Similarity and Style in Electronic Dance Music Drum Rhythms

Daniel Gómez Marín

PHD THESIS UPF / 2018

THESIS DIRECTOR

Dr. Sergi Jordà
Music Technology Group
Dept. of Information and Communication Technologies

Universitat Pompeu Fabra Barcelona

Acknowledgments

This thesis is dedicated to Eulalia, Elisa and León, my beautiful family, who have always encouraged curiosity and endurance with great kindness.

During this doctoral process I became very close to my PhD fellows Ángel Faraldo, Cárthach O'Nuánain and Martin Hermant as we became intimate friends, sharing the intensity of a PhD process. Together we won prizes, traveled, taught, discussed, struggled and laughed. Our team became a reality mostly by the sharpness of Dr. Sergi Jordà who decided it was a good idea to have us four as his PhD students with the co-supervision of Perfecto Herrera. I am very thankful to both Sergi and Perfe as their guides and challenges to my ideas traced the many roads that finally materialized as this thesis. I want to acknowledge everyone who kindly took their time to participate in the different listening experiments, and specially the musicians Sano and Boska who generously shared their talent and time to reflect on the systems that I worked on.

During my life as a PhD student I was surrounded by marvelous people who made the process possible, and whom I also want to thank. Andrea who became my family and support. The guys from the Veni Vidi Bici bicycle club, with whom I explored so many kilometers of the beautiful roads in Cataluña. Nico, Cata, GC and Cosmo who introduced me to the calm of Montseny and the beauty of Sant Pol de Mar. All the MTG staff, specially Xavier Serra, Sebas Mealla, Marius Miron, David D'Almazzo and Alastair Porter. During the writing stage, Paula shed a bright light to process, reminding me the art of day-by-day construction and giving me her patient love.

Abstract

This thesis presents original research carried out in the topic of electronic dance music (EDM) drum sequencing, a fundamental and yet underdeveloped subject in the music production literature. The work undertaken is focused in two main areas: similarity between drum patterns and modeling of drumming style. The study of pattern similarity is rooted in current knowledge on human processing of monophonic rhythms, and is expanded until a model capable of predicting similarity sensations of polyphonic drum rhythms is reached. With this model, *RhythmSpace*, a graphical system for the continuous real-time exploration of drum pattern collections, is developed. The second area of research, drumming style modeling, is approached from a statistical perspective, developing a generative model capable of learning styles from examples and creating original drum patterns in the learned styles. This model allows high-level musical flexibility, letting a musician combine and transform styles in real-time during the generative process. Taking advantage of this model, a style-based drum machine application, *DrDrums*, is implemented and evaluated in subject-based experiments.

Resumen

Esta tesis presenta una investigación original llevada a cabo en el área de la secuenciación de baterías de música electrónica de baile (EDM), un tema fundamental y al mismo tiempo poco desarrollado en la literatura de producción musical. El trabajo realizado se enfoca en dos áreas: la similitud entre patrones de batería y los estilos en la composición de patrones percusivos. El estudio de la similitud entre patrones se fundamenta en el conocimiento actual del procesamiento humano de patrones monofónicos, y es expandido hasta alcanzar un modelo capaz de predecir sensaciones de similitud en ritmos polifónicos. Con este modelo se ha creado *RhythmSpace*, un sistema gráfico para la exploración en tiempo real de colecciones de patrones de batería. La segunda área de investigación, el estilo de composición de baterías, es abordada desde una perspectiva estadística, desarrollando un modelo generativo capaz de aprender estilos desde ejemplos y luego crear patrones originales en los estilos aprendidos. Este modelo estadístico permite una flexibilidad musical de alto nivel, haciendo posible que un músico combine y transforme estilos en tiempo real durante el proceso generativo. Usando este modelo se implementa *DrDrums*, una máquina de ritmos con inteligencia de estilo, que es evaluada experimentalmente con sujetos.

Contents

	Pàg.
Abstract	ix
Resumen	x
1. INTRODUCTION	1
1.1. Motivation	1
1.2 Context	2
1.3 Existing technologies for EDM drum production	11
1.4 Opportunities and challenges	13
1.5 Thesis Outline	14
2. STATE OF THE ART	17
2.1 Introduction	17
2.1.1 What is Rhythm?	18
2.2 Human Rhythm Processing	19
2.2.1 Pulse and meter	20
2.2.2 Syncopation	25
2.2.2.1 Monophonic syncopation	26
2.2.2.2 Polyphonic syncopation	28
2.3 Similarity in Percussive Rhythms	30
2.3.1 Notions of similarity	31
2.3.1.1 Dimensionality reduction techniques	36
2.3.2 Monophonic rhythmic similarity	37
2.3.3 Polyphonic rhythmic similarity	42
2.3.4 Rhythm spaces	44
2.4 Style in EDM	45
2.4.1 Definitions of style	45
2.4.2 How to study EDM styles	48
2.5 EDM Production and Drum Sequencing	50
2.6 Generative Music Sequencing	55
2.6.1 Generative drum sequencing	58
2.7 Conclusions	61

3 TOWARDS METRICS OF RHYTHM SIMILARITY 65
3.1 Introduction 65
3.2 Experiment 1: Monophonic similarity and syncopation 69
3.2.1 Methods 73
3.2.1.1 Participants 73
3.2.1.2 Material 73
Objective Distance Metrics 73
Stimuli 76
3.2.1.3 Procedure 77
3.2.2 Results 78
3.2.3 Discussion 81
3.3 Using Beat-Induced Similarity Ratings to Define Similarity Metrics 84
3.3.1 Introduction 84
3.3.2 Expanded metrics for monophonic similarity 85
3.3.2.1 Syncopation and awareness distance (SAD) 86
3.3.2.2 Pattern coincidence and awareness metric (PAD) 87
3.3.2.3 Deducing the weights for PAD and SAD 87
3.3.2.4 Example to compute the metrics 89
3.3.4 Discussion and conclusions 90
3.4 Symbolic Descriptors for Polyphonic Drum Similarity 90
3.4.1 Symbolic drum pattern descriptors 93
3.4.1.1 Number of Instruments (NOI) 96
3.4.1.2 hisync, midsync and losync 96
3.4.1.3 Polysync 96
3.4.1.4 hiD, midD, loD 97
3.4.1.5 losyness, midsyness, hisyness 96
3.4.1.6 stepD 97
3.4.1.7 lowness, midness, hiness 97
3.5 Experiment 2: Objective Similarity in Alf Gabrielsson's Rhythm Spaces 97
3.5.1 Methods 99
3.5.1.1 Materials 99
3.5.1.2 procedure 100
3.5.2 Results and discussion 100
3.6 Experiment 3: Generalizing Polyphonic Descriptors 102

for EDM
3.6.1 Methods 103
3.6.1.1 participants 103
3.6.1.2 Material 103
3.6.1.3 Procedure 105
3.6.2 Results 105
3.6.2.1 From EDM to GE1 108
3.6.2.2 Relevant set of descriptors 109
3.6.3 Discussion 109
3.7 Conclusions 111

