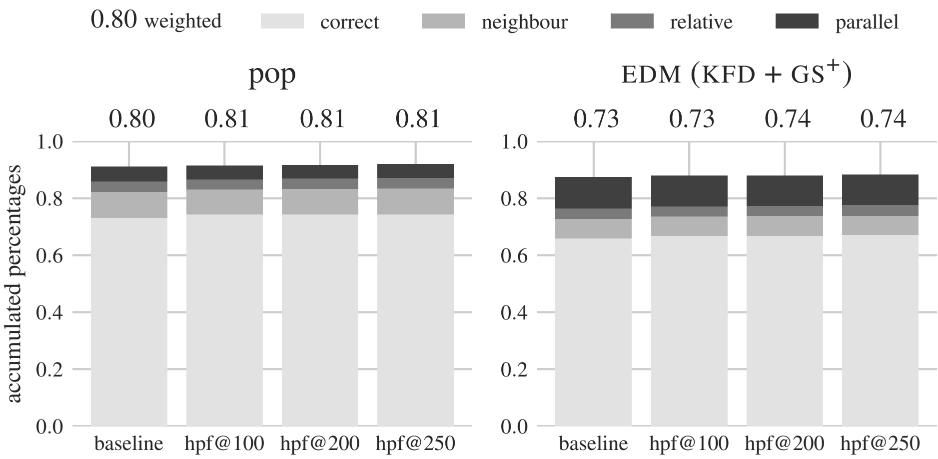


FIGURE 6.11: Effect of high-pass filtering with various cut-off frequencies, over the improved EDMA method described in the previous section.

6.4.1 High-Pass Filtering

As we have just noted, in our second published experiment we inserted a 3rd order IIR high-pass filter prior to the spectral transformation. The intention behind this step was to minimise the effect of bass-drum sounds and other low-frequency effects tuned differently from the key of the observed excerpt. Figure 6.11 shows the effect of this filtering stage in the improved algorithm presented in the previous section (EDMA with spectral whitening and detuning detection), with the filter's cut-off frequency set to 100, 200 and 250 Hz. A timid increment of nearly 1% is produced in the correctly classified instances when setting the cut-off frequency to 100 Hz —although this is not reflected in the weighted scores.[104]

In order to avoid possible overfitting effects, the evaluations throughout this section are carried on a test collection comprising the GS$^+$ and KFD datasets, excluding the Beatport dataset, used for key-profile extraction. The popular music aggregated data-set, on the other hand, remains identical. The baseline results for the remainder of this argumentation are provided by the leftmost bars of the plots in Figure 6.11, where the improved EDMA method recently described is used as a reference for further variations. As it can be observed, while these results align with those presented in Figure 6.8, the evaluation on EDM diverges slightly, given the modification of the evaluation test collection, and reflecting the positive effect of removing the challen-

[104]In our original publication, however, the cut-off frequency was set to 200 Hz. The difference between both sources is an effect of the narrative chosen for this chapter. Whereas in Faraldo et al. (2017) we experimented directly with the key profiles introduced in the next block, in this dissertation we favour a sequential narration, building our method upon the steps presented in previous sections.

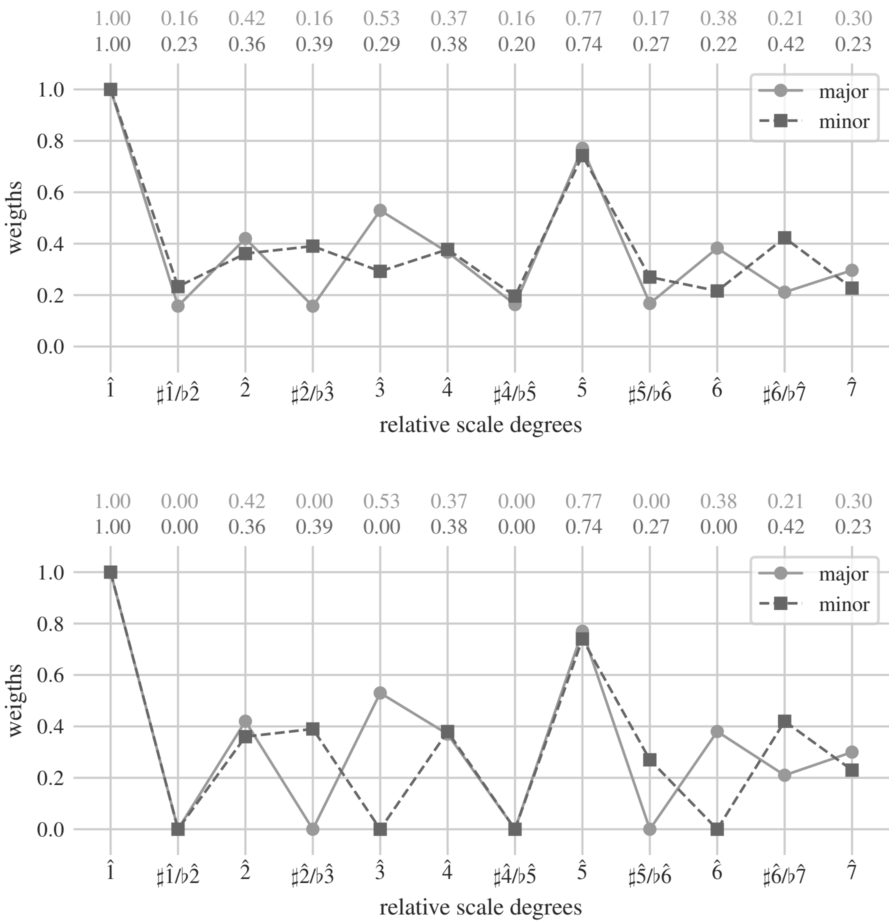


FIGURE 6.12: Major and minor key profiles obtained from a sub-collection of the BP dataset (BRAW), and with zeroed non-diatonic degrees (BGATE).

ging BP dataset, to which our method is not well prepared. In Section 6.5 we present a more thorough evaluation of our best-performing methods across all the datasets discussed from various musical styles, allowing a more mindful assessment of the performance of our algorithms compared to existing state of the art solutions.

6.4.2 BRAW and BGATE

With the intention to obtain more balanced profiles for major and minor modalities, we repeated the median profile extraction operation on a subset of BP. More precisely, we gathered a collection of 600 tracks —half major, half minor— with a confident level of annotation, which were correctly classified with other tonality profiles, including the Krumhansl & Kessler probe tone weightings, the modifications

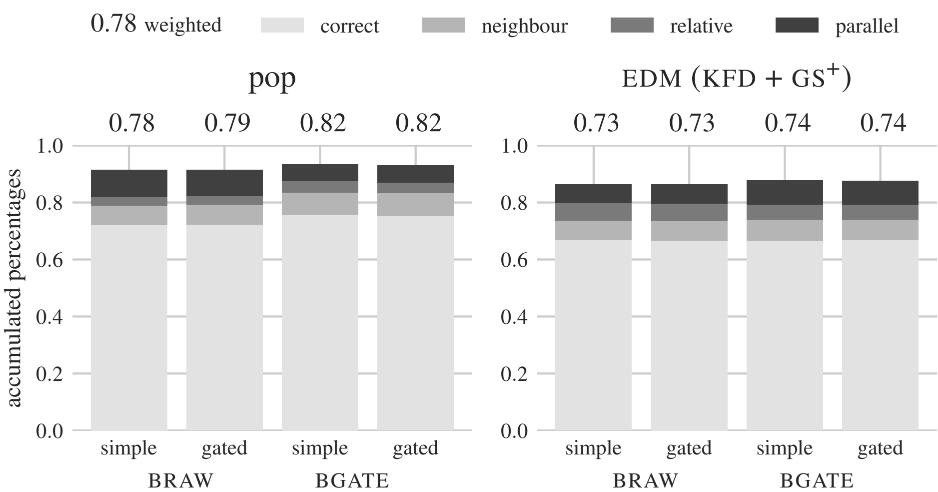


FIGURE 6.13: Evaluation of our method with the newly created BRAW and BGATE profiles, with and without a chromagram gating function.

introduced by Temperley (1999), and our own EDMA pair. The resulting two profiles, to which we refer as BRAW, are shown in Figure 6.12 (top). Since this new profile extraction operation involved 'controlled' audio files, with a confident performance across various profiles in the literature, theoretical tonal hierarchies manifest clearly in the major profiles, with almost a constant weight, just under 0.2, for all chromatic, non-tonal degrees (e. g. compare with the basic space by Lerdahl in Figure 3.2). On the contrary, the tonal hierarchy is far from evident in the minor BRAW profile, where apart from the tonic diad $\hat{1}$-$\hat{5}$, weights are distributed in a flatter way.

In order to compensate for the lack of difference in the minor profile, we derived a second distribution, by zeroing the weights of non-modal degrees, as 'suggested' by the major *braw* profile. This manual operation zeroed the $\flat\hat{2}$, $\flat\hat{3}$, $\sharp\hat{4}$ and $\flat\hat{6}$ degrees in major; and $\flat\hat{2}$, $\sharp\hat{4}$, $\natural\hat{6}$ in minor, obtaining the new BGATE key profiles, as illustrated at the bottom of Figure 6.12. The derivation process of the BGATE profiles, clearly insinuates an analogous operation in the chromagram calculation, where we inserted a HPCP thresholding function just before the detuning detection stage. This tonal 'noise gate' simply zeroes the bins with a total energy below a selected threshold in the averaged chromagram, ideally obtaining chromagrams closer to theoretical tonal hierarchies. We set the initial threshold value to 0.2, according to the weights of the chromatic degrees in the major BRAW profile.

Figure 6.13 shows the evaluation results with these newly created profiles, with and without the tonal gating function, with a threshold set to 0.2. Regarding pop music, the EDMA profiles seem to work just as good as the new profiles, which only offer

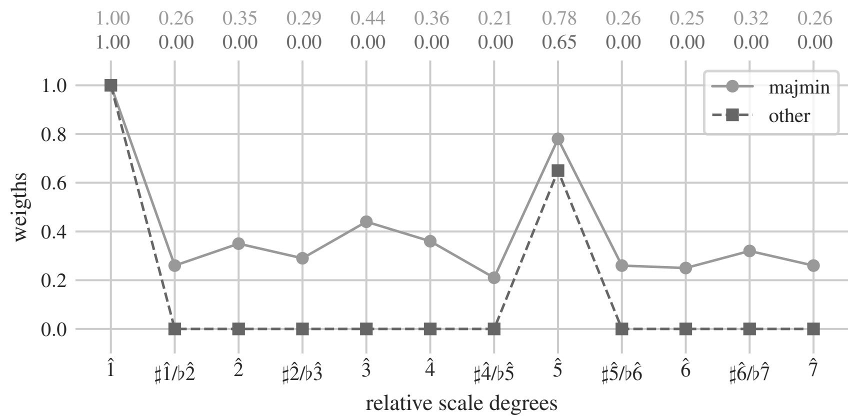


FIGURE 6.14: Additional profiles accounting for tonically clear but modally ambiguous distributions.

a small improvement when using the BGATE profiles without the tonal noise gate. Similarly, the improvement on the electronic dance music data is only visible with the BGATE profiles, with a comparable performance between the raw aggregated chromagram and the gated one. To assess the performance of the gated chromagram we experimented with various other thresholds (0.2, 0.25, 0.3, 0.4 and 0.5), obtaining progressively lower scores, concluding that the chromagram gating function provides at best a neutral effect. Therefore, in subsequent evaluations we prescind from this additional processing stage.

6.4.3 Additional Profiles for Ambiguous EDM Tracks

In addition to the two newly created profiles, in order to minimise parallel errors, we obtained a third profile from a group of difficult minor tracks wrongly estimated as major with the BGATE profiles, (Faraldo et al., 2017). This additional profile is shown in Figure 6.14 ('majmin'). However, as we have seen in Chapter 5, these more difficult tracks are most likely modally ambiguous tracks (and not merely items in minor), which present a clear tonic but a varying degree of openness regarding their principal modal sign. Accordingly, in this work we have labelled these tracks with the 'majmin' string (e.g. A minor | A major), highlighting the modal ambiguity of these tracks in line with our annotation and evaluation methodologies. Similarly, we added a simpler profile with energy concentrating on the first and fifth degrees, leaving all other chromas neutrally at zero (Figure 6.14, 'other'). With this fourth profile we intend to detect tracks that do not convey a modality at all, likely corresponding to drone monotonic fragments. As one last step, we heuristically set a confidence estimation threshold value of 0.5, below which tracks are assigned the 'no-key' label.

Figure 6.15 shows the results of analysing our test collections with the new profile additions. From left to right, bars in each plot represent an accumulation of (a) the binary method (BGATE), with (b) one additional profile to detect major/minor ambiguous tracks (majmin), (c) plus the extra monotonic profile to discover amodal tracks (other), and last, (d) a correlation confidence threshold to produce 'no-key' labels. In the following paragraphs, we refer to this four-profile variation as the BGATE$^+$ method.

The pop music test collection seems unaffected by the addition of new modal labels, what might at least indicate that the newly introduced profiles do no produce a negative effect in musics with a clear modal definition. This can be better seen in the modal confusion matrix of Figure 6.16, where the four possible single modal labels 'major' (I), 'minor' (i), 'other' (1) and 'no-key' (X) are measured across all possible estimations. The matrix shows that the BGATE$^+$ produced only two 'no-key' estimations. These correspond to "Nuthin' but a G Thang" by Dr. Dre and Snoop Dog (from the RS dataset), —perhaps due to its predominantly spoken rap texture— and The Police's "Don't Stand so Close to Me" (from the BB collection), possibly as an effect of the alternating semitone modulation between verse and chorus (E$\flat$ minor $\leftrightarrow$ D major) in the final aggregated chromagram. Regarding modally undefined tracks, the estimated errors seem to vaguely point to rap-oriented songs, such as "Brass Monkey" by the Beastie Boys (BB) or Eminem's "Lose Yourself", from the BB and RS datasets, respectively. Other errors could be produced by tracks with little harmonic change, or with melodies insistently centred on a single tone, like some forms of —especially early— rock'n'roll (e.g. "Born to Cry", by Dion & The Belmonts).