4. GENERATIVE TOOLS FOR EDM RHYTHM 115
4.1 Introduction 115
4.1.1 Fundamentals of DrDrums 116
4.1.2 Fundamentals of RhythmSpace 116
4.1.3 Technologies of DrDrums and RhythmSpace 118
4.2 Agnostic Density Transformer 118
4.2.1 Agnostic density transformation example 122
4.2.2 Considerations of real-time interaction with density 123
4.3 Drumming Style Extractor 125
4.3.1 Pre-processing the patterns 126
4.3.2 Analyzing the patterns 128
4.3.3 Generalizing stylistic music knowledge extraction 131
4.4 Drum Pattern Generator 133
4.4.1 Using style information to generate a pattern 134
4.4.2 Style interaction concepts 136
4.4.2.1 Style combinations 136
4.4.2.2 Commonness 138
4.4.2.2.1 Sigmoid commonness 139
4.4.2.2.2 Power commonness 140
4.4.2.3 Inducing variation 141
4.4.2.4 inducing variation by timbre grouping 145
4.4.3 The order effect 147
4.4.4 Zero frequency states 147
4.5 DrDrums 148
4.5.1 DrDrums' new style creation 150
4.5.2 Main controls 151
4.5.3 Generation of patterns in style 151
4.5.4 Style combination 152

4.5.5 Inducing variation 152
4.5.6 Density post processing 153
4.6 Quantitative Evaluation of DrDrums 154
4.6.1 Method 155
4.6.1.1 Participants 155
4.6.1.2 Material 156
4.6.1.3 Procedure 156
4.6.2 Results 157
4.7 Qualitative Evaluation of Dr Drums 158
4.7.1 Method 159
4.7.1.1 Participants 159
4.7.1.2 Material 159
4.7.1.3 Procedure 159
4.7.2 results 160
4.7.3 Summary and discussion 164
4.8 Rhythm Spaces 167
4.8.1 Requirements of a rhythm space 169
4.8.2 Rhythm interpolation 171
4.8.2.1 Step-by-step naive algorithm (STENA) 173
4.8.2.2 Step-by-step event to onsets (STEVO) algorithm 175
4.8.2.3 STEVO-D with fixed density profile 176
4.8.3 Interpolation algorithm discussion 177
4.9 RhythmSpace Application 179
4.10 Conclusions 182

5 CONCLUSIONS AND FUTURE WORK 185
5.1 Original Contributions 188
5.1.1 Algorithms 188
5.1.2 Corpora 189
5.1.3 Applications 189
5.1.4 Experimental methods 190
5.1.5 Software Libraries 190
5.2 Further Research 190
5.3 Closing Remarks 191

APPENDIX A: LIST OF PUBLICATIONS 193

APPENDIX B: GLOSSARY 195

BIBLIOGRAPHY 197

1. INTRODUCTION

"*To go out into space today means to go further into rhythm. Far from abandoning rhythm, the Futurist producer is the scientist who goes deeper into the break, who crosses the threshold of the human drummer in order to investigate the hyperdimensions of the dematerialized Breakbeat.*"

Kodwo Eshun - More Brilliant Than The Sun

1.1 Motivation

As a musician and an engineer I have always been seduced by the interconnections between music and technology. I have been inspired by the endless ways in which they feedback one another, especially when it favors aesthetic creations or scientific discoveries. Some of these interconnections are clearly manifested in electronic dance music (EDM), especially in the way this music has been inspired by technology at many different aesthetic levels, also how it is dependent on technology for its existence and evolution, or in the way EDM has been incorporated as a cultural technology used to incite human desire to move and dance. These different relations present EDM as a scenario in which musical and technological practices coalesce extensively, a vast setting which can be analyzed from musical, technological or cultural perspectives suggesting a fertile ground for research.

As a music lover myself, I have always been inspired by some powerful experiences when music reveals itself as much more than just an interleaved collection of sound events, acquiring a meaning way beyond its components and triggering sensations and ideas otherwise unreachable. In such occasions, I have envisioned music as being produced by beautiful dynamic systems having a life on their own, capable of creating rhythms and textures that expand the limits of our cognition. These moments have deeply impacted my creative practice, as I have been self-driven to seek ways to materialize such imagined music systems, to design artificial musicians that form a sort of automated organizations capable of generating beautiful music.

Pursuing these ideas, I joined a team of people from different institutions as the Music Technology Group, Johannes Kepler University, Red Bull Music Academy, Steim, ReacTable and Native Instruments which had started the European Community funded Giant Steps research project[1], where EDM was the central topic of research. Within this group I found a fertile ground to reflect and work on musical meta control and music intelligence, as means to amplify EDM creation and performance. This situation seemed ideal for working and learning with this new team, as its members had developed experience in materializing musical questions into useful machines and software applications. Once in this context, I confirmed rhythm and drumming as main transversal topics within EDM. Therefore, I directed my interest towards the human cognition of rhythm with the aim to learn from it and, once with that knowledge, use it to create new smart agents for the production and performance of EDM drum rhythms.

1.2 Context

Progressively, since its early days, a global and diverse community has flourished around EDM converting it into a cultural, aesthetic and

[1] http://giantsteps-project.eu

technological phenomena that mobilizes millions of people around the world. It is a hot cauldron bringing together a variety of actors: Scientists and engineers who design software, musical instruments and interfaces which are used by sound technicians, musicians, music producers and sound designers, for a variety of audiences, that go from big open air festivals to the intimacy of rehearsal rooms or porTable electronic devices. EDM is also disseminated in many different formats that range, for example, from the soundtrack of a video-game to a live transmission via the internet or a vinyl pressing. Specialized magazines, scientific journals, reporters and book authors provide news and critiques of the aesthetic, technological and cultural transformations of EDM to a captive crowd of music lovers, practitioners and dancers. The EDM industry, in all of its many dimensions, represents a huge and global economic cluster creating employment driven by technological and artistic advances (International Music summit report, 2017)[2].

One of the many reasons behind the consolidation of EDM as a global phenomenon is the evolution of the digital environments to create, transform and produce music. Nowadays, with a computer and an off-the-shelf digital audio workstation (DAW) software, anyone can get access to high quality tools sufficient to work in the creative side of the EDM industry. This possibility has opened the door of music production to individuals of diverse musical backgrounds thriving to get involved in EDM production, and aspiring to develop professional artistic careers. The accessibility of digital musical technology, has introduced a very low entrance fee to music production, in much cases favoring the success of individuals, despite their lack of formal musical training, in developing fruitful careers. Many established producers have not studied in a music conservatory, and some of these artists have concluded that the skills needed to become a successful EDM producer are not necessarily learned through formal musical training or at least not as it is currently institutionalized. This idea has been manifested in different ways by some

[2] http://www.internationalmusicsummit.com/business-report/

of EDM’s early pioneers:

> “*There are people who've been to college to study music and they can't make a simple rhythm track, let alone a hit record. It's weird [...] because now a little kid can pick up a computer, get lucky with it, and write a hit.”*
> Farley ‘Jackmaster’ Funk cited by Brewster & Broughton, (1999)

A similar idea is also noted by Marshall Jefferson when presenting the song ‘On and on’ by Jesse Saunders, as one of the most inspirational music pieces in EDM:

> “*That was the single most important record to me of the 20th century because it let the non musician know that he could make music.”*
> Marshall Jefferson, in Pump up the volume (Bidder,2001).

It is taken to an extreme by Stephen Morris, of the band New Order, when discussing the importance of learning an instrument as a prerequisite for succeeding with a career in EDM:

> “*The ones who will succeed are the ones who understand technology. You don’t need to be musical; musicality is actually a disadvantage.”*
> Stephen Morris, interviewed by Mills & Menagh (1990, p-80)

Without necessarily going to the extreme of Morris, it is a fact that contemporary musical tools have allowed gifted and dedicated individuals, as Marshall Jefferson himself, to be self-taught in the art of creating EDM at a professional level, producing tracks that transcend the home studio, having unimaginable impact in dance clubs all over the world.

The opinions of Funk, Morris and Jefferson convey a certain notion of distance from traditional music and academia to EDM, as a genre and also from its compositional practices. This notion is perhaps due to the functional prerequisite, the *dance* dimension, of EDM which implies its main focus is to make people move and it is achieved by the appropriate use of percussive rhythms (Witek et al., 2014a). Although some classical music forms owe their origin to medieval dances such as the Allemande, the Gigue or the Gavotte, we have to wait until the 20th century to witness rhythm and percussion become a central feature, in pieces as Ionisation (Varèse, 1931) or the Rítmicas (Roldán, 1930). However, these pieces were not intended be danced but to be listened to in the stillness of a concert hall. On the contrary, EDM is completely focused on rhythm and percussion, and is explicitly made for dancing, so it lays in a territory unexplored in classical music.

In fact, a proper EDM *dance track* is characterized by its finely crafted sounds and timbres and a powerful drum section designed for inducing body movement (Collins, 2008). Other musical dimensions such as the harmonic, melodic or even the structural ones, are many times non-existent in many EDM tracks or, in other cases, reduced to extreme simplicity. Chord progressions, melodic elaborations or counterpoint are definitely the least developed musical characteristic in EDM if compared to jazz, pop or classical music (Faraldo et al., 2016). This is why the main activities of a professional EDM producer are operating musical gear such as synthesizers, audio effects, sequencers and drum machines; transforming timbres and textures using that gear; and mainly, creating thriving *drum tracks*. As a consequence, the use of technological musical devices dedicated to drum synthesis and arrangement, such as drum machines, is essential as these instruments bear most of the weight of the structure and the dance functionality of EDM. This is clearly noted by Man Parrish referring to the use of drum machines by early EDM composers:

"*They're using this drum machine and it is actually a viable piece of equipment that you can actually make records out of it, because people hit the floor and dance to it."*
Man Parrish, in 808 documentary (Baker & Dunn, 2015).

This central role of machine-based drumming technology is acknowledged across different styles of EDM, from acid house music to hip-hop. Two of their foundational producers, Irwin Larry Eberhart II (the acid house artist known as Chip E) and Hank Shocklee, the producer of the hip-hop band Public Enemy, note in relation to the importance of drum machines in EDM:

[To compose an acid house song] "The first thing you have to do is to start with a strong kick drum and then you got to have a bassline. And from there you build on it with snare drums, you build on it with the hi-hat, you build on it with the rimshot, with the claps."
Chip E, in Pump up the volume (Bidder,2001).

"*In New York at the time man, every record had to have an 808*[3] *in it in order for it to have any sort of success in the dance floor."*
Hank Shocklee, in 808 documentary (Baker & Dunn, 2015).

There is no doubt that rhythm and drums incite body movement and people's urge to dance (Witek et al., 2014a), and as such, detailed care of percussion in EDM productions is microscopical. As will be discussed below, producing a *drum track* is a fine craft that goes beyond the use of technology, as very specific musical skills are required (contradicting the radical technology-above-musicality position of Morris), specially the

[3] The TR-808 drum machine was released by the Japanese company Roland in 1980. It was a foundational instrument for EDM production and performance.

skills related to the concepts of musical *style* and percussive *variation*. EDM is a very specialized music, meaning that within it, there are many different styles which are directly linked to the drum patterns used in the track (Collins, 2013; Butler, 2006). Each style (i.e. house, techno, breakbeat or trance, among many others) has in itself many subdivisions and ramifications, all of them with defined musical attributes. It is not surprising that, being the drums one of the most prominent feature of EDM, they have a big responsibility of conveying style information (Adamo, 2010; Brown and Griese, 2000; Emmerson, 2013; Hewitt, 2009; Snoman, 2012). Therefore, one definitive skill of EDM producers is to develop a deep knowledge of the boundaries between styles and the rules that define how drum patterns in a specific style are composed.

However, despite their importance, it is not easy to find structured guidelines to learn how a specific drumming style works, or even more, why a rhythm of a given style works. There is a lack of clear and simple music theory frameworks to explain and learn rhythm composition for EDM; tools that musicians could use when addressing drums, as it occurs for example with the circle of fifths when creating chord progressions. Instead, EDM drumming styles are learned by imitating the drum patterns in the records and in the production literature. In fact, the way an EDM drumming style is presented in production books (Adamo, 2010; Brown and Griese, 2000; Emmerson, 1988; Hewitt, 2009; Snoman, 2012), is by offering a single monolithic drum pattern as a means to introduce a musician into the comprehension of a given style. So it takes a great effort and, given the lack of musical tools, it is many times done as a trial-and-error procedure, to learn the features of a given style and then to imitate it making sure that a *drum track* is not only musically correct but also that it sounds as expected and makes people want to dance.