With EDM collections, on the other hand, the addition of new labels progressively lowers the performance of the algorithm. This, again, can only be timidly seen in Figure 6.15, although it is more clear in the relative confusion matrix in Figure 6.17. We have incorporated the Beatport dataset in these visualisation, since most of the difficult tracks with alternative labelling belong to this collection. This relative confusion matrix shows how the new labels mostly fail when attempting to classify existing tracks. For example, our method only managed to correctly assign a 'no-key' label to five tracks,[105] whereas all other atonal labels are wrongly assigned to tonical classes. Regarding the 'other' modal variants, only 13 items have been correctly placed compared to the 44 taken as minor, and 4 estimated as atonical tracks. This, at least, suggests paths to continuing this work in various ways. On the one hand, our methodology with simple additional profiles and atonal confidence threshold seems insufficient to address the degree of tonal complexity presented by many EDM samples. A possible solution would be to extract profiles from shorter

[105] With Beatport unique track identifiers 4372159, 298989, 765583, 923844 and 3400782.

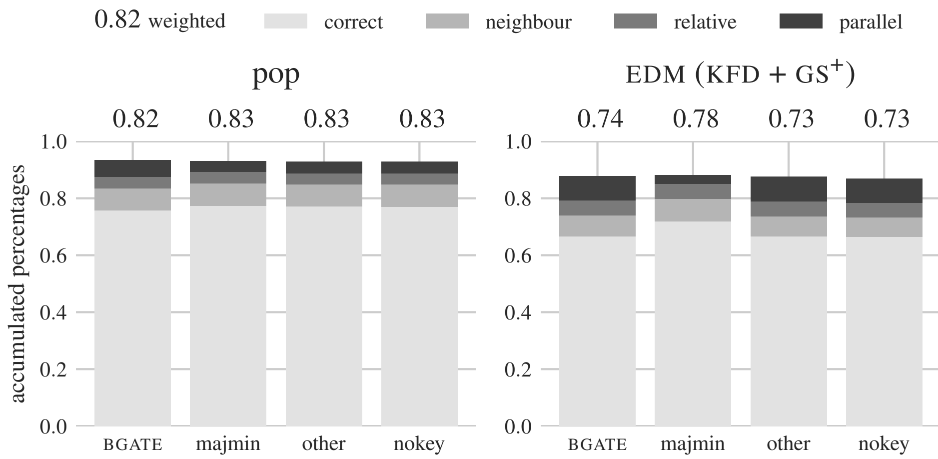


FIGURE 6.15: Evaluation of the BGATE profiles (with high-pass filter and without tonal gate) with up to four profiles: (a) BGATE uses only the major and minor profiles; (b) 'majmin' adds an additional profile accounting for modally ambiguous tracks; (c) 'other' incorporates the simple monotonic profile, and (d), 'nokey', additionally reports atonical tracks when the profile-correlation scores under 0.5.

fragments, aligned with hypermetrical units, what could in turn provide an idea of tonal change besides the global estimation. We believe that this approach could reduce the negative effect of chroma aggregation, which seems to create profiles with tonal noise in at least some of the more difficult tracks. On the other hand, the Beatport dataset might need further inspection of items with 'no-key' and 'other' labels, perhaps establishing different tonal subclasses. This could be achieved by a division based on additional modal details (e.g. monotonic, locrian, whole-tone, et cetera), and perhaps also based on subgenre differentiation, in order to define clearer groups of tonal behaviour within EDM tracks.

6.5 Final Evaluation

In this last section, we present the results of several algorithms on the various datasets used throughout this dissertation, including euroclassical music (*The Well-Tempered Clavier*, WTC), Western popular music —comprising the Billboard (BB), Beatles (BTL) and Rolling Stones Magazine (RS) collections— as well as electronic dance music, including the KeyFinder annotations (KFD), plus our two recently created corpora: the GiantSteps+ (GS$^+$) and Beatport (BP) datasets. The algorithms reported include commercial solutions tailored to EDM, such as KeyFinder (KFA), Mixed-In-Key (MIK), and Traktor (TK), together with two published algorithms, namely, the methods by Gómez (2006a) and Noland & Sandler (2007), implemented as an

	I	♭II	II	♭III	III	IV	♭V	V	♭VI	VI	♭VII	VII	i	♭ii	ii	♭iii	iii	iv	♭v	v	♭vi	vi	♭vii	vii	1	♭2	2	♭3	3	4	♯4	5	♭6	6	♭7	7	X
X																																					
1																																					
i	3			14		4			2		5	1	117	1	1			5		3				3	3											1	1
I	657	4	3	2		20	3	50	3		4	10	35	1	4		2	5	1	4		25		5						1							1

FIGURE 6.16: Relative key confusion matrices of BGATE+ with all the combined pop datasets. It is worth noting that all the annotations are labelled as single major or minor tracks.

	I	♭II	II	♭III	III	IV	♭V	V	♭VI	VI	♭VII	VII	i	♭ii	ii	♭iii	iii	iv	♭v	v	♭vi	vi	♭vii	vii	1	♭2	2	♭3	3	4	♯4	5	♭6	6	♭7	7	X
X	13																								29												5
1	3							1					44					2		2	1				13	1		1	1	1							4
i	45	5	4	136	2	32	2	13	41		39	1	1397	13	12	1	8	88	3	79	6	4	6	27	193	4	5	8	1	9	6	11	8	1	3	17	39
I	481	3	2	3	1	25	1	37		1	1	3	23	4	7		8	5	1	5		15		4	36	1		2	7	2		3				8	5

FIGURE 6.17: Relative key confusion matrices of BGATE+ with all the combined EDM datasets. Recall that KFD provides single major/minor annotations, whereas BP and GS+ provide an extended vocabulary with ambiguous 'majmin' and ambivalent bimodal tracks, plus the 'other' and 'no-key' labels.

Essentia key extractor (ES) and the QM Key Detector (QM), respectively. All of these systems have already been introduced in Section 4.3, where we offered a preliminary evaluation, and are shown here to facilitate comparability with two of our algorithms, the binary classifier BGATE^2 and the multi-profile method BGATE$^+$, which provides additional modal verbose and 'no-key' labels, as discussed in the previous block. From all other methods tested, only MIK produces labels for atonal tracks, although QM writes an additional tag 'unknown', when it can not confidently estimate the key of the sound file.

Table 6.2 shows the evaluation results of all the described algorithms on the available test collections annotated with a single global key (WTC, BB, BTL, RS, KFD). Since these annotations and the estimations from most tested algorithms are restricted to a binary single-key output (with the exception of BGATE$^+$ and, to a lesser extent, MIK), the different evaluation strategies discussed should provide identical results. Therefore, the laxer evaluation method used in this table —assessing positively modal ambiguity and tonical ambivalence— will only have a visible effect in the results of BGATE$^+$. It is for this reason that we also evaluate this particular method with a stricter MIREX protocol (shown in a smaller font size), which imposes a penalty to tracks ambiguously labelled or with additional modal tags. Our newly created datasets (BP and GS$^+$) are assessed independently in Tables 6.4 and 6.5, under slightly diverging evaluation methodologies, including the stricter binary MIREX evaluation system just mentioned, and our laxer strategy (described in Section 6.2), with both restricted and extended modal vocabularies, as it will be explained shortly. Besides correctly classified items, all the tables in this section report the percentage of correctly identified tonic and mode, as well as typical regional errors.

A first observation stemming from Table 6.2, is that our two methods provide the highest marks in all datasets except for KeyFinder (KFD). This is particularly noticeable in the WTC dataset, where BGATE$^+$ presents an identical performance to BGATE^2, without producing any relative or parallel errors, obtaining the highest weighted scores. The performance of Essentia (ES) is comparably just as good, with a higher percentage of correctly classified items. The main difference between ES and our methods seems to lay in the ratio of non-neighbour errors (other), where our algorithms minimise the amount of Essentia's errors, at the cost of producing a few more fifth mislabels.

The popular music datasets present lower scores in general, likely due to the higher timbral complexity of these musics. This is especially visible in the collection from The Beatles (BTL), where algorithms show a drop of about a 10% compared to the other two popular music datasets. The only exception to this pattern is provided by the QM Key Detector, which behaves exactly in the opposite tendency. In any case,

		correct items			*typical errors*				
set	*method*	*tonic*	*mode*	key	*fifth*	*relative*	*parallel*	*other*	score
WTC	BGATE^2	.8541	.9791	*.8541*	.1250	.0	.0	.0208	**.9167**
	BGATE$^+$	.8541	.9791	*.8541*	.1250	.0	.0	.0208	**.9167**
	(BGATE$^+$ strict)	.8541	.9791	*.8541*	.1250	.0	.0	.0208	**.9167**
	ES	.8646	.9271	**.8646**	.0625	.0208	.0	.0521	*.9021*
	QM	.7812	.8958	*.7812*	.0625	.0521	.0	.1042	*.8281*
	KFA	.5000	.7604	*.5000*	.2604	.1146	.0	.1250	*.6646*
	MIK	.8125	.8958	*.8021*	.0937	.0625	.0104	.0312	*.8698*
	TK	.8125	.8542	*.8125*	.0417	.1250	.0	.0208	*.8708*
BB	BGATE^2	.8256	.8688	**.7712**	.0784	.0416	.0544	.0544	**.8338**
	BGATE$^+$	.8224	.8784	**.7808**	.0816	.0416	.0146	.0544	**.8424**
	(BGATE$^+$ strict)	.8224	.7936	*.7024*	.0752	.0416	.1200	.0608	*.7764*
	ES	.6704	.8576	*.6416*	.1488	.0288	.0288	.1520	*.7304*
	QM	.6448	.7952	*.5696*	.1120	.0528	.0752	.1904	*.6565*
	KFA	.6480	.6416	*.5488*	.0832	.1760	.0992	.0928	*.6630*
	MIK	.7936	.7840	*.7072*	.0656	.7040	.8640	.7040	*.7784*
	TK	.7680	.7456	*.6688*	.0640	.0832	.0992	.0848	*.7456*
BTL	BGATE^2	.7542	.8603	**.6816**	.1061	.0503	.0726	.0894	**.7642**
	BGATE$^+$	.7542	.8770	**.7039**	.1006	.0503	.0503	.0949	**.7793**
	(BGATE$^+$ strict)	.7542	.7765	*.6033*	.1006	.0503	.1508	.0949	*.6989*
	ES	.5530	.8659	*.5307*	.1061	.0447	.2235	.2961	*.6017*
	QM	.6816	.8044	*.6145*	.1117	.0335	.0670	.1732	*.6938*
	KFA	.5754	.6201	*.4693*	.1117	.1899	.1061	.1229	*.6033*
	MIK	.7542	.7542	*.6425*	.0726	.0838	.1117	.0894	*.7263*
	TK	.6871	.6704	*.5642*	.0670	.1117	.1323	.1341	*.6559*
RS	BGATE^2	.8400	.8900	**.7800**	.0550	.0200	.0600	.0850	**.8255**
	BGATE$^+$	.8300	.9000	**.8000**	.0450	.0200	.0300	.1050	**.8345**
	(BGATE$^+$ strict)	.8300	.7450	*.6500*	.0400	.0200	.1800.	.1100	*.7120*
	ES	.5800	.8100	*.5100*	.2100	.0350	.0700	.1750	*.6395*
	QM	.6350	.7650	*.5400*	.1100	.0400	.0950	.2150	*.6260*
	KFA	.6750	.6550	*.5350*	.0750	.1100	.1400	.1400	*.6335*
	MIK	.8050	.7950	*.7050*	.0450	.0550	.1000	.0950	*.7640*
	TK	.7450	.7450	*.6450*	.0600	.0600	.1000	.1350	*.7130*
KFD	BGATE^2	.7024	.7725	*.6202*	.0902	.0581	.0821	.1493	*.6992*
	BGATE$^+$	.7014	.7525	*.6142*	.0872	.0581	.0872	.1533	*.6927*
	(BGATE$^+$ strict)	.7014	.7304	*.5932*	.0851	.0571	.0821	.1563	*.6745*
	ES	.3878	.6462	*.3377*	.1794	.0591	.0501	.3737	*.4551*
	QM	.3767	.5792	*.3387*	.1413	.1293	.0381	.3527	*.4557*
	KFA	.7084	.8737	**.6663**	.1062	.0341	.0421	.1513	**.7381**
	MIK	.7575	.8677	**.7054**	.0802	.0331	.0521	.1293	**.7658**
	TK	.7064	.8206	*.6543*	.0912	.0541	.0521	.1483	*.7265*

TABLE 6.2: Comparative results of our methods, BGATE^2 (with binary output) and BGATE$^+$ (with multiple labels), along state of the art algorithms in test collections annotated with a binary single-key. These comprise euroclassical music (WTC), pop (BB, BTL, RS) and EDM (KFD). In tiny font, we show the results of BGATE$^+$ with a strict, single-key and unambiguous evaluation method.

name	*abr.*	*modal vocabulary*	*ambiguous multi-labelling*
all-ambiguous	AA	{major, minor, other, no-key}	✓
all-single	AS	{major, minor, other, no-key}	✗
binary-ambiguous	BA	{major, minor}	✓
binary-single	BS	{major, minor}	✗

TABLE 6.3: Evaluation criteria used in our assessment of the Beatport and GiantSteps+ datasets. They result from the combination of two modal vocabularies, and the recognition —or not— of multiple annotations denoting modal ambiguity and/or tonical ambivalence.

our methods outperform all other algorithms in the three popular music corpora. Our binary system BGATE^2 presents correct key estimates ranging from 68.2% in BTL to 78% in the RS dataset, with global scores of 0.76 and 0.825 points, respectively. Our four-profile method performs even better, with 78% (BB), 70% (BTL), and 80% (RS) correctly estimated keys. The good performance of BGATE$^+$ echoes the claims by Temperley & De Clercq (2013), calling for annotation —and evaluation— methods tailored to the modal idyosincracies of popular music. Moreover, it indicates that our perhaps excessive care in addressing modal and tonical ambiguity does indeed produce a positive effect in musics where these practises are well documented. As a matter of fact, the ambiguous 'majmin' profile mostly operates in rock'n'roll samples, presenting the types of modal ambiguity described in Section 2.3. Furthermore, the visible drop of BGATE$^+$ when evaluated with a stricter MIREX method (with down to 15% less correctly classified keys in RS), seems to corroborate the adequate behaviour of our ambiguous labelling and evaluation system. On the other hand, the lower results of our algorithms in KFD, might suggest that these types of ambiguity occur in electronic dance music only as part of a larger battery of tonal practises, not just reducible to major/minor modal ambiguity. In this latter music collection, MIK clearly provides the best results, with 70% correctly classified instances, followed by the KeyFinder application (KFA), likely as a consequence of overfitting the data.[106] In this collection, none of our methods manages to classify more than 62% of the items correctly (BGATE^2). It is interesting to note, however, that the stricter evaluation method on BGATE$^+$, only scores $\approx 2\%$ below the laxer assessment (compared to the 8–15% difference in popular music), suggesting that modally ambiguous tracks might be annotated as 'major' in the KeyFinder collection.