Taking a closer look, the work of creating drum arrangements in a DAW is called *drum sequencing*, which implies fixating single drum events, which are called *drum onsets*, in time. Drum sequencing is most commonly done through an interface with the metaphor of a *piano roll*, a multilayered time grid swept by a "playhead" at an isochronous rate, where layers (rows) symbolize instruments and steps (columns) symbolize time divisions. In this grid, a dot in a certain position represents a specific drum onset at a given time (see Figure 1.1).

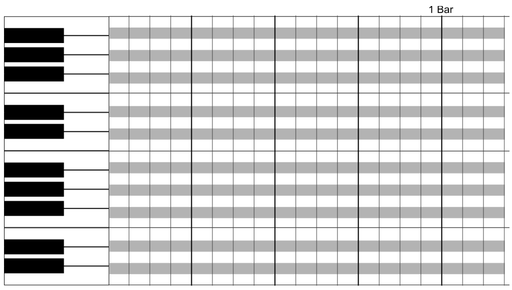

Figure 1.1 The scheme of a piano roll interface. On the left there are the disposition of the notes of a piano. On the right the space to place the notes of the musical arrangement where durations are depicted as rectangles of different length and dynamics as different shades in the rectangle..

Superficially, this representation suggests drum sequencing can be approached as placing beads (onsets) in graph paper[4]. But, in reality, the repetition of the sequence, the time resolution of the grid and the type of sounds being sequenced define how the position of the onsets affects our sensation and appraisal of a sequenced rhythm. In fact, as it will be presented in detail in Chapters 2 and 3 of this thesis, the effect of a drum rhythm goes beyond the graphic representation of events in time and is mediated by our cognition of rhythm. That is, a rhythm-processing mechanism is triggered in our brain, as a response to hearing a sequence of patterned sound events, determining the way this rhythm affects us, a mechanism which is invisible in the piano roll. This is the reason why a simple transformation as displacing a drum event a single step in a

[4] In fact it is very interesting to see how this piano roll sequencing metaphor has been repeatedly amplified to the physical domain in different interactive projects, proving both the conceptual success of the metaphor but also the difficulty to create rhythm sequences by trial-and-error "bead scattering".

sequencing grid can dramatically affect the sensation of a drum pattern or why, on the contrary, several onsets can be changed while still retaining the essence of a pattern. The way in which onsets affect the sensation of rhythm is not evident in the graphic representation of a drum pattern in a piano roll, as it takes higher level interpretations. So, paradoxically, the functionality of onsets in a rhythm pattern is sometimes unknown by musicians practicing EDM drum production, specially those which have not developed clear notions of rhythm. This problem is clearly addressed by DeSantis (2015) when describing the sometimes cumbersome process of drum sequencing:

> *"Your programmed drum beats tend to use the available instruments in "expected" ways: Hi-hats keep time, kick drums emphasize the beats, snares/claps are placed on or around the backbeats (beats 2 and 4). But in the music you admire, you sometimes hear other, more creative ways of working with drums. Sometimes it almost sounds like the drums are playing a "melody" of their own. But when you try to create patterns like this, it just sounds random and chaotic."* (DeSantis, 2015)

As mentioned, along with sequencing and understanding styles, the concept of variation is also fundamental in EDM and it comes in right after a promising drum pattern is created. When a short compelling rhythmic idea is defined, the next step is to construct a number of variations from it which are later assembled together, forming a *drum track*, in order to recreate a sensation of progress and evolution while strictly maintaining a central idea (DeSantis, 2015). An idea clearly exemplified by renowned EDM producer Jori Julkkonen:

> "*House music and Techno music, it's all about having this one bar looping endlessly and doing variations on that. For me that's the definition of House." Jori Julkkonen, in* 808 documentary (Baker & Dunn, 2015).

As suggested, an appropriate *drum track* in an EDM *dance track* is then a concatenation of patterns, carefully crafted and connected to maintain a emotive contour, offering diverse expressive cues to dancers. It is a delicately crafted sound structure directed straight to the sensorimotor systems of dance-avid crowds.

In practice, both essential concepts of drum style and pattern variation pose tremendous challenges for an EDM producer: It is hard to evolve a drum pattern template into a new one, trying both to infuse the identity of the author while maintaining its dance induction functionality. It is even more challenging when the intention is of a higher scale and the pattern must derive into a series of meaningful transformations in order to create the flowing musical structure of a *dance track*.

Some specific needs for EDM drum production are thus identified, which will be revisited throughout this thesis. EDM composers need to be skilled in the:

- Creation of drum patterns which comply to EDM styles, as that is the best way to ensure that a pattern is well-formed in a specific tradition and works for the purpose of inciting dance.

- Transformation of existing drum patterns and their concatenation in a higher order structure to create fluid *Drum Tracks* which are the support for a *Dance Track*.

- Performance of the above activities, imparting to them the producer's own character and style.

1.3 Existing Technologies for EDM Drum Production

Given the complexity of producing a *drum track*, different commercial products have been developed to help producers in this task. Some of these products come in the form of pre-recorded drum patterns to be loaded into a DAW, as a symbolic sequence or as audio recordings. Other software-based products allow the transformation of existing sequences. Pre-recorded musical material often comes in packages offering not only drum sequences, but also chord progressions and bass and lead melodies bundled by musical style. These packages are distributed royalty-free so they can be used straight to a commercial composition, with the obvious downside of potentially appearing in a lot of other musicians' productions and resigning some portion of the creative process. Although these products blur the line between musical creation and assembly, they serve the purpose of giving an uninspired, untrained or hasty musician, the opportunity to get some music done in a short time, with the advantage of being able to create music in a style without really having to understand it, and the disadvantage of musical fragments coming in finite numbers.

This way of composing music is based on coupling sequences of musical building blocks in different instrumental layers, without the need to understand their rhythmic or any other musical aspect. Composition becomes an exercise of free-form matching; again, as in sequencing, possibly undertaken as a trial-and-error activity which, this time, is limited by the number of building blocks available in the hard drive. Thus, as a consequence, there is a tendency to overpopulation of musical building blocks in a musicians' production toolkit, ranging from thousands of drum sound snippets to super-sized collections of drum, bass, chords and melodic sequences. Much as this compositional methodology serves the purpose of bypassing musical and stylistic knowledge, it still imposes huge constraints in archiving systems. Specifically, search mechanisms of musical material in a computer (i.e. the file finders) have no musical knowledge embedded in them, thus forcing the search of, for example, a drum pattern collection, to be alphabetical, ignoring the rhythmic or timbral properties in which a proper musical search should be based on. An activity as frustrating and senseless as searching a huge color palette by the color name[5].