As advanced earlier, Tables 6.4 and 6.5 show the results of the seven discussed algorithms on the Beatport and GiantSteps+ datasets, respectively. Since these corpora present important challenges to the various algorithms, we offer four different evalua-

[106]We remind the reader that Sha'ath gathered the KeyFinder dataset (KFD) in order to improve his key estimation application KeyFinder (KFA).

		correct items			typical errors				
set	*eval*	*tonic*	*mode*	key	*fifth*	*relative*	*parallel*	*other*	score
BGATE^2	AA	.6871	.6144	*.5013*	.0781	.0505	.1857	.1843	*.5926*
	AS	.6750	.5902	*.4778*	.0768	.0599	.1972	.1884	*.5735*
	BA	.7228	.6541	**.5337**	.0831	.0537	.1891	.1401	**.6291**
	BS	.7099	.6282	*.5086*	.0817	.0637	.2013	.1447	*.6088*
BGATE$^+$	AA	.6810	.6669	*.5606*	.0801	.0464	.1205	.1925	*.6386*
	AS	.6669	.5787	*.4724*	.0727	.0505	.1945	.2100	*.5628*
	BA	.7351	.8330	**.7010**	.1005	.0588	**.0341**	.1056	**.7757**
	BS	.7189	.7214	*.5911*	.0911	.0639	.1278	.1261	*.6814*
ES	AA	.3627	.5962	*.2692*	.1790	.0417	.0935	.4166	*.3899*
	AS	.3553	.5787	*.2564*	.1729	.0478	.0989	.4240	*.3770*
	BA	.3847	.6347	**.2865**	.1905	.0444	.0981	.3804	**.4148**
	BS	.3768	.6160	*.2729*	.1841	.0509	.1039	.3883	*.4010*
QM	AA	.4556	.6050	*.3721*	.1507	.0612	.0834	.3324	*.4826*
	AS	.4462	.5868	*.3580*	.1433	.0666	.0882	.3439	*.4673*
	BA	.4814	.6440	**.3961**	.1605	.0652	.0852	.2930	**.5130**
	BS	.4713	.6246	*.3811*	.1526	.0709	.0903	.3052	*.4967*
KFA	AA	.6366	.6676	*.5047*	.1083	.0525	.1319	.2026	*.6010*
	AS	.6258	.6480	*.4852*	.1063	.0606	.1406	.2073	*.5847*
	BA	.6700	.7101	**.5369**	.1152	.0558	.1331	.1589	**.6379**
	BS	.6590	.6898	*.5165*	.1132	.0645	.1426	.1633	*.6209*
MIK	AA	.7046	.7140	*.5915*	.0794	.0511	.1131	.1649	*.6692*
	AS	.6965	.6945	*.5734*	.0774	.0565	.1231	.1696	*.6536*
	BA	.7437	.7609	**.6303**	.0847	.0546	.1134	.1170	**.7117**
	BS	.7356	.7407	*.6114*	.0826	.0603	.1243	.1214	*.6956*
TK	AA	.6164	.6696	*.5074*	.1063	.0794	.1090	.1978	*.6062*
	AS	.6023	.6460	*.4865*	.1023	.0875	.1157	.2079	*.5871*
	BA	.6478	.7122	**.5397**	.1131	.0845	.1081	.1546	**.6432**
	BS	.6347	.6877	*.5179*	.1089	.0931	.1168	.1633	*.6236*

TABLE 6.4: Results of our methods BGATE^2 and BGATE$^+$, along state of the art systems in the Beatport (BP) dataset. We present four different evaluation results, accounting for different vocabularies and consideration of tonal ambiguity: all-ambiguous (AA), all-single (AS), binary-ambiguous (BA), and binary-single (BS).

tion variations, as summarised in Table 6.3. First, we consider all the key labels as they are annotated, including atonical tracks and fragments annotated as ‘other’, as much as ambiguously tagged items (i.e. with two labels indicating either tonical ambivalence or modal ambiguity). This ‘all-ambiguous’ evaluation is indicated as AA in both tables. Secondly, we evaluate the methods considering all possible labels but penalising ambiguously annotated tracks, indicated as ‘all-single’ (AS). Given that all

		correct items			*typical errors*				
set	*eval*	*tonic*	*mode*	key	*fifth*	*relative*	*parallel*	*other*	score
BGATE^2	AA	.8333	.8067	*.7450*	.0433	.0433	.0883	.0800	*.7973*
	AS	.7967	.7383	*.6633*	.0567	.0583	.1333	.0883	*.7358*
	BA	.8368	.8491	**.7842**	.0456	.0456	.0526	.0719	**.8312**
	BS	.7996	.7786	*.6995*	.0598	.0615	.1002	.0791	*.7678*
BGATE$^+$	AA	.8283	.8000	*.7467*	.0400	.0400	.0817	.0917	*.7950*
	AS	.7900	.7083	*.6417*	.0533	.0533	.1483	.1033	*.7140*
	BA	.8533	.8971	**.8381**	.0457	.0457	**.0152**	.0552	**.8777**
	BS	.8152	.7924	*.7181*	.0610	.0610	.0971	.0629	*.7863*
ES	AA	.4733	.6650	*.3983*	.1700	.0767	.0750	.2800	*.5213*
	AS	.4533	.6117	*.3533*	.1583	.0850	.1000	.3033	*.4780*
	BA	.4842	.7000	**.4193**	.1789	.0807	.0649	.2561	**.5460**
	BS	.4640	.6450	*.3726*	.1670	.0896	.0914	.2794	*.5012*
QM	AA	.5117	.6333	*.4683*	.1117	.1250	.0433	.2517	*.5703*
	AS	.4917	.5700	*.4133*	.1033	.1367	.0783	.2683	*.5217*
	BA	.5237	.6678	**.4938**	.1178	.1318	.0299	.2267	**.5982**
	BS	.5026	.6011	*.4359*	.1090	.1441	.0668	.2443	*.5469*
KFA	AA	.7900	.8500	*.7250*	.0917	.0317	.0650	.0867	*.7933*
	AS	.7583	.7800	*.6517*	.0933	.0500	.1067	.0983	*.7347*
	BA	.7982	.8947	**.7632**	.0965	.0333	.0351	.0719	**.8284**
	BS	.7645	.8225	*.6872*	.0984	.0527	.0773	.0844	*.7677*
MIK	AA	.8567	.8800	*.8033*	.0550	.0200	.0533	.0683	*.8475*
	AS	.8250	.8000	*.7167*	.0617	.0383	.1083	.0750	*.7807*
	BA	.8664	.9279	**.8471**	.0580	.0211	.0193	.0545	**.8863**
	BS	.8327	.8451	*.7570*	.0651	.0405	.0757	.0616	*.8169*
TK	AA	.7950	.8117	*.7350*	.0567	.0567	.0600	.0917	*.7923*
	AS	.7633	.7383	*.6600*	.0567	.0717	.1033	.1083	*.7305*
	BA	.8053	.8544	**.7737**	.0596	.0596	.0316	.0754	**.8277**
	BS	.7715	.7786	*.6960*	.0598	.0756	.0756	.0931	*.7636*

TABLE 6.5: Comparative results of our two methods BGATE^2 and BGATE$^+$, along state of the art systems in the GiantSteps+ (GS$^+$) dataset. We include four different evaluation results, accounting for different vocabularies and inclination towards tonal ambiguity: all-ambiguous (AA), all-single (AS), binary-ambiguous (BA), and binary-single (BS).

the algorithms except for BGATE$^+$ use a simpler binary dictionary (major, minor), we present correspondingly reduced assessments, filtering out entries labelled as ‘other’ or ‘no-key’ in our corpora (which add to 90 excerpts in BP and 29 tracks in GS$^+$). These two additional evaluations —with and without multi-label consideration— are indicated as ‘binary-ambiguous’ (BA) and ‘binary-single’ (BS), respectively.

From the four evaluation modes just described, the 'binary-ambiguous' vocabulary (BA) presents the highest scores in both datasets for all the algorithms assessed. This conforms to our predicted behaviour, since this evaluation scenario filters out difficult tracks with additional labels, assessing positively the modal ambiguity implied by annotations such as, for example, "G minor | G major". This is particularly noticeable in the reduction of parallel errors with BGATE+ in both datasets (highlighted in bold font in the tables). However, other evaluation methods follow different tendencies, depending on the dataset under consideration. Whereas in the Beatport collection, the restricted 'binary-single' evaluation method (comparable to the MIREX protocol) provides better results than with larger modal dictionaries, the GiantSteps+ dataset receives more favourably the 'all-ambiguous' evaluation strategy. A plausible explanation to this might be found in the fact that the Beatport dataset has been analysed without the amount of detail devoted to GS+, although it presents a higher degree of tonal complexity (probably favoured by the randomised recollection process described in Section 5.2.1). This is probably behind the lower performance in BP when assessing extended modal vocabularies, suggesting that this corpus could benefit from a new revision of its items annotated with lower degrees of confidence (see Table 5.2). In any case, the overall considerably lower results compared to the GiantSteps+ collection, seems to confirm that the Beatport corpus constitutes a challenging collection for tonality estimation in EDM, as supported by the figures in Table 6.4, where our BGATE+ method obtains the first position with a score of 0.77, followed by MIK with 0.71 points in the 'binary-ambiguous' evaluation. Regarding the extended modal vocabulary 'all-ambiguous', the results are inverted, and BGATE+ scores ≈ 0.64 points, slightly behind the mark of 0.67 by Mixed-In-Key.

On the other hand, the GiantSteps dataset is clearly biased towards minor modalities, as discussed in Sections 4.1.3 and 5.2.2, presenting tonal tracks corrected in user fora, and generally containing less alternative modal labels. These facts are likely behind the performance boost in all scenarios (compared to the Beatport dataset), achieving scores as high as 0.886 and 0.877 by MIK and BGATE+, respectively, around 10% higher than in the Beatport dataset with our 'binary-ambiguous' evaluation method. The additional modal labels, however, diminish the performance of BGATE+ down to 0.79 (all-ambiguous) and 0.71 (all-single), an effect that is less perceivable in all the other solutions tested.

To conclude, Figure 6.18 summarises the results discussed, organised and aggregated by musical style. Although providing good results on euroclassical and popular musics, the methods published by Gómez (2006b) (ES) and Noland & Sandler (2007) (QM) present clear difficulties in EDM, independently from the evaluation vocabulary chosen. In contrast, all other methods under evaluation, presumably tailored to EDM, provide fair results across the various musical repertoires considered. As discussed in

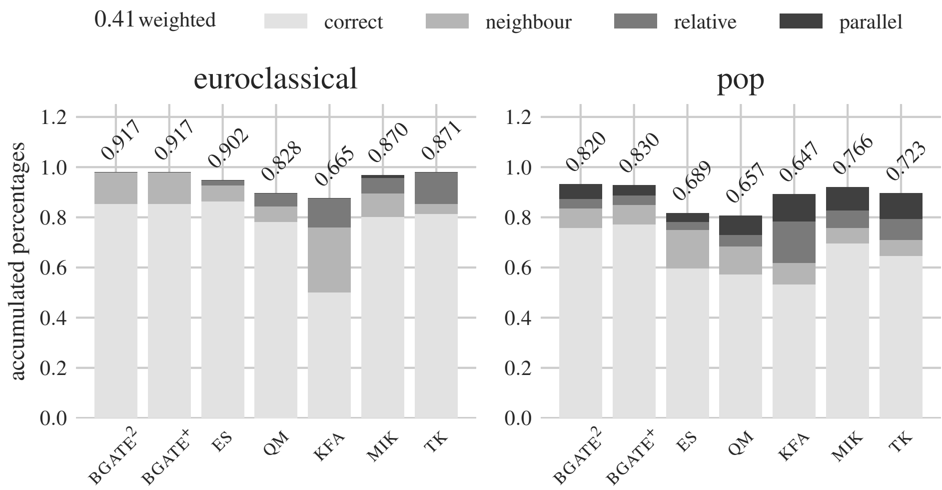


FIGURE 6.18: Overall results in three different musical styles. The euroclassical plot refers to the WTC dataset, whereas the results for pop include BB, BTL and RS. Regarding EDM, the four different levels of evaluation discussed in the section are provided for the aggregation of BP, GS$^+$ and KFD.

previous paragraphs, our method outperforms all other algorithms in euroclassical and popular music repertoires. Regarding EDM, on the other hand, the evaluation results seem to be highly dependent on the evaluation methodology chosen. Our BGATE$^+$ system performs comparably to the state of the art when assessed with ambiguous labels, obtaining the second position with the 'all-ambiguous' convention (0.689 vs. 0.74 by MIK), and the highest overall score with the 'binary-ambiguous' evaluation (0.79 points). This aligns with our efforts to develop annotation and estimation strategies accounting for more varied modal practises. Similarly, the restricted and unambiguous binary vocabulary evaluation ('binary-single') places BGATE$^+$ in the second rank, with 0.72 points, just below Mixed-In-Key. Last, the 'all-single' assessment situates our methods behind the other solutions tailored to electronic dance music, including KeyFinder and Traktor, likely due to the enlarged vocabulary of the BGATE$^+$ algorithm. However, to end with a positive note, the additional verbose provided by this method could easily illuminate future work in this area, by helping detecting difficult tracks and revising unconfident annotations.

***

In this chapter, we have discussed our variations on template matching automatic key-finding algorithms, attempting to improve the performance in the specific domain of electronic dance music. As we have tried to highlight in previous chapters, this meta-genre imposes specific challenges, from the choice of signal processing parameters to the acknowledgement of its particular musical characteristics, what becomes even more apparent after the evaluation results presented in this chapter. Throughout the chapter, we have tried to incorporate some of the knowledge distilled in Chapter 5, especially regarding the recognition of tracks presenting some degree of tonal ambiguity. Although we have not managed to obtain a balance between correct classification and finer modal detail —as it was originally intended— our methods perform fairly well when assessing larger modal vocabularies, slightly below current state of the art methods present —and hidden— in commercial applications. On the other hand, EDM evaluations with binary dictionaries seem to provide a good scenario for our algorithms, which achieve the best results on 'binary-ambiguous' evaluations, obtaining a second position with 'binary-single' labels (i.e. MIREX). Furthermore, our final evaluation has revealed that our methods accomodate exceptionally well to other musical genres, such as euroclassical and popular music, providing state of the art performance, what could make them optimal choices for style-agnostic key estimation endeavours —although that was not part of our initial goals. In the next chapter, we present our concluding reflections, addressing some of the issues raised in this chapter, and pointing at future lines of work. Additionally, we provide a summary of the main contributions stemming from this dissertation.

Chapter 7
Conclusion

By going further, make your way
Till looking back at where you've wandered,
You look back on that path you may
Not set foot on from now onward.
Antonio Machado, *Fields of Castile* (1912)

Throughout this dissertation, I have described the main research path explored during my doctorate research, accomplished in the course of four years at the Music Technology Group from Universitat Pompeu Fabra, in Barcelona. As stated in the Introduction, the operational frame of my research was given by the GiantSteps project, a collective international effort to bring the powers of computational knowledge-extraction and summarisation into the reality of practising EDM makers. Throughout its existence, the project produced outcomes in areas such as music information retrieval, human-computer interaction, or knowledge visualisation. As such, it has materialised in the work of consortium partners[107] and fellow doctorandi, who have dealt with aspects of timbre and concatenative synthesis (Ó Nuanáin et al., 2017) and rhythmic spaces (Gómez-Marín et al., 2016). My study, complementarily, revolved around the identification of tonal practises in EDM, in order to implement better informed algorithms for tonality estimation, an endeavour that is received with interest in DJs and producers circles. Surely the combination of my personal interests, my academic background as a musician, and my obvious limitations with information retrieval expertise, have contributed to make this thesis exactly what it is. In any case, I have tried to compensate my weaknesses with a solid theoretical background on tonality, documenting and evidencing the need to study tonal practises within the musical contexts in which they are developed. Similarly, I have shown how the

[107] http://www.giantsteps-project.eu/#/downloads/deliverables

available test datasets and evaluation strategies seem to neglect this assumption, by borrowing models from euroclassical tonality without much questioning. I have underlined the importance of understanding, from perceptual, interpretive, and aesthetic viewpoints, the importance of tonal ambiguity in musical discourse —something that is neither reflected in current evaluation methods— proposing means to address this issue in data recollection strategies and evaluation methodologies. In this respect, this study contributes two datasets comprising over 2,000 audio excerpts with tonal annotations of varying degrees of detail. I regard this as an important contribution on its own, with a prospective effect in the work of fellow researchers working in the areas of music information retrieval and electronic dance music. I have provided musical analyses and insights of what I found to be the most relevant aspects of pitch configurations in EDM, some of which could have a straightforward applicability in key estimation methods aimed at this particular meta-genre, presenting my own key-finding algorithms, mostly adapting already existing methodologies.

To conclude this dissertation, in the following section I summarise in more detail the main contents of each chapter, emphasising the original contributions stemming from this research. Section 7.2, additionally, points at potential lines of continuing this work, both in the fields of musical analysis and computational key estimation.

7.1 Summary and Contributions

I started this thesis by declaring my motivations and research goals, for which I tried to present EDM as an interesting musical domain, posing specific challenges both to the music analyst and the engineer. I stressed the significance of automatic key estimation among EDM practitioners, trying to underline the fact that, despite a common preconception of this music as tonally uninteresting, there are potential indicators of idiosyncratic configurations. Such peculiar configurations most likely stem from compositional techniques centred on multi-track sequencers, and I declared my intention to study their potential effect in configuring novel tonal arrangements, and to implement tonality estimation methods taking advantage of them.

Chapters 2 and 3 presented the theoretical foundations of the dissertation, covering aspects of music theory and computational key estimation, respectively. In Chapter 2 (Fundamentals of Tonality), I presented the basic tonal terminology that, in varying degrees, has been used throughout this dissertation. I reviewed the basic workings of tonality in euroclassical music, what is typically assumed as the 'yardstick' upon which any other tonal practises are measured, and on which a great amount of literature is available. However, I also tried to provide insights into popular music theory —an area of study which has been increasingly attended in the past 20 years—

highlighting the aspects that situate popular music modality in a clearly differentiated arena from euroclassical binary tonality. I have also provided a summary of the coverage on tonality in EDM research, underlining the vagueness with which the topic is normally addressed in scholar literature, perhaps with the exception of the single publication by Wooller & Brown (2008). Furthermore, I introduced the musical effect referred to as *harmonic mixing* —one of the direct applications of key estimation systems amongst DJs and producers— which is typically thought of in sequential terms —very much like modulation— to provide a dramatic tensional curve to large mixes or DJ sets, and for which all the existing applications provide a simple binary vocabulary (based on minor and major modes), offering no insight into other potentially significant tonal marks.

Complementarily, Chapter 3 (Tonality and Computers) addressed the area of computational key estimation, with a short introduction to its perceptual reality, and to how it has been modelled in cognitive psychology, mostly as statistical tonal hierarchisation through exposure to music (Krumhansl, 1990). I presented a short discussion on early computational methods for key finding on symbolic musical representations, before entering into the main body of the chapter, discussing the particularities of key estimation procedures from audio signals, covering aspects of signal processing and focusing on template matching approaches. This way, I intended to set the basis for my own computational methods for key identification, in line with the notion of tonal hierarchies in music theory and music psychology.

After establishing the scientific basis of my research, Chapter 4 stood as the central turning point in the thesis, introducing the methodological ground over which I proposed my first contributions. I started the chapter reporting on existing musical collections for computational tonality estimation, including euroclassical, popular and electronic dance music styles, showing that most datasets with explicit key information follow a binary major/minor modal system (with the exception of the corpora by Temperley & De Clercq (2013) and Burgoyne et al. (2011), mostly aimed at chord recognition endeavours). I also described typical evaluation methodologies, normally based upon weighted rating of correctly estimated keys or in tonally related regions, offering a critique of the MIREX evaluation system, which is decidedly biased towards euroclassical music, as evidenced by the inclusion of new datasets in the competition. I observed that there has not been much activity in the task, probably due to the lack of challenging datasets, a situation that it is apparently changing in recent years. Furthermore, I proposed a revision of assumptions in simple evaluation methods, regarding the quality and duration of tracks, as well as the weighting of flat-side neighbouring keys.

The main contributions of Chapter 4 include the description of a new EDM key dataset and a preliminary evaluation of existing methods and collections, as condensed in the following points:

(1) The GiantSteps key dataset, a collection of 600 two-minute excerpts, comprising over 15 different EDM subgenres, with a global-key label per track, obtained with an automated approach based of HTML parsing of web fora, and multiple user annotations.

(2) A preliminary evaluation of existing key estimation algorithms, including commercial applications tailored to EDM, supporting the study of tonal idiosyncrasies in electronic dance music, and the development of better informed algorithms. This initial evaluation showed that the current state of the art is able to classify accurately around 70% in EDM corpora (compared to the 90% achieved over euroclassical music) leaving considerable room for improvement.

After presenting the basic methodological framework, Chapters 5 and 6 condensed the principal contributions of this dissertation. In Chapter 5 (A Study of Tonal Practises in EDM), I described two additional datasets, in an attempt to balance existing collections with better and more numerous labels, and a finer degree of tonal detail. Furthermore, taking from the newly introduced collections, I elaborated a study of tonal practises in EDM, based on a simple taxonomy of likely tonal configurations. This was proposed as a simple annotation method for electronic dance music, although it could be of potential utility beyond the reach of this meta-genre. The main contributions of the chapter are summarised as follows:

(3) A lax annotation framework giving account of tonically ambiguous and modally ambivalent fragments, especially useful to describe music with bimodal openness and other popular music modal features, allowing for finer modal descriptions while being easily parseable by computer.

(4) Two new manually annotated datasets, obtained with the help of two external collaborators, fully reformatted and revised. These two collections (referred to as the Beatport and GiantSteps datasets throughout this text), add to more than 2,000 EDM audio snippets, with varying degrees of modal annotations, key changes, and pitch-class set descriptions, spanning over 15 different subgenres.

(5) A study of tonal practises in EDM based on the newly curated collections, focusing on global characterisation of musical fragments, in resonance with what Tagg called the extended present, providing evidence of the expressive role of tonal ambiguity in electronic dance music, as well as presenting simple statistics of its main modal scales, pitch cardinalities and tonal distributions.

In Chapter 6 (Automatic Key Estimation in EDM), additionally, I explained the key-finding methods developed in the course of my research, revolving around the creation of specific tonality profiles for EDM, based on the statistical analysis of some of the corpora described in previous chapters. These include,

(6) Two iterations on binary tonality profiles as described in Faraldo et al. (2016a) and Faraldo et al. (2017), providing binary classification into major and minor modalities, offering a visible improvement over previous methods for global-key estimation, and getting a performance ratio just below state of the art commercial applications.

(7) One additional approach based on multiple profiles, which tries to give account of major/minor ambiguous fragments and other difficult tracks with a four-profile system, providing additional labelling of atonal or atonical tracks (no-key), ambiguous tracks (majmin) as well as reductive modal practises (other).

(8) A laxer evaluation method, aligned with the annotation framework presented in the previous chapter, which regards multiple interpretation as a valid indicator of the inherently ambiguous modality of EDM and, which in turn, provides further verbose details that could be useful in musicological analysis and harmonic mixing endeavours.

As a complement to the work reported in the body of the dissertation, I have prepared three appendices with additional information. In Appendix A, I list the publications stemming from the research described. Appendix B presents a convenient summary of the typesetting conventions used throughout the thesis, and is intended as a reference guide while consulting or reading the manuscript. Most importantly, Appendix C describes the additional materials created in the course of my research, including audio datasets, musical analyses, parsing tools for annotation and evaluation, as well as my own key estimation methods.

7.2 Future Work

A detailed list of all the experiments and analyses I would have liked to undertake would probably span through as many pages as my whole dissertation. However, time and human capacity are limited, and I regard the contents of this thesis as a humble examination, upon which other lines of investigation could be drawn.

Regarding tonal analysis of EDM, for example, I feel I have only touched the surface of what could have been studied, given the practical orientation of my enquiry towards computational key determination. However, although my endeavour has proven useful for this purpose, tonality analysis proper should give account of actual time, and consider pitch relationships embodied within metrical, timbral, structural and emotional marks. Furthermore, my analyses avoided any large scale structural implications —whether regarding complete audio tracks or full DJ sets— which are the actual places where a lot of the musical, narrative and emotional powers of EDM unfold. Besides, a tonality study of EDM should probably proceed by subgenres, since there are styles that clash almost frontally, regarding their tonal configuration. Although I have tried to focus on practises mostly disconnected from song-oriented styles, truth is that these differences should be further acknowledged and studied. In any case, any genuine line of investigation should necessarily integrate the music makers (and possibly the dancers too) to their discourses and working methodologies, in order to establish a dialogue with real creative processes, eventually assessing the relevance of the type of claims made throughout this work.

On the other hand, my approaches to computational key estimation should be taken as a timid attempt to provide evidence about the utility of integrating expert music-theoretical with engineering approaches to information extraction. In this line, a methodology based on hypermetrical key-detection could bridge the gap between local and global estimations, paving the way for studies of tonal structure in larger musical units, and eventually improving the performance of the methods presented in EDM. Recent research suggests that end-to-end systems might be the ultimate approach to computational tonality induction (e.g. Korzeniowski & Widmer, 2016, 2017). However, machine learning methodologies still need the degree of modal specification that I was seeking to provide, so perhaps this could constitute a natural continuation of the research contained herein.

***

At this point my narration reaches its end. As such, it constitutes a durable trace of my four years at the Music Technology Group, where I have learnt uncountable things from the projects, the methods and, especially, from the people I have met in this period. Some of the things I have learnt will stay with me for a long time, and they will hopefully manifest transmuted into different realities, knowledge, music and research. As for the rest of humanity, I only wish this was not done completely in vain.

Barcelona, 21st December, 2017

Ángel Faraldo

Appendix A

Publications by the Author

Faraldo, A., Jordà, S., Herrera, P. (2017). A Study of Tonal Practises in Electronic Dance Music. *Proceedings of 9th European Music Analysis Conference, Extended Abstract*. Strasbourg, France. [Not cited.]