Growing on top of those building block systems, other types of products are based on transforming this building blocks in real time. Software as Stylus RMX[6], offer a limited number of variations of a loop, meter change of a loop, changing the density of a pattern by eliminating or adding drum onsets, and transforming the micro timing variations of the onsets. All operations are however style agnostic so they are dependent only on the sequence itself without any explicit relation to a meta-structure as a style. Other tools for music composition are based on learning from the performer as the Continuator (Pachet, 2003), or modeling styles as the Drummer in Garage Band[7].

Drums are also being explored from different scientific perspectives which can potentially impact music production by expanding current knowledge on drum manipulation and therefore be implemented in new software applications. There are independent scientific communities devoted to the study of drums and rhythm, as automatic drum transcription from an engineering perspective; stylistic drum generation from the perspective of artificial intelligence; mental and aural processing of polyphonic drumming from a psychological and cognitive perspective. All this different scientific knowledge can eventually be implemented in current tools for EDM drum production leading to industrial improvements and the expansion of the features of percussive musical instruments.

---

[5] https://en.wikipedia.org/wiki/List_of_colors:_A–F
[6]https://www.spectrasonics.net/products/stylusrmx.php
[7]https://www.apple.com/lae/mac/garageband/

1.4 Opportunities and Challenges

As has been presented, EDM production offers a complex and rich scenario with needs from its practitioners, which are not solved in traditional music academies nor can be studied from manuals. This whole scenario is ready to be approached by new creative methodologies and specialized systems with knowledge on how an EDM rhythm is constructed in different styles, systems that can transform rhythms based on meaningful rhythmic frameworks and, that in general, can stimulate the creative process of composing an EDM track.

There is also an opportunity to improve computer based search mechanisms for musical building blocks, such as drum samples or drum sequences focused on musical qualities. Musically informed browsers, offering representations of drum sequence collections where musicians can explore the elements by their rhythmic qualities, could improve the accuracy, extent and depth by which a music collection is examined. This, in turn, would have a positive impact in the time spent exploring a collection, the depth of musical associations within a musical collection, and could thus improve the production process.

For achieving these potential tools, cognitive and perceptual perspectives on how rhythm, in the context of EDM, is processed by musicians, constitutes the main framework around which this thesis develops. This framework may inform new systems on how rhythmic material is selected, concatenated, contrasted, and later, used in musical compositions. As it will be presented in the next chapter, cognitive research on rhythm is still in expansion, so questions as the existence of metrics for assessing the similarity in monophonic and polyphonic percussion arrangements are still to be researched and clarified. Although the basic cognitive principles of how rhythm is processed are well known, there are still open questions on what influences two rhythms to feel alike and how much impact this has on the construction of EDM *dance tracks*.

There are aesthetic and technical advantages in researching EDM drum production as it allows for new creative possibilities to emerge which can support the amateur starting musician as well as the seasoned producer. Such systems could allow to focus on the flow of the creative process by exploring rhythmic alternatives faster and in new ways. Also, the idea of instruments that make use of the concept of style and have the flexibility of managing it as any other constitutive element (i.e. as it already happens with tempo or timbre), allows for flexible and novel music production strategies, which in turn can amplify the practice of EDM.

This thesis presents an attempt to improve the current tools used for EDM production, loading them with specialized knowledge on rhythm, so that this knowledge is available for musicians to use in their production workflow. In order to achieve this, current results in the fields of generative systems, rhythm cognition and music representation are used to carry out some original research.

1.5 Thesis Outline

In the following chapters I will present the work carried out to investigate how EDM composition systems, loaded with EDM rhythm knowledge, can be designed and constructed based on the scientific state of the art of rhythm cognition and generative algorithms.

Chapter 2 will present the research context in which this thesis is developed. Different aspects of rhythm will be reviewed, exposing how rhythm processing in humans is modeled from different perspectives, with a special interest in theories which study musical and percussive rhythm. The main aspects of human rhythm processing, as pulse meter and syncopation will be discussed, focusing on quantitative models used to describe them. One of the advantages of studying these models is the possibility to quantify properties that describe the way in which they are processed by our cognitive system. Later, that information can be used as a means to compare among different drum patterns' properties. This possibility leads to the study of rhythm similarity based on monophonic and polyphonic stimuli. Some unsolved questions on rhythm similarity will be identified and discussed, suggesting a research path that will be undertaken in Chapters 3 and 4. Starting from existing theories on conceptual similarity, we will advance the way they can help to formalize rhythm similarity in small dimensional spaces. With the tools and framework to measure similarity in rhythm, we will head our way to automatically generating meaningful and style-coherent rhythm sequences. The relation of generative tools and the concept of style will be discussed, exploring different engineering approaches to rhythm generation along with currently existing musical tools that implement such approaches.

A roadmap will be extracted, in chapter 2, that will trace a route for the development of this thesis. Experiments on monophonic and polyphonic rhythmic similarity will be evidenced as crucial to round up a theory that could link human similarity sensations with objective features extracted from drum patterns. These experiments and their results will be presented and discussed in chapter 3. Next, in chapter 4, the use of statistical modeling for the generation of short musical building blocks as drum sequences, will be explored. In this chapter, the implementation of style-based drum generative systems will be explained in detail, especially evidencing the ways in which the results from the experiments presented in chapter 3 affect the conception of these generative systems.

A final chapter will sum up the main outcomes and the contributions resulting from this research. These contributions will be considered in relation to the scientific context of chapter 2 and with the results presented in chapters 3 and 4. The research paths that I have not undertaken because they demanded a departure from the main topic will also be acknowledged.

2. STATE OF THE ART

2.1 Introduction

This chapter presents the scientific context in which this thesis is grounded. The ideas that establish the foundations for the development of this thesis are discussed, as well as the challenges they pose for advancing towards the construction of expert rhythmic musical agents. The chapter concludes featuring research activities that will be developed and discussed in depth in chapters 3 and 4.

The notion of rhythm will be the starting point, denoting the necessary components for a rhythm to be identified as such. The idea of rhythm is modeled by its relation to the human mind and body, as motion, time awareness and musicality. Although there are many modalities in which rhythm is manifested (as visual, tactile or auditory), this thesis is only concerned with auditory rhythm, activated by an acoustical signal, processed by the hearing system and passed on as electrical stimuli to the brain. A section dedicated to human rhythm cognition will present how rhythms are processed after being transduced from acoustical signals by our hearing system. The concepts of pulse, meter and syncopation will be discussed from different perspectives to define a common ground, to round up a unifying cross-disciplinary view. The review of these concepts will be presented in section 2.2 describing how they allow the emergence of quantitative relations between rhythmic patterns and other phenomena of rhythm cognition.

The possibility of extracting quantitative information from audio containing rhythmic patterns, opens the door to compare them, an activity belonging to the realm of rhythm similarity, which will be discussed in section 2.3. In this section, the notion of cognitive spaces and their role in similarity will be reviewed and main rhythm similarity studies will be discussed, reflecting on different experimental results that can lead to the construction of similarity metrics. It will be noted how explicit areas need to be expanded to achieve comprehensive metrics that can link human similarity sensations with objective information measured directly from rhythmic patterns.

In section 2.4 a general notion of musical style will be introduced and with it, the special case of style in EDM will be reviewed describing how the idea of style can be adapted to the distinctive characteristics of EDM. The special case of style in EDM drum sequencing will be reviewed in section 2.5 and generative approaches to drum sequencing will be reviewed in section 2.6. In this last section the main characteristics of diverse musical generative systems will be commented, presenting the type of algorithms used, their advances and limitations.

The last section of this chapter draws conclusions from each of the topics introduced. The conclusions section outlines specific knowledge that can be used for the design and construction of expert drumming agents. It also presents areas of knowledge and open questions that will be addressed during the course of this thesis. Explicitly, the conclusions section delineates the different activities that compose this thesis and which are presented and developed throughout the rest of the chapters.

2.1.1 What is Rhythm?

To start, the genealogy of the word rhythm is traced to the latin *rhythmus* that comes from the greek *ῥυθμός*. Different meanings are "a regular pattern of change, especially one that happens in nature" (Cambridge dictionary), "a strong, regular repeated pattern of movement or sound" (Oxford dictionary). Other interpretations of the word *ῥυθμός* specially derived from philosophy are “the particular manner of flowing” (Adkins, 1962). A more precise interpretation regarding how a musical *pattern* becomes evident by repetition is given by Schuback (2003): “Rhythm is discontinuity in continuity. What in music is called rhythm is properly an unrhythm, that is, a ‘break,’ an interruption, a rift, a breathing or caesura of and in continuity”. For this thesis musical rhythm is going to signify the articulation between sounds and silences which, by their structure, allow to elicit regularity, repetition and thus a sense of musical flow.

2.2 Human Rhythm Processing

Different disciplines study the way in which humans process rhythm. On one side there is auditory cognition which seeks to understand the mechanisms involved in recognition and processing of acoustic stimuli; there is also neuroscience, which explores brain activation patterns; and from a musical point of view, there are explorations from music cognition on how the specific case of musical rhythm is structured on listeners. All these views try to come up with an explanation of the processes involved in triggering and representing a “rhythmic sensation” derived from an acoustic stimulus that contains strong traces of repetition, with the special cases of music and speech. There is a general agreement among these disciplines on the basic elements involved in rhythm processing which are: the emergence of a periodic *pulse* sensation based on a synchronization with the different rhythmic levels of an acoustic event, along with expectancies and anticipations of the future development of the acoustic signal; the emergence of a *meter* in the form of perceptual accentuations of certain events in the signal; and finally, a general emergent sensation of balance/predictability or imbalance/surprise based on how the incoming auditory events comply (or not) with the the predictions and expectancies at different levels. There is also concurrence on modeling rhythmic processing as a feedback loop where the current acoustic events are used to predict the next cycle of the rhythm and then testing what is predicted against the incoming acoustic signal and adjusting the prediction if necessary. Even more, some suggest these rhythmic mechanisms are innate but refinable through experience and musical training (London, 2012). Once this iterative process of pulse acquisition, metric entrainment, anticipation and testing is set in motion, its outcome can be used in other high level processes such as dancing, playing an instrument, attending a concert or discerning a spoken dialect.

2.2.1 Pulse and meter

Representations of temporal aspects of music can be derived after some "regularity" in the flow of events is detected. Such regularity sensation is an emergent perceptual phenomenon called the pulse, which is evidenced when we, voluntarily or not, end up tapping the foot or nodding the head to a music piece. Even when an acoustic stimulus is unaccented, as the *tics* of a clock where all events are identical, our mind imposes some sort of organization, arranging the potentially infinite indistinguishable pulses into structured groups and thus transforming a "*tick, tick, tick, tick*" sequence into a more structured "*tick, tock, tick, tock*" and imposing a grouping sensation which is not explicitly present in the signal (Bolton 1894; Meuman, 1894; Cohen, 1957, page 136; Cooper et al, 1963). It has been shown that this capacity of *entraining* to a pulse is present since childhood (Drake et al., 2000), even in babies which synchronize sucking to auditory stimulus (Pouthas, 1995), and also on animal species as sea lions (Cook et al., 2013) or cockatoos (Fitch, 2013; Patel et al., 2009). There is neural evidence of entrainment, as neurons in the auditory cortex have been observed to synchronize with periodical acoustic stimuli (Nozaradan et al., 2011, 2012), and this entrainment may underlie cognitive functions such as the perception of beat in music (Lehman et al., 2016). However, entrainment is not entirely a bottom up process built directly from the analysis of an acoustic signal by our brain's time keeping mechanisms. There might be an alternative mechanism based on long term memory, in which a repertoire of entrainment responses is stored and recalled when subjects are exposed to a rhythm. It is suggested that this memory mechanism is the first one triggered, when adapting to a rhythmic acoustic signal and decoding the periodicities of a signal. Then, if the metric repertoire fails, the mechanism for deducing the pulse form periodicities is activated (London, 2012). There are also experiments on preference for double or triple metrics in babies, depending on the metrical dominance in their mother tongues, suggesting language modulates a cognitive bias to select the firsthand hypothesis to entrain to a pulse (Patel & Daniele, 2003).

As mentioned, music is a special case of an auditory event, as it most frequently comes in the form of sound sequences where pitch, timbre, amplitude and time dimensions are organized in such a way that it is capable of eliciting a pulse sensation. This works as in the *ticks* of the clock, but has additional perceptual and cognitive effects as the definition of a melody, the sensation of a structure or even tonality. There are different models that intend to explain the processes that are triggered in the presence of a periodic sound source as music. Some of those models address note duration (Longuet-Higgins and Lee, 1984), pulse saliency (Parncutt, 1994) or dynamic attending (Jones and Boltz 1989), using techniques as neural networks (Gasser et al., 1999) or explicit rules (Desain and Honing, 1999; Eck, 2001). None of these models is bound to specific neural processes, although there have been found neural responses in the auditory cortex synchronizing with periodicities of an acoustic signal and maintaining the periodicities, even in the absence of a stimulus. That is, acoustic periodicities trigger a timekeeping mechanism that, once set in motion, is constantly looking ahead, inducing neural activity in parallel, thus not completely dependent, with the music. It is a cognitive mechanism that goes beyond the surface of the sound, creating new information derived from the acoustic stimulus but not necessarily present in it. The neural activation signal is not a single pulse, but rather peaks at different frequencies related metrically to the acoustic stimulus, but all having a frequency relation with the pulse sensation (Lehman et al 2016).