Faraldo, A., Jordà, S., Herrera, P. (2017). A Multi-Profile Method for Key Estimation in EDM. *Proceedings of the 2017 AES International Conference on Semantic Audio*. Erlangen, Germany. [Cited in pages 9, 78, 83, 143, 157, 158, 161, and 177.]

Faraldo, A., Jordà, S., Herrera, P. (2017). The House Harmonic Filler: Interactive Exploration of Chord Sequences by Means of an Intuitive Representation. In *Proceedings of 3rd International Conference on Technologies for Music Notation and Representation*. A Coruña, Spain. [Not cited.]

Faraldo, A., Jordà, S., Herrera, P. (2016). Fkey-Edm: An Algorithm for Key Estimation in Electronic Dance Music. Extended Abstract accompanying the MirEx submission. In *Proceedings of the 17th International Society for Music Information Retrieval Conference*. New York, USA. [Cited in pages 101, 102, 103, and 156.]

Jordà, S., Gómez-Marín, D., Faraldo, Á., Herrera, P. (2016). Drumming with Style: From User Needs to a Working Prototype. In *Proceedings of the New Interfaces For Musical Expression Conference*. Brisbane, Australia. [Not cited.]

Faraldo, A., Gómez, E., Jordà, S., Herrera, P. (2016). Key Estimation in Electronic Dance Music. In *Proceedings of the 38th European Conference on Information Retrieval*, 335–347. Padua, Italy. [Cited in pages 9, 78, 80, 83, 143, 148, 149, 156, 157, and 177.]

Ó Nuanáin, C., Hermant, M., Faraldo, Á., Gómez, D. (2015). The EEEAR: Building a Real-Time MIR-Based Instrument from a Hack. Late-Breaking Demo paper. In *Proceedings of the 16th International Society for Music Information Retrieval Conference*. Málaga, Spain. [Not cited.]

Faraldo, Á., Ó Nuanáin, C., Gómez, D., Herrera, P., Jordà, S. (2015). Making Electronic Music with Expert Musical Agents. Late-Breaking Demo paper. In *Proceedings of the 16th International Society for Music Information Retrieval Conference*. Málaga, Spain. [Not cited.]

Knees, P., Faraldo, Á., Herrera, P., Vogl, R., Böck, S., Hörschläger, F., Le Goff, M. (2015). 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*, 364–370. Málaga, Spain. [Cited in pages 83, 92, 93, 96, 98, 117, and 184.]

Appendix B

Musical Typesetting Conventions

In the course of our explanation, reference to various musical objects could be ambivalent. For example, the letter 'A' in a sentence, could be interpreted as an indefinite article, a single note name, a major triad chord or a musical key. With the intention of minimising this possible ambivalence while keeping the readability and flow, the following typesetting conventions are strictly applied throughout the text:

- Single pitches and chroma names are written in lower-case sans-serif letters (a–g) followed by a flat (♭) or sharp (♯) alteration if needed.
- Octave indexes follow the pitch letter as a subscript.
- Reference tuning standard pitch is $\mathsf{a}_4 = 440$ Hz.
- Pitch-class integers are spelled in duodecimal notation (0–9, Ƭ, Ɛ), to facilitate the synthetic expression and manipulation of pitch-class sets. Pitch class 0 corresponds to chroma c, and subsequent integers imply a chromatic raise completing all twelve semitones in an octave.
- Pitch-class sets and note aggregates are represented in curly brackets, without spaces or commas between the different components, e.g. {0237ƬƐ}, {ceg}.
- Keys are written in sans-serif upper-case letters (A–G), followed by an alteration if needed, and modal labels separated by single spaces, e.g. A minor harmonic, B♭ mixolydian, G other monotonic.
- Similarly, chord roots are capitalised and followed by a chord-type shorthand without spaces (e.g. B♭7, C♯maj9). In order to minimise confusion with keys and single pitches, major and minor chords are always followed by their type label (Gmaj, A♭min).

Most of the times, reference to melodic sequences or chord progressions is given in relative terms, without referencing the specific tonic (chroma) of a passage, providing a useful level of abstraction to particular renditions of musical sequences. Although some authors make relative notation mode-dependent (using the same degree labels or numbers to represent different intervalic relations depending on the modal context) (e.g. Moore, 1992, 1995), in order to avoid confusion between different modes and relative notations, we adopt what has been referred to as the 'ionian reference model' (e.g. Tagg, 2014), which takes this scale pattern as the labelling reference for all other relative degrees:

- Relative scale degrees and compound intervals are specified with circumflex accents over Arabic numerals. Degrees corresponding to a ionian scale are written without an alteration, comprising major and perfect intervals (e.g. $\hat{1}$, $\hat{3}$, $\hat{5}$, $\hat{7}$). On the other hand, minor and diminished intervals are indicated with a flat symbol preceding the degree label (e.g. minor third = $\flat\hat{3}$, diminished fifth = $\flat\hat{5}$); augmented intervals are written with a sharp symbol (e.g. augmented fourth = $\sharp\hat{4}$).
- Similarly, relative chord functions are written as Roman numerals in sans-serif font, preceded by an alteration to indicate non-ionian degrees (i.e. minor and altered intervals). Major chord functions are capitalised, whereas minor degrees are written in lower case, since this notation does not leave room left for ambiguity (e.g. I, ii, ♯IV, ♭VII).

Appendix C

Datasets and Online Resources

In the course of our research, we have generated a number of resources, including the various datasets and computational methods described, representing the material evidence of this study, and a necessary complement to it. This appendix describes these additional resources, mentioned throughout this thesis, which are mainly intended to promote experimental reproducibility, encouraging further research in computational and musicological analysis of electronic dance music.

C.1 Available Datasets

Throughout this dissertation, we make explicit reference to three newly created EDM tonal datasets, namely, the GiantSteps key dataset (GS^{K}), the Beatport dataset (BP) and the GiantSteps+ dataset (GS^{+}). Although the data from these three collections is highly related (they all provide single key annotations), each collection presents a slightly diverging approach, either in the recollection of the data or in the analytical methodology. It is mainly for this reason that they have been published separately.

As advanced in the main body of the dissertation, examples and analyses presented throughout the text are entirely taken from our contributing datasets, what might constitute a reason on its own to download the corresponding audio files.

The GiantSteps Key Dataset

The original GiantSteps key dataset comprises 604 single key annotations of two-minute EDM excerpts downloaded from Beatport, an online music store for DJs and producers, obtained with a semi-automated procedure.[108] This dataset was originally

[108] https://www.beatport.com

published in 2015, together with a corpus of tempo annotations from the same online repository, and it is described in detail by Knees et al. (2015), which is the preferred reference when using this dataset. The data is available as a Github repository,[109] including the single key annotations together with scripts to download the linked audio files. Additionally, the Johannes Kepler University in Linz, provides an alternate download portal, offering some additional benchmarking with various commercial and research algorithms.[110]

The Beatport EDM Key Dataset

The Beatport EDM key dataset includes 1,486 additional samples from Beatport, with key annotations, comments and confidence levels generously provided by Eduard Mas Marín, and thoroughly revised and expanded according to our annotation framework by the author of this document.

The Beatport dataset contains the original MP3 audio snippets, associated with individual text files containing the global key annotations, and the original metadata from Beatport. We also provide an excel document with track information, key labels and additional comments, which can be parsed with the MIRAN library to generate annotations according to various criteria, as we explain in the next section.

The Beatport EDM Key Dataset is published with a unique Digital Object Identifier (DOI), 10.5281/zenodo.1101082, as an open access resource in Zenodo,[111] a research data repository supported by the European Union. Parts of this dataset have been previously published in a Github repository.[112] However, we strongly encourage potential users to download the current updated version. If this dataset were used in further research, we would appreciate the citation of the current doctoral dissertation and/or the provided DOI.

The GiantSteps+ EDM Key Dataset

The third dataset discussed consists in a revision of 500 items from the original GiantSteps key dataset, with updated genre information and metadata, 63 corrections over the initial single-key annotations, and more detail of analysis including key changes, pitch-class set descriptions, additional modal labels, comments, and confidence levels. This supplementary analytical information is provided as an excel spreadsheet. As with the Beatport EDM key dataset, we prepared text files with global

[109] https://github.com/GiantSteps/giantsteps-key-dataset
[110] http://www.cp.jku.at/datasets/giantsteps
[111] https://zenodo.org/record/1101082
[112] https://github.com/GiantSteps/giantsteps-mtg-key-dataset

key annotations acompanying the individual audio tracks, together with the original Beatport metadata. The GS$^+$ dataset is hosted in Zenodo, with DOI 0.5281/zenodo.1095691.[113] Although the initial GS^K dataset has been published beforehand, we encourage researchers to use this updated version, correspondingly referencing the current work.

C.2 Computational Resources

Throughout our research we have developed computational tools to analyse, parse, evaluate and summarise our data. Almost the totality of our research is implemented in the Python programming language, with the sole exception of a few classes written in C++, implementing our key estimation methods in MTG's *Essentia*.[114] Most of our computational efforts have been organised in a fully operational Python library, MIRAN, comprising various modules and simple command-line programs to execute recurrent operations. The Python modules facilitate the following operations:

1. Functionality to download tracks and stems from Beatport.
2. Evaluation definitions including the MIREX standard and our proposed methodologies.
3. Various formatting functions to convert annotation formats from popular EDM key estimation applications.
4. Key estimation algorithms as described in Chapter 6.
5. Utilities to parse excel spreadsheets, in order to facilitate the parsing and analysis of the data contained in our accompanying datasets.
6. Plotting functions to obtain key distributions, tonality profiles and confusion matrices.

The included command-line tools automate some common tasks, such as (a) downloading online audio data, (b) performing key detection and (c) evaluation, (d) finding hypermeters in audio files, (e) process large amounts of data using vamp-plugins, or (f) converting between different audio formats.

The MIRAN toolbox (DOI 10.5281/zenodo.1101110) can be downloaded directly from GitHub,[115] where library dependencies and installation instructions are also provided.

[113] https://zenodo.org/record/1095691
[114] http://essentia.upf.edu
[115] https://github.com/angelfaraldo/miran

References

Aljanaki, A., Soleymani, M., Wiering, F., & Veltkamp, R. C. (2014). MediaEval 2014: A Multimodal Approach to Drop Detection in Electronic Dance Music. In *MediaEval 2014 Workshop Proceedings*. Barcelona. [Cited in page 9.]

Amb (2010a). Ambiguity [Def. 1]. In *Oxford Dictionary of English*. Oxford University Press. [Cited in page 137.]

Amb (2010b). Ambivalence [Def. 1]. In *Oxford Dictionary of English*. Oxford University Press. [Cited in page 137.]

Babbitt, M. (1955). *Some Aspects of Twelve-Tone Composition*, vol. 12. International Music Association. [Cited in page 29.]

Bailey, D. (1993). *Improvisation: Its Nature and Practice in Music*. London: Da Capo Press. [Cited in page 31.]

Bandera, C. D., Tardón, L. J., Barbancho, I., & Sammartino, S. (2011a). Mirex 2011: Audio Key Detection System Based on Probability Density Functions. In *Proceedings of the Music Information Retrieval Evaluation Exchange*. [Cited in page 96.]

Bandera, C. D., Tardón, L. J., Barbancho, I., & Sammartino, S. (2011b). Mirex 2011: Audio Key Detection System with Statistical Key Profiles. In *Proceedings of the Music Information Retrieval Evaluation Exchange*. [Cited in page 96.]

Barbancho, A. M., Barbancho, I., Tardón, L. J., & Molina, E. (2013). *Database of Piano Chords: An Engineering View of Harmony*. Springer-Verlag. [Cited in page 84.]

Barr, T. (2000). *Techno: The Rough Guide*. London: Rough Guides. [Cited in page 8.]

Béhague, G. & Schechter, J. M. (2012). Ecuador. In *Grove Music Online - Oxford Music Online*. Oxford University Press, accessed 23 August 2017, http://www.oxfordmusiconline.com/subscriber/article/grove/music/08534. [Cited in page 45.]

Bello, J. P. & Pickens, J. (2005). A Robust Mid-level Representation for Harmonic Content in Music Signals. In *Proceedings of the International Society for Music Information Retrieval Conference*, pp. 304–311. [Cited in pages 67 and 73.]

Bernardes, G. & Davies, M. (2016). Audio Key Finding in the Tonal Interval Space. In *Proceedings of the Music Information Retrieval Evaluation Exchange*. [Cited in pages 97 and 99.]

Bernardes, G., Davies, M. E. P., & Guedes, C. (2017a). A Perceptually-Motivated Harmonic Compatibility Method for Music Mixing. In *Proceedings of the International Symposium on CMMR*, pp. 104–115. [Cited in pages 13 and 51.]

Bernardes, G., Davies, M. E. P., & Guedes, C. (2017b). Automatic Musical Key Estimation with Adaptive Mode Bias. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 1, pp. 316–320. [Cited in page 79.]

Bertin-Mahieux, T., Ellis, D. P., Whitman, B., & Lamere, P. (2011). The Million Song Dataset. In *Proceedings of the International Society for Music Information Retrieval Conference*, pp. 591–596. [Cited in page 82.]

Bogdanov, D., Wack, N., Gómez, E., Gulati, S., Herrera, P., Mayor, O., Roma, G., Salamon, J., Zapata, J., & Serra, X. (2013). ESSENTIA: An Audio Analysis Library for Music Information Retrieval. In *Proceedings of the International Society for Music Information Retrieval Conference*, pp. 493–498. [Cited in pages 102, 103, and 145.]

Bower, C. M. (1984). The Modes of Boethius. *The Journal of Musicology*, *3*(3), 252–263. [Cited in page 25.]

Brewster, B. & Broughton, F. (2000). *Last Night a DJ Saved my Life: The History of the Disk Jockey*. New York: Grove Press. [Cited in pages 2, 3, and 8.]

Brown, J. C. (1991). Calculation of a Constant Q Spectral Transform. *The Journal of the Acoustical Society of America*, *89*(1), 425–434. [Cited in page 70.]

Brown, J. C. (1992). An Efficient Algorithm for the Calculation of a Constant Q Transform. *The Journal of the Acoustical Society of America*, *92*(5), 2698–2701. [Cited in page 70.]

Burgoyne, J. A., Wild, J., & Fujinaga, I. (2011). An Expert Ground-Truth Set for Audio Chord Recognition and Music Analysis. In *Proceedings of the International Society for Music Information Retrieval Conference*, pp. 633–638. [Cited in pages 82, 92, 114, and 175.]

Butler, M. J. (2006). *Unlocking the Groove: Rhythm, Meter and Musical Design in Electronic Dance Music*. Bloomington & Indianapolis: Indiana University Press. [Cited in pages 2, 4, 6, 7, 8, 11, and 46.]

Butler, M. J. (2012). Introduction. In M. J. Butler (Ed.) *Electronica, Dance and Club Music*, chap. Introducti, pp. xi – xxxii. Farnham, Surrey, UK: Ashgate. [Cited in page 3.]

Cannam, C., Benetos, E., Mauch, M., Davies, M. E. P., Dixon, S., Landone, C., Noland, K., & Stowell, D. (2016). Mirex 2016: Vamp Plugins from the Centre for Digital Music. In *Proceedings of the Music Information Retrieval Evaluation Exchange*. [Cited in pages 97, 99, and 103.]

Cannam, C., Mauch, M., & Davies, M. (2013). Mirex 2013 Entry: Vamp Plugins From the Centre for Digital Music. In *Proceedings of the Music Information Retrieval Evaluation Exchange*. [Cited in pages 97 and 98.]

Cannam, C., Mauch, M., & Davies, M. (2014). Mirex 2014 Entry: Vamp Plugins From the Centre for Digital Music. In *Proceedings of the Music Information Retrieval Evaluation Exchange*. [Cited in page 97.]

Cannam, C., Mauch, M., & Davies, M. (2015). Mirex 2015 Entry: Vamp Plugins From the Centre for Digital Music. In *Proceedings of the Music Information Retrieval Evaluation Exchange*. [Cited in pages 97 and 99.]

Catteau, B., Martens, J., & Leman, M. (2007). A Probabilistic Framework for Audio-Based Tonal Key and Chord Recognition. *Advances in Data Analysis*, (2005), 1–8. [Cited in page 67.]

Chai, W. & Vercoe, B. (2005). Detection of Key Change in Classical Piano Music. In *Proceedings of the International Society for Music Information Retrieval Conference*, pp. 468–473. [Cited in pages 68 and 77.]

Chester, A. (1970). Second Thoughts on a Rock Aesthetic: The Band. *New Left Review*, (62), 75. [Cited in page 41.]

Chew, E. (2000). *Towards a Mathematical Model of Tonality*. Ph.D. thesis, Massachusetts Institute of Technology. [Cited in pages XXI, XXI, 56, 57, and 63.]

Chew, E. (2001). Modeling Tonality: Applications to Music Cognition. *Proceedings of the 23rd Annual Meeting of the Cognitive Science Society*, pp. 206–211. [Cited in page 63.]

Cho, T. & Bello, J. P. (2014). On the Relative Importance of Individual Components of Chord Recognition Systems. *IEEE Transactions on Audio, Speech and Language Processing*, *22*(2), 477–492. [Cited in page 73.]

Chordia, P. & Senturk, S. (2013). Joint Recognition of Raag and Tonic in North Indian Music. *Computer Music Journal*, *37*(3), 82–98. [Cited in page 77.]

Chuan, C.-h. & Chew, E. (2005a). Audio Key Finding Using FACEG: Fuzzy Analysis with the CEG Algorithm. In *Proceedings of the Music Information Retrieval Evaluation Exchange*. [Cited in page 96.]

Chuan, C. H. & Chew, E. (2005b). Polyphonic Audio Key Finding Using the Spiral Array CEG Algorithm. *IEEE International Conference on Multimedia and Expo, ICME 2005*, *2005*, 21–24. [Cited in pages 56 and 77.]

Collins, N. (2012). Influence in Early Electronic Dance Music: An Audio Content Analysis Investigation. In *Proceedings of the International Society for Music Information Retrieval Conference*, pp. 1–6. [Cited in page 9.]

Cont, A. (2008). ANTESCOFO: Anticipatory Synchronization and Control of Interactive Parameters in Computer Music. In *Proceedings of the International Computer Music Conference*. Belfast. [Cited in page 9.]

Cooley, J. W. & Tukey, J. W. (1965). An Algorithm for the Machine Calculation of Complex Fourier Series. *Mathematics of Computation*, *19*(90), 297–301. [Cited in page 68.]

Cope, D. (1991). *Computers and Musical Style*. Oxford: Oxford University Press. [Cited in page 58.]

Danielsen, A. (Ed.) (2010). *Musical Rhythm in the Age of Digital Reproduction*. Farnham, Surrey, UK: Ashgate. [Cited in page 46.]

Davies, M. E. P. & Böck, S. (2014). Evaluating the Evaluation Measures for Beat Tracking. In *Proceedings of the International Society for Music Information Retrieval Conference*, Ismir, pp. 637–642. [Cited in page 94.]

De Clercq, T. & Temperley, D. (2011). A Corpus Analysis of Rock Harmony. *Popular Music*, *30*(1), 47–70. [Cited in page 83.]

Di Giorgi, B. (2013). Automatic Chord Recognition Based on the Probabilistic Modeling of Diatonic Modal Harmony. Tech. rep., Politecnico di Milano, Milano. [Cited in pages 84 and 92.]

Doehring, A. (2015). Andrés's 'New For U': New for Us. On Analysing Electronic Dance Music. In R. von Appen, A. Doehring, D. Hemls, & A. F. Moore (Eds.) *Song interpretation in 21st-Century Pop Music*. Farnham, Surrey, UK: Ashgate. [Cited in pages 3, 4, 46, and 47.]

Doll, C. (2011). Rockin' Out: Expressive Modulation in Verse-Chorus Form. *Music Theory Online*, *17*(3). [Cited in page 40.]

Downie, J. S. (2003). Music Information Retrieval. *Annual Review of Information Science and Technology*, *37*, 295–340. [Cited in page 8.]

Downie, J. S. (2008). The Music Information Retrieval Evaluation Exchange (2005–2007): A Window into Music Information Retrieval Research. *Acoustical Science and Technology*, *29*(4), 247–255. [Cited in page 93.]

Downie, J. S., Ehmann, A. F., Bay, M., & Jones, M. C. (2010). The Music Information Retrieval Evaluation eXchange: Some Observations and Insights. In *Studies in Computational Intelligence*, vol. 274, pp. 93–115. [Cited in page 93.]

Drabkin, W. (2012). Pantonality. In *Grove Music Online - Oxford Music Online*. Oxford University Press, accessed 3 October 2017, http://www.oxfordmusiconline.com.sare.upf.edu/subscriber/article/grove/music/20835. [Cited in page 32.]

Dressler, K. & Streich, S. (2007). Tuning Frequency Estimation Using Circular Statistics. In *Proceedings of the International Society for Music Information Retrieval Conference*, pp. 2–5. [Cited in pages 71 and 152.]

Driedger, J., Müller, M., & Disch, S. (2014). Extending Harmonic-Percussive Separation of Audio Signals. In *Proceedings of the International Society for Music Information Retrieval Conference*, pp. 611–616. [Cited in page 71.]

Everett, W. (2004). Making Sense of Rock's Tonal Systems. *Music Theory Online*, *10*(4), 1–14. [Cited in pages 40 and 41.]

Faraldo, Á., Gómez, E., Jordà, S., & Herrera, P. (2016a). Key Estimation in Electronic Dance Music. In *Proceedings of the 38th European Conference on Information Retrieval*, pp. 335–347. [Cited in pages 9, 76, 78, 79, 141, 146, 147, 153, 155, and 177.]

Faraldo, Á., Jordà, S., & Herrera, P. (2016b). Fkey-Edm: an Algorithm for Key Estimation in Edm. In *Proceedings of the Music Information Retrieval Evaluation Exchange*. [Cited in pages 97, 98, 99, and 153.]

Faraldo, Á., Jordà, S., & Herrera, P. (2017). A Multi-Profile Method for Key Estimation in EDM. In *Proceedings of the AES International Conference on Semantic Audio*. Erlangen. [Cited in pages 9, 76, 79, 141, 156, 157, 160, and 177.]

Fikentscher, K. (2000). *You Better Work! - Underground Dance Music in New York City*. Hanover, US: Wesleyan University Press. [Cited in pages 2 and 8.]

Fineman, M. (1996). *The Nature of Visual Illusion*. Dover publ edn. [Cited in page 138.]

Fitzgerald, D. (2010). Harmonic/Percussive Separation Using Median Filtering. *Proceedings of the International Conference on Digital Audio Effects*, (1), 10–13. [Cited in page 71.]

Forte, A. (1973). *The Structure of Atonal Music*. Yale and London: Yale University Press. [Cited in pages 21, 29, and 30.]

Fujishima, T. (1999). Realtime Chord Recognition of Musical Sound: A System Using Common Lisp Music. In *Proceedings of the International Computer Music Conference*, vol. 9, pp. 464–467. [Cited in pages 68, 71, 73, and 77.]

Garcia, L.-M. (2005). On and On: Repetition as Process and Pleasure in Electronic Dance Music. *Music Theory Online*, *11*(4). [Cited in pages 7, 46, and 86.]

Gebhardt, R., Davies, M., & Seeber, B. (2015). Harmonic Mixing Based on Roughness. In *Proceedings of the International Conference on Digital Audio Effects*. Trondheim. [Cited in pages 13 and 51.]

Gebhardt, R. B., Davies, M. E., & Seeber, B. U. (2016). Psychoacoustic approaches for harmonic music mixing. *Applied Sciences*, *6*(5), 1–21. [Cited in pages 13 and 51.]

Glazyrin, N. (2014). Towards Automatic Content-Based Separation of DJ Mixes Into Single Tracks. In *Proceedings of the International Society for Music Information Retrieval Conference*, pp. 149–154. [Cited in page 9.]

Gómez, E. (2005). Key Estimation from Polyphonic Audio. In *Proceedings of the Music Information Retrieval Evaluation Exchange*. [Cited in page 96.]

Gómez, E. (2006a). *Tonal Description of Music Audio Signals*. Ph.D. thesis, Universitat Pompeu Fabra. [Cited in pages 10, 53, 70, 71, 76, 81, 92, 93, 102, 103, 114, 118, 151, and 162.]

Gómez, E. (2006b). Tonal Description of Polyphonic Audio for Music Content Processing. *INFORMS Journal on Computing*, *18*(3), 294–304. [Cited in pages 72, 74, 76, 77, 78, 101, 102, 103, 146, 147, 148, and 169.]

Gómez-Marín, D., Jordà, S., & Herrera, P. (2016). Rhythm Spaces. In *Proceeedings of the 4th International Workshop on Musical Metacreation*. [Cited in pages 9 and 173.]

Goto, M., Hashiguchi, H., Nishimura, T., & Oka, R. (2002). RWC Music Database: Popular, Classical and Jazz Music Databases. In *Proceedings of the International Society for Music Information Retrieval Conference*, pp. 287–288. [Cited in page 82.]

Goto, M., Hashiguchi, H., Nishimura, T., & Oka, R. (2006). AIST Annotation for RWC Music Database. In *Proceedings of the International Society for Music Information Retrieval Conference*, pp. 359–360. [Cited in page 82.]

Gulati, S., Bellur, A., Salamon, J., H.G., R., Ishwar, V., Murthy, H. A., & Serra, X. (2014). Automatic Tonic Identification in Indian Art Music: Approaches and Evaluation. *Journal of New Music Research*, *43*(1), 53–71. [Cited in page 93.]

Harte, C. (2010). *Towards Automatic Extraction of Harmony Information from Music Signals*. Ph.D. thesis, Queen Mary University of London. [Cited in pages 82, 152, and 153.]

Harte, C. & Sandler, M. (2009). Automatic Chord Recognition Using Quantised Chroma and Harmonic Change Segmentation. In *Proceedings of the Music Information Retrieval Evaluation Exchange*. [Cited in page 73.]

Harte, C., Sandler, M., & Gasser, M. (2006). Detecting Harmonic Change in Musical Audio. In *Proceedings of the 1st ACM Workshop on Audio and Music Computing Multimedia*. New York, New York, USA: ACM Press. [Cited in page 146.]

Heittola, T. & Klapuri, A. (2002). Locating Segments with Drums in Music Signals. In *Proceedings of the International Society for Music Information Retrieval Conference*. [Cited in page 9.]

Hockman, J., Davies, M., & Fujinaga, I. (2012). One in the Jungle: Downbeat Detection in Hardcore, Jungle, and Drum and Bass. In *Proceedings of the International Society for Music Information Retrieval Conference*, pp. 169–174. [Cited in page 9.]

Honingh, A. & Bod, R. (2011). In Search of Universal Properties of Musical Scales. *Journal of New Music Research*, *40*(1), 81–89. [Cited in page 19.]

Honingh, A., Panteli, M., Brockmeier, T., López Mejía, D. I., & Sadakata, M. (2015). Perception of Timbre and Rhythm Similarity in Electronic Dance Music. *Journal of New Music Research*, *44*(4), 1–18. [Cited in page 9.]

Hörschlager, F., Vogl, R., Böck, S., & Knees, P. (2015). Addressing Tempo Estimation Octave Errors in Electronic Music by Incorporating Style Information Extracted from Wikipedia. In *Proceedings of the Sound and Music Computing Conference*. [Cited in page 9.]