This neurological confirmation of neural peaks present at frequencies related to an auditory event is very akin with the Dynamic Attending conceptual model (Jones, 1976, 1987, 1990), that describes the effect of auditory stimuli on time perception, based on musical examples. From a neurological perspective there is evidence of a synchronicity of neural activity with musical rhythm, which is phase locked to the pulse and interconnects the auditory cortex with motor areas of the brain as the basal ganglia (Grahn and Brett, 2007), even without movement being involved (Merchant et al., 2015). This neural activation is believed to function as an anticipation mechanism which could help predictive movement planning in time-sensitive scenarios (Fujioka et al 2012). Other marks of rhythm-related neural activity have also been found in subjects evaluated in non-attentitve scenarios independently from their degree of musical expertise. This activity suggests these neural responses to be general and autonomous, irrespective of any musical practice (Bouwer et al, 2014) (Fischer et al., 2010). It seems then, that, provided a short exposition to musical input, from 5 to 10 notes according to Desain and Honing (1999), musicians and non-musicians generate a neural predictive model, even when ignoring auditory information by being inattentive to music (Honing, 2012).

The Dynamic Attending model, developed by Jones (1976, 1987, 1990; Jones & Boltz, 1989) and further refined with some collaborators (Large and Jones 1999; Drake, Jones and Baruch, 2000), proposes that rhythm processing is based on an array of available oscillators which synchronize at different frequencies, equal, above and below a reference level (usually the pulse), when a periodic stimulus is present. The synchronization is based on the *attunement* of the listener, who becomes phase-locked to different time spans explicitly marked (or not) in the periodic stimulus, one of them being the pulse. In short, one of the oscillations activated with a periodic acoustic event, as a musical piece, is synchronized in phase and period with it and is the carrier of the pulse sensation. Above this period (i.e. with a lower frequency) other oscillations can be activated and, by shifting the attention to them, higher structural properties of the acoustic event, grouping various pulses such as the end of a musical phrase, can be anticipated. Attention can also be shifted below this period and a so-called *analytical attunement* occurs which is related to events at the sub-pulse level as timbral variations in a conga pattern or syllables in spoken word.

The dynamic attending model is based on the possibility to shift the focus of attention between the different referent levels. This implies a cognitive limitation encountered in the complexity of the sound that is being attended: when the auditory events become more complex to process, attunement becomes harder, thus affecting the capacity to perform attentive shifts, where only one or two levels remain constant and the others become blurred. This capacity to attune to different focal levels shifts with age can also be the consequence of a training process as an individual can learn the periodicities of a complex rhythm and identify acoustic markings that signal phase-locking (Jones, 1976). Pulse entrainment can be seen as a process of dynamic attending, which is by definition the focalization to most salient temporal locations for events, accenting some of the most relevant acoustic activity at a given time. However, the unaccented events are not ignored, they are "reorganized" generating a sensation that is called *meter*: the superimposition of periodic accents to the sounds in an acoustic signal and the consequent groupings (London, 2012).

From a perceptual perspective, meter elicits only two metrical levels, accented and unaccented events, which affect the perception of an acoustical signal. This implies meter is nothing but a binary measure which differentiates between *strong* and *weak* sound events, as when one counts a ternary division of a pulse as "**ONE**, two, three, **ONE**, two, three", where the ONEs are strong while two and three are weak (London, 2012) (see Figure 2.1). This metric entrainment seems to be universal, however the quality of the entrainment is modeled by the user itself by its own "ability to generate metric patterns (an ability that may vary with age, talent, training, and enculturation), and the lack of interference from subsequent musical stimuli (interference here meaning the emergence of a pattern of alternate metric cues)" (London 2012).

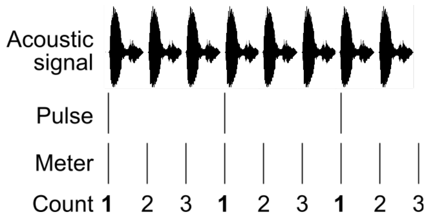

Figure 2.1. A repetitive acoustic signal (top) and the hypothetical induced pulse (center) and meter (bottom).

From a theoretical musicological point of view, Lerdahl and Jackendoff (1985) propose a more nuanced metric structure composed of more than two (strong and weak) types of accents. In their General Theory of Tonal Music (GTTM) they present a hypothetical multileveled model of metrical hierarchies that is based on the same nesting principles as the dynamic attending model (a reference level can be concatenated or subdivided in predefined proportions, see Figure 2.2 left). In their model, consecutive pulses are grouped while also preserving the metrical subdivisions of the pulse (i.e. binary or ternary), creating an evenly distributed time grid that spans for several pulses, in which the duration of the smallest period is the same as the time elapsed between lowest metrical events. In this structure, metrical accentuation depends on the relation of each accent with the highest metrical grouping. Although completely theoretical, this nested metric presents an amplified version of the strong-weak psychological differentiation of meter, where the inter-pulse activity is weighted depending on their position within the pulse, and the weight of the pulses are also dependent on their position within the above-pulse structure.

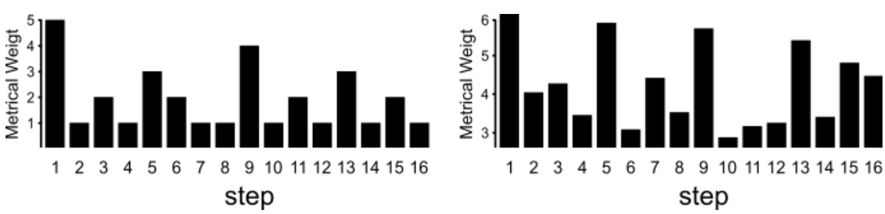

Figure 2.2. Metrical weights as proposed by LHL (left) and as found by Palmer and Krumhansl (right).