Hu, X. & Kando, N. (2012). User-Centered Measures vs. System Effectiveness in Finding Similar Songs. In *Proceedings of the International Society for Music Information Retrieval Conference*, Ismir, pp. 331–336. [Cited in page 94.]

Hyer, B. (2012). Tonality. In *Grove Music Online - Oxford Music Online*. Oxford University Press, accessed 23 August 2017, http://www.oxfordmusiconline.com/subscriber/article/grove/music/28102. [Cited in pages 31 and 32.]

Izmirli, Ö. (2005a). An Algorithm for Audio Key Finding. In *Proceedings of the Music Information Retrieval Evaluation Exchange*. [Cited in pages 77, 78, 96, and 98.]

Izmirli, Ö. (2005b). Template Based Key Finding from Audio. In *Proceedings of the International Computer Music Conference*. [Cited in pages 66, 70, 75, 76, 81, and 101.]

Jacobson, K., Davies, M., & Sandler, M. (2007). Toward Textual Annotation of Rhythmic Style in Electronic Dance Music. In *Audio Engineering Society Convention 123*. [Cited in page 9.]

Jansson, A. & Weyde, T. (2012). Mirex 2010 : Key Recognition With Zweiklang Profiles. In *Proceedings of the Music Information Retrieval Evaluation Exchange*. [Cited in page 97.]

Karakurt, A., Senturk, S., & Serra, X. (2016). MORTY: A Toolbox for Mode Recognition and Tonic Identification. [Cited in page 93.]

Kelz, R., Dorfer, M., Korzeniowski, F., Böck, S., Arzt, A., & Widmer, G. (2016). On the Potential of Simple Framewise Approaches to Piano Transcription. In *Proceedings of the International Society for Music Information Retrieval Conference*. [Cited in page 70.]

Kirss, P. (2007). Audio Based Genre Classification of Electronic Music. Tech. rep., University of Jyväskylä. [Cited in page 9.]

Klapuri, A. (2008). Multipitch Analysis of Polyphonic Music and Speech Signals Using an Auditory Model. *IEEE Transactions on Audio, Speech and Language Processing*, *16*(2), 255–266. [Cited in page 71.]

Knees, P., Faraldo, Á., Herrera, P., Vogl, R., Böck, S., Hörschläger, F., & Le Goff, M. (2015). Two Datasets for Tempo Estimation and Key Detection in Electronic Dance Music Annotated from User Corrections. In *Proceedings of the International Society for Music Information Retrieval Conference*, pp. 364–370. Málaga. [Cited in pages 79, 87, 88, 92, 94, 113, and 184.]

Korzeniowski, F. & Widmer, G. (2016). Feature Learning for Chord Recognition: The Deep Chroma Extractor. In *Proceedings of the International Society for Music Information Retrieval Conference*. [Cited in pages 68 and 178.]

Korzeniowski, F. & Widmer, G. (2017). End-to-End Musical Key Estimation Using a Convolutional Neural Network. In *Proceedings of the European Signal Processing Conference*. [Cited in pages 68, 75, 76, 79, 82, and 178.]

Kostka, S. & Payne, D. (1995). *Workbook for Tonal Harmony*. New York: McGraw-Hill. [Cited in page 62.]

Krumhansl, C. L. (1990). *Cognitive Foundations of Musical Pitch*. New York: Oxford University Press. [Cited in pages 28, 32, 50, 55, 58, 59, 90, and 175.]

Krumhansl, C. L. (2004). The Cognition of Tonality – as We Know it Today. *Journal of New Music Research*, *33*(3), 253–268. [Cited in page 54.]

Krumhansl, C. L. & Cuddy, L. L. (2010). A Theory of Tonal Hierarchies in Music. In M. Riess Jones, R. R. Fay, & A. N. Popper (Eds.) *Springer Handbook of Auditory Research*, *Springer Handbook of Auditory Research*, vol. 36, pp. 51–85. New York: Springer. [Cited in page 54.]

Krumhansl, C. L. & Kessler, E. J. (1982). Tracing the Dynamic Changes in Perceived Tonal Organization in a Spatial Representation of Musical Keys. *Psychological Review*, *89*, 334–368. [Cited in page 55.]

Krumhansl, C. L. & Shepard, R. N. (1979). Quantification of the Hierarchy of Tonal Functions within a Diatonic Context. *Journal of experimental psychology: Human perception and performance*, *5*(4), 579–594. [Cited in page 55.]

Lee, K. (2006). Automatic Chord Recognition from Audio Using Enhanced Pitch Class Profile. In *Proceedings of the International Computer Music Conference*. [Cited in pages 71 and 73.]

Leimeister, M., Gärtner, D., & Dittmar, C. (2014). Rhythmic Classification of Electronic Dance Music. In *Audio Engineering Society Conference: 53rd International Conference: Semantic Audio*. [Cited in page 9.]

Lerdahl, F. (1988). Tonal Pitch Space. *Music Perception*, *5*(3), 315–350. [Cited in pages XXI and 56.]

Lerdahl, F. (2001). *Tonal Pitch Space*. Oxford: Oxford University Press. [Cited in pages 36, 37, and 50.]

Lilja, E. (2009). *Theory and Analysis of Classic Heavy Metal Harmon*. Vantaa: IAML Finland. [Cited in pages 42, 43, 44, and 137.]

López-Serrano, P., Dittmar, C., Driedger, J., & Müller, M. (2016). Towards Modeling and Decomposing Loop-Based Electronic Music. In *Proceedings of the International Society for Music Information Retrieval Conference*, pp. 502–508. [Cited in page 9.]

MacMullen, J. (2003). Requirements Definition and Design Criteria for Test Corpora in Information Science. Tech. rep., School of Information and Library Science, University of North Carolina at Chapel Hil. [Cited in page 80.]

Mauch, M. (2010). *Automatic Chord Transcription from Audio Using Computational Models of Musical Context*. Ph.D. thesis, Queen Mary University of London. [Cited in page 83.]

Mauch, M., Cannam, C., & Davies, M. (2009a). OMRAS2 Metadata Project 2009. In *Proceedings of the International Society for Music Information Retrieval Conference, Late Breaking Demo*. [Cited in pages 84 and 92.]

Mauch, M. & Dixon, S. (2010a). Approximate Note Transcription for the Improved Identification of Difficult Chords. In *Proceedings of the International Society for Music Information Retrieval Conference*, pp. 135–140. [Cited in pages 71, 73, and 151.]

Mauch, M. & Dixon, S. (2010b). Simultaneous Estimation of Chords and Musical Context from Audio. *IEEE Transactions on Audio, Speech and Language Processing*, *18*(6), 1280–1289. [Cited in pages 67 and 73.]

Mauch, M., Noland, K., & Dixon, S. (2009b). Using Musical Structure to Enhance Automatic Chord Transcription. In *Proceedings of the International Society for Music Information Retrieval Conference*, pp. 231–236. [Cited in page 82.]

Mcfee, B., Raffel, C., Liang, D., Ellis, D. P. W., Mcvicar, M., Battenberg, E., & Nieto, O. (2015). librosa: Audio and Music Signal Analysis in Python. In *Proceedings of the Python in Science Conference*, pp. 1–7. [Cited in page 65.]

McLeod, K. (2001). Genres, Subgenres, Sub-Subgenres and More: Musical and Social Differentiation Within Electronic/Dance Music Communities. *Journal of Popular Music Studies*, *13*(1), 59–75. [Cited in page 3.]

McVicar, M., Santos-Rodríguez, R., Ni, Y., & De Bie, T. (2014). Automatic Chord Estimation from Audio: A Review of the State of the Art. *IEEE Transactions on Audio, Speech and Language Processing*, *22*(2), 556–575. [Cited in page 73.]

Meyer, L. (1973). *Explaining Music*. Berkeley: University of California Press. [Cited in page 11.]

Meyer, L. B. (1957). Meaning in Music and Information Theory. *The Journal of Aesthetics and Art Criticism*, *15*(4), 412–424. [Cited in page 55.]

Middleton, R. (1990). *Studying Popular Music*. McGraw-Hill Education (UK). [Cited in pages 38 and 39.]

Middleton, R. & Manuel, P. (2015). Popular Music. In *Grove Music Online - Oxford Music Online*. Oxford University Press, accessed 3 October 2017, http://www.oxfordmusiconline.com.sare.upf.edu/subscriber/article/grove/music/43179. [Cited in page 2.]

Moore, A. (1992). Patterns of Harmony. *Popular Music*, *11*(1), 73–106. [Cited in pages 39, 40, 42, 45, and 182.]

Moore, A. (1995). The So-Called 'Flattened Seventh' in Rock. *Popular Music*, *14*(2), 185–201. [Cited in pages 40, 42, 46, and 182.]

Moore, A. F. (2003). Introduction. In A. F. Moore (Ed.) *Analyzing Popular Music*, chap. Introducti, pp. 1–15. Cambridge: Cambridge University Press. [Cited in page 38.]

Moore, A. F. (2012). *Song Means: Analysing and Interpreting Recorded Popular Song*. Farnham, Surrey, UK: Ashgate. [Cited in pages 6, 40, 42, and 45.]

Müller, M. & Ewert, S. (2010). Towards Timbre-Invariant Audio Features for Harmony-Based Music. *IEEE Transactions on Audio, Speech and Language Processing*, *18*(3), 649–662. [Cited in pages 71, 73, and 151.]

Ni, Y., McVicar, M., De Bie, T., & Santos-Rodriguez, R. (2012). An End-to-End Machine Learning System for Harmonic Analysis of Music. *IEEE Transactions on Audio, Speech & Language Processing*. [Cited in pages 67, 70, 71, and 73.]

Noland, K. (2009). *Computational Tonality Estimation: Signal Processing and Hidden Markov Models*. Ph.D. thesis, Queen Mary University of London. [Cited in page 53.]

Noland, K. & Sandler, M. (2006). Key Estimation Using a Hidden Markov Model. In *Proceedings of the International Society for Music Information Retrieval Conference*, pp. 121–126. [Cited in pages 68, 70, and 78.]

Noland, K. & Sandler, M. (2007). Signal Processing Parameters for Tonality Estimation. In *Proceedings of the Audio Engineering Society Convention*. [Cited in pages XXII, 76, 81, 92, 98, 102, 103, 108, 162, and 169.]

Noland, K. & Sandler, M. (2009). Influences of Signal Processing, Tone Profiles, and Chord Progressions on a Model for Estimating the Musical Key from Audio. *Computer Music Journal*, *33*(1), 42–56. [Cited in pages 75 and 77.]

Ó Nuanáin, C., Herrera, P., & Jordà, S. (2017). Rhythmic Concatenative Synthesis for Electronic Music: Techniques, Implementation, and Evaluation. *Computer Music Journal*, *32*(2), 21–37. [Cited in page 173.]

Oramas, S. & Sordo, M. (2016). Knowledge Is Out There: A New Step in the Evolution of Music Digital Libraries. *Fontes Artis Musicae*, *63*(4), 285–298. [Cited in page 9.]

Panteli, M., Bogaards, N., & Honingh, A. (2014). Modeling Rhythm Similarity for Electronic Dance Music. In *Proceedings of the International Society for Music Information Retrieval Conference*. [Cited in page 9.]

Papadopoulos, H. (2010). *Joint Estimation Of Musical Content Information From an Audio Signal*. Ph.D. thesis, Université Pierre et Marie Curie. [Cited in page 67.]

Papadopoulos, H. & Peeters, G. (2007). Large-Scale Study Of Chord Estimation Algorithms Based On Chroma Representation and HMM. *Content-Based Multimedia Indexing*, pp. 53–60. [Cited in pages 67 and 73.]

Papadopoulos, H. & Peeters, G. (2009). Local Key Estimation Based on Harmonic and Metric Structures. *Proceedings of the International Conference on Digital Audio Effects*, pp. 1–8. [Cited in page 68.]

Pauwels, J., Martens, J., & Peeters, G. (2012). The Ircamkeychord Submission for Mirex 2012. In *Proceedings of the Music Information Retrieval Evaluation Exchange*. [Cited in page 97.]

Pauwels, J. & Peeters, G. (2013). The Ircamkeychord Submission for Mirex 2013. In *Proceedings of the Music Information Retrieval Evaluation Exchange*. [Cited in page 97.]

Pauws, S. (2004). Musical Key Extraction from Audio. In *Proceedings of the International Society for Music Information Retrieval Conference*. [Cited in pages 70, 74, 76, 81, 93, and 101.]

Pauws, S. (2005). Keyex: Audio Key Extraction. In *Proceedings of the Music Information Retrieval Evaluation Exchange*. [Cited in page 96.]

Peeters, G. (2006a). Chroma-Based Estimation of Musical Key from Audio-Signal Analysis. In *Proceedings of the International Society for Music Information Retrieval Conference*, pp. 155–120. [Cited in pages 66, 81, and 101.]

Peeters, G. (2006b). Musical Key Estimation of Audio Signal Based on Hidden Markov Modeling of Chroma Vectors. In *Proceedings of the International Conference on Digital Audio Effects*, pp. 127–131. Montreal. [Cited in pages 68, 71, 73, 76, and 152.]

Peeters, G. (2010). Mirex 2010 Audio Key Detection Task: Ircamkeymode. In *Proceedings of the Music Information Retrieval Evaluation Exchange*. [Cited in page 96.]

Peeters, G. (2011). Mirex 2011 Audio Key Detection Task: Ircamkeymode. In *Proceedings of the Music Information Retrieval Evaluation Exchange*. [Cited in page 96.]

Peeters, G. (2012). Mirex 2012 Audio Key Detection Task: IrcamKeyMode. In *Proceedings of the Music Information Retrieval Evaluation Exchange*. [Cited in page 97.]

Peeters, G. & Cornu, F. (2013). Mirex 2013 Audio Key Detection Task: Ircamkeymode. In *Proceedings of the Music Information Retrieval Evaluation Exchange*. [Cited in page 97.]

Peeters, G. & Fort, K. (2012). Towards a (Better) Definition of the Description of Annotated Mir Corpora. In *Proceedings of the International Society for Music Information Retrieval Conference*, pp. 25–30. [Cited in page 80.]

Piston, W. (1991). *Harmony*. Spanish edition and translation by Milán Amat J.L. Madrid: Norton. [Cited in page 27.]

Pollack, A. W. (1999). *Notes on ... Series*. Soundscapes.info, accessed 1 February 2015, http://www.icce.rug.nl/˜soundscapes/DATABASES/AWP/awp-notes_on.shtml. [Cited in pages 84 and 92.]

Purwins, H. & Blankertz, B. (2005). CQ-Profiles for Key Finding in Audio. In *Proceedings of the Music Information Retrieval Evaluation Exchange*. [Cited in page 96.]

Purwins, H., Blankertz, B., & Obermayer, K. (2000). A New Method for Tracking Modulations in Tonal Music in Audio Data. *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks*, *6*, 270–275. [Cited in pages 75 and 98.]

Purwins, H., Blankertz, B., & Obermayer, K. (2001). Constant Q Profiles for Tracking Modulations in Audio Data. *Proceedings of the International Computer Music Conference*, *1*(20), 274–276. [Cited in page 77.]

Raffel, C., Mcfee, B., Humphrey, E. J., Salamon, J., Nieto, O., Liang, D., & Ellis, D. P. W. (2014). mir_eval: A Transparent Implementation of Common MIR Metrics. In *Proceedings of the International Society for Music Information Retrieval Conference*, pp. 367–372. [Cited in page 100.]

Rameau, J.-P. (1971). *Treatise on Harmony*. English edition and translation by Gosset, P. New York: Dover Publications, Inc. [Cited in pages 27 and 33.]

Ratcliffe, R. (2013). Analytical Précis of Chime by Orbital: Towards an Analysis of Electronic Dance Music. *eOREMA*. [Cited in pages 46, 47, and 49.]

Rebelo, A., Fujinaga, I., Paszkiewicz, F., Marcal, A. R., Guedes, C., & Cardoso, J. S. (2012). Optical Music Recognition: State-of-the-Art and Open Issues. *International Journal of Multimedia Information Retrieval*, *1*, 173–190. [Cited in page 9.]

Reti, R. (1958). *Tonality, Atonality, Pantonality: A Study of Some Trends in Twentieth Century Music*. London: Barrie and Rockliff. [Cited in pages 22 and 31.]

Reynolds, S. (1998). *Energy Flash: A Journey Through Rave Music and Dance Culture*. London: Picador. [Cited in pages 2 and 8.]

Riemann, H. (1903). Harmony Simplified or The Theory of the Tonal Functions of Chords. [Cited in pages 33 and 56.]

Rietveld, H. (1998). *This is our House: House music, Cultural Spaces, and Technology*. Brookfield: Ashgate. [Cited in pages 2 and 8.]

Röbel, A. & Rodet, X. (2005). Efficient Spectral Envelope Estimation and its Application to Pitch Shifting and Envelope Preservation. In *Proceedings of the International Conference on Digital Audio Effects*, pp. 30–35. Madrid. [Cited in pages 71 and 151.]

Rocha, B., Bogaards, N., & Honingh, A. (2013). Segmentation and Timbre- and Rhythm-Similarity in Electronic Dance Music. In *Proceedings of the Sound and Music Computing Conference*, pp. 754–761. [Cited in page 9.]

Rocher, T., Robine, M., Hanna, P., & Strandh, R. (2010). Dynamic Chord Analysis. In *Proceedings of the Music Information Retrieval Evaluation Exchange*. [Cited in page 96.]

Salamon, J. & Urbano, J. (2012). Current Challenges in the Evaluation of Predominant Melody Extraction Algorithms. In *Proceedings of the International Society for Music Information Retrieval Conference*, pp. 5–10. [Cited in page 94.]

Scarfe, T., Koolen, W., & Kalnishkan, Y. (2014). Segmentation of Electronic Dance Music. *International Journal of Engineering Intelligent Systems for Electrical Engineering and Communications*, *22*(3). [Cited in page 9.]

Schaeffer, P. (2017). *Treatise on Musical Objects: Essays Across Disciplines*, vol. 20. University of California Press. [Cited in page 4.]

Schaffrath, H. & Huron, D. (1995). The Essen Folksong Collection in the Humdrum Kern Format. *Menlo Park, CA: Center for Computer Assisted Research in the Humanities*. [Cited in page 61.]

Schedl, M., Gómez, E., & Urbano, J. (2014). Music Information Retrieval: Recent Developments and Applications. *Foundations and Trends in Information Retrieval*, *8*(2-3), 127–261. [Cited in page 8.]

Schellenberg, E. G. & Von Scheve, C. (2012). Emotional Cues in American Popular Music: Five Decades of the Top 40. *Psychology of Aesthetics, Creativity and the Arts*, *6*(3), 196–203. [Cited in pages 51 and 131.]

Schoenberg, A. (1974). *Harmonielehre*. Spanish edition and translation by Barce R. Madrid: Real Musical. [Cited in pages 27, 36, and 38.]

Scholz, R., Ramalho, G., & Cabral, G. (2016). Cross-task Study on Mirex Recent Results: an Index for Evolution Measurement and Some Stagnation Hypotheses. In *Proceedings of the International Society for Music Information Retrieval Conference*, pp. 372–378. [Cited in page 94.]

Schwarz, D. & Rodet, X. (1999). Spectral Envelope Estimation and Representation for Sound Analysis – Synthesis. *Proceedings of the ICMC*, pp. 351–354. [Cited in page 71.]

Serra, X. (2014). Creating Research Corpora for the Computational Study of Music: the case of the CompMusic Project. In *AES Semantic Audio Conference*, pp. 1–9. [Cited in page 80.]

Sha'ath, I. (2011). Estimation of Key in Digital Music Recordings. Tech. rep., Birkbeck College, University of London. [Cited in pages XXII, 9, 74, 77, 86, 92, 102, 103, 104, 148, 149, and 151.]

Shapiro, P. (Ed.) (2000). *Modulations: A History of Electronic Music - Throbbing Words on Sound*. New York: Distributed Art Publishers. [Cited in page 8.]

Sheh, A. & Ellis, D. P. W. (2003). Chord Segmentation and Recognition Using EM-Trained Hidden Markov Models. *Proceedings of the International Society for Music Information Retrieval Conference*, pp. 185–191. [Cited in page 67.]

Shepard, R. N. (1964). Circularity in Judgments of Relative Pitch. *The Journal of the Acoustical Society of America*, *36*(12), 2346–2353. [Cited in page 71.]

Shmulevich, I. & Yli-Harja, O. (2000). Localized Key Finding: Algorithms and Applications. *Music Perception*, *17*(4), 531–544. [Cited in page 64.]

Sicko, D. (1999). *Techno Rebels: The Renegades of Electronic Funk*. New York: Billboard Books. [Cited in page 8.]

Smith, J., Burgoyne, J. A., Fujinaga, I., De Roure, D., & Downie, J. S. (2011). Design and Creation of a Large-Scale Database of Structural Annotations. In *Proceedings of the International Society for Music Information Retrieval Conference*, pp. 555–560. [Cited in page 82.]

Snoman, R. (2009). *Dance Music Manual: Tools, Toys and Techniques*. Oxford, UK: Focal Press, second edi edn. [Cited in page 6.]

Solberg, R. T. (2014). “Waiting for the Bass to Drop”: Correlations Between Intense Emotional Experiences and Production Techniques in Build-up and Drop Sections of Electronic Dance Music. *Dancecult: Journal of Electronic Dance Music Culture*, *6*(1), 61–82. [Cited in pages 8 and 46.]

Spicer, M. (2004). (Ac) cumulative Form in Pop-Rock Music. *Twentieth Century Music*, *1*(1), 29–64. [Cited in pages 4, 7, 11, 46, 48, and 134.]

Stephenson, K. (2002). *What to Listen for in Rock: A Stylistic Analysis*. New Haven & London: Yale University Press. [Cited in pages XXI, 41, 42, 44, 46, and 64.]

Straus, J. N. (2005). *Introduction to Post-Tonal Theory*. New Jersey: Pearson Prentice Hall, third edit edn. [Cited in pages 21 and 29.]

Tagg, P. (1994). From Refrain to Rave: The Decline of Figure and the Rise of Ground. *Popular Music*, *13*(02), 209–222. [Cited in pages 2, 39, 40, 47, 48, and 134.]

Tagg, P. (2012). *Music’s Meaning - A Modern Musicology for Non-Musos*. New York & Huddersfield: The Mass Media Music Scholars’ Press. [Cited in pages 13, 17, 41, and 128.]

Tagg, P. (2013). *Trouble with Tonal Terminology*. Online resource, accessed 6 February 2017, http://tagg.org/xpdfs/Aharonian2011.pdf. [Cited in page 17.]

Tagg, P. (2014). *Everyday Tonality II - Towards a Tonal Theory of What Most People Hear*. New York & Huddersfield: The Mass Media Music Scholar’s Press. [Cited in pages 13, 17, 18, 22, 26, 31, 38, 40, 41, 45, 128, 134, 176, and 182.]

Temperley, D. (1999). What's Key for Key? The Krumhansl-Schmuckler Key-Finding Algorithm Reconsidered. *Music Perception: An Interdisciplinary Journal*, *17*(1), 65–100. [Cited in pages XXI, 59, 60, 61, 75, 98, 108, 148, 149, 151, and 159.]

Temperley, D. (2001). *The Cognition of Basic Musical Structures*. Cambridge, Massachusetts: MIT Press. [Cited in pages XXI, XXI, 42, 58, 60, 61, and 114.]

Temperley, D. (2007a). *Music and Probability*. Cambridge, Massachusetts: MIT Press. [Cited in pages 58, 61, 62, 63, and 77.]

Temperley, D. (2007b). The Melodic-Harmonic 'Divorce' in Rock. *Popular Music*, *26*(2). [Cited in page 46.]

Temperley, D. (2007c). The Tonal Properties of Pitch-Class Sets: Tonal Implication, Tonal Ambiguity, and Tonalness. *Computing in Musicology*, *15*(Tonal Theory for the Digital Age), 24–38. [Cited in page 49.]

Temperley, D. (2011). Scalar Shift in Popular Music. *Music Theory Online*, *17*(4), 1–19. [Cited in page 40.]

Temperley, D. & De Clercq, T. (2013). Statistical Analysis of Harmony and Melody in Rock Music. *Journal of New Music Research*, *42*(3). [Cited in pages XXII, 46, 52, 61, 77, 78, 80, 83, 85, 92, 100, 111, 114, 141, 166, and 175.]

Terhardt, E. (1974). Pitch, Consonance, and Harmony. *Journal of the Acoustical Society of America*, *55*(5), 1061–1069. [Cited in page 26.]

Thornton, S. (1995). *Club Cultures: Music, Media and Subcultural Capital*. Oxford, UK: Polity Press. [Cited in page 2.]

Trehub, S. E., Becker, J., & Morley, I. (2015). Cross-Cultural Perspectives on Music and Musicality. *Philosophical transactions of the Royal Society of London*, *370*(1664). [Cited in page 19.]

Tymoczko, D. (2002). Stravinsky and the Octatonic: A Reconsideration. *Music Theory Spectrum*, *24*(1), 68–102. [Cited in pages 32 and 48.]

Tzanetakis, G. (2012). MARSYAS Submission to Mirex 2012. In *Proceedings of the Music Information Retrieval Evaluation Exchange*. [Cited in page 97.]

Ueda, Y., Uchiyama, Y., Nishimoto, T., Ono, N., & Sagayama, S. (2010a). HMM-Based Approach for Automatic Chord Detection Using Refined Acoustic Features. *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, pp. 5518–5521. [Cited in page 71.]

Ueda, Y., Uchiyama, Y., Ono, N., & Sagayama, S. (2010b). Mirex 2010: Joint Recognition of Key and Chord From Music Audio Signals Using Key-Modulation HMM. In *Proceedings of the Music Information Retrieval Evaluation Exchange*. [Cited in page 96.]

Urbano, J., Bogdanov, D., & Herrera, P. (2014). What Is the Effect of Audio Quality on the Robustness of MFCCs and Chroma Features. In *Proceedings of the International Society for Music Information Retrieval Conference*. [Cited in pages 101 and 106.]

Vorobyev, Y. & Coomes, E. (2012). *Beyond Beatmatching: Take Your DJ Career to the Next Level*. Mixed In Key. [Cited in pages 10, 50, and 51.]

Whittall, A. (2011). Tonality. In A. Latham (Ed.) *The Oxford Companion to Music*. Oxford University Press, Accessed 3 October 2017, http://www.oxfordmusiconline.com.sare.upf.edu/subscriber/article/opr/t114/e6829. [Cited in pages 31 and 33.]

Wooller, R. & Brown, A. R. (2008). A Framework for Discussing Tonality in Electronic Dance Music. In *Sound: Space - The Australasian Computer Music Conference*, pp. 91–95. [Cited in pages 2, 48, 49, 51, 52, 79, and 175.]

Yadati, K., Larson, M., Liem, C. C. S., & Hanjalic, A. (2014). Detecting Drops in Electronic Dance Music: Content Based Approaches to a Socially Significant Music Event. In *Proceedings of the International Society for Music Information Retrieval Conference*, pp. 143–148. [Cited in page 9.]

Zhu, Y. (2005). An Audio Key Finding Algorithm. In *Proceedings of the Music Information Retrieval Evaluation Exchange*. [Cited in page 96.]

Zhu, Y., Kankanhalli, M. S., & Gao, S. (2005). Music Key Detection for Musical Audio. In *Proceedings of the International Multimedia Modelling Conference*, pp. 30–37. [Cited in page 71.]