In order to clarify the existing ambiguity between the binary and the multileveled views of meter, evidence has been gathered from experimentation with musicians and non-musicians, confirming how "listeners, like composers, have abstract knowledge of multileveled metrical hierarchies in musical composition" (Palmer and Krumhansl, 1990). Based on the results of three systematic experiments, Palmer and Krumhansl provide experimental support for the existence of a metrical structure similar to the one proposed by Lerdahl & Jackendoff (Figure 2.2 left). Their analyses evidence different levels of metrical hierarchies in between strong and weak accentuations, specially in the context of western classical music. It still remains a question if these structures exist beyond western tonal music, or if music from other cultures is grounded on different metrical principles, and as such imposes different perceptual structures to listeners and practitioners.

2.2.2 Syncopation

Once the mechanisms of pulse and meter entrainment superimpose sensations of regularity and hierarchy, as a natural and predictive behavior when stimulated with a periodic musical acoustic event, a rich interaction between the temporal elements of the music and our hierarchical temporal expectations takes place. The most relevant interactions related to rhythm are caused by a complex and ongoing interplay of confirming and challenging the elicited expectations or hierarchies. As the sensory activations of rhythm seem to be a natural human phenomenon, confirmation-and-challenge interplay is a cross cultural phenomenon also experienced in other musical dimensions different from rhythm. In the rhythmic realm, the concept of syncopation is related with the challenge or reinforcement by the music of our inner pulse and meter predictions. This concept is important as syncopated rhythms are said to elicit higher level musical responses as pleasure, motor activation and even dance (Witek et al., 2014a). Syncopation is based on the idea that metric hierarchies within a rhythmic event are an expression of the expectancy to perceive a sound; that is, once we are metrically entrained we have higher expectations of incoming sounds to be temporally aligned with higher metrical positions. When the opposite occurs and a sound is aligned with a low metrical position and a silence is aligned with the next high metrical position, a syncopation is said to occur (Longuet-Higgins and Lee, 1982) (See Figure 2.3).

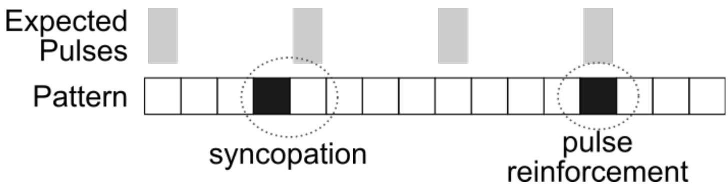

Figure 2.3. Pulse expectation and syncopations. Positions of expected pulses in grey, notes in black. A syncopated note on the left and a pulse reinforcing note on the right.

2.2.2.1 Monophonic syncopation

One of the first systematic studies on syncopation (Longuet-Higgins and Lee, 1984), proposes a formula to measure the syncopation of a monophonic musical phrase based on weights which are assigned to each note in a musical phrase. These weights are analogous to the metrical hierarchy of Lerdahl and Jackendoff, which is defined as a nested structure: "the weight of a given note or rest is the level of the highest metrical unit that it initiates. (The level of the topmost metrical unit is arbitrarily set equal to 0, and the level of any other unit is assigned the value n-1, where n is the level of its "parent" unit in the rhythm)" (Figure 2.4 top).

Syncopation is then defined with a formula: "If R is a rest or a tied note, and N is the next sounded note before R, and the weight of N is no greater that the weight of R, then the pair (N,R) is said to constitute a syncopation (Figure 2.4). The "strength" of the syncopation is the weight of R minus the weight of N" (Longuet-Higgins and Lee, 1984). Given that a metrical hierarchy is established, weights can be assigned to each note and rest within a musical phrase based on its position: if a note is followed by a rest in a position of a higher metrical weight then a syncopation occurs (Figure 2.4). Syncopation is a violation of the expectancy to find a note at a high weight position, which is replaced by a silence.

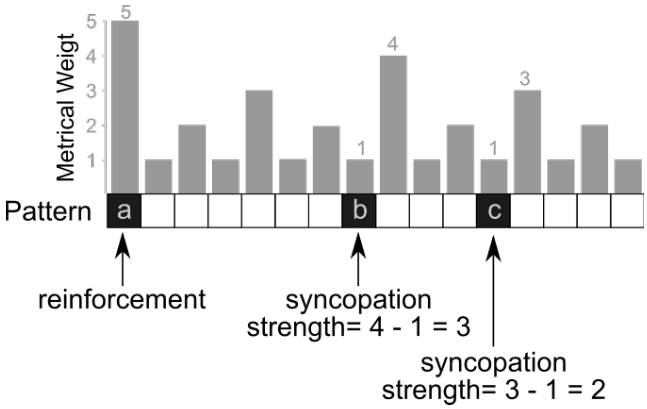

Figure 2.4. A syncopated pattern composed of three notes a, b and c. Note a is not syncopated as it is reinforcing the pulse. Note b is syncopated with a strength of 3 and note c is syncopated with a strength of 2.

Syncopation also goes beyond the realm of music theory and has been proven to elicit, in the auditory cortex, specific neural patterns that influence motor and synchronization functions. Fitch and Rosenfeld (2007) found psychological relevance of Longuet-Higgins and Lee's model (LHL) of measuring syncopation. They experimented with subjects, having them perform different rhythm tasks of pulse tracking, rhythm reproduction and delayed recognition, and found that LHL's metric was a strong predictor of the participant's performance. In their experiments, as the syncopation index increased, the average performance of the subjects in the different tasks decreased in accuracy. They conclude that syncopated rhythms in general, were more difficult to perceive and produce by musicians, and that measures of "complexity" increased with increasingly syncopated rhythms. On the same path, brain activation peaks have been measured when a prediction of an upcoming rhythmic event is generated but is violated by the actual acoustical event, with the magnitude of the peak being proportional to the magnitude of the violation (Näätänen et al., 2007; Polich, 2007). The musical concept of syncopation as proposed and measured by LHL is found to influence sensorimotor performance and to be a cause of arousal in the timekeeping mechanisms of the brain.

2.2.2.2 Polyphonic syncopation

The research mentioned so far has a remarkable particularity: it deals with monophonic syncopation only. This condition is a very straightforward way to present an acoustic signal to subjects in an experiment, but it is also a reduced version of how real music is experienced. It is completely fair to use such material for achieving a great deal of experimental control of the stimuli, but, from a musical point of view, it leaves few clues on how different layers of instruments reproducing different rhythms, which in turn induce and challenge a beat, are processed. Such is the case of real life polyphonic drum ensembles, either acoustic or synthetic, as they are more likely to be present in EDM in the aforementioned *drum tracks* and *drum breaks*.

Although there is still much room for the comprehension of polyphonic rhythm phenomena, there are some studies that have focused on syncopation using polyphonic real-life musical rhythms. Bouwer et al. (2014) used a polyphonic music-like drum-kit stimuli, omitting the kick drum and hi-hat at different metrical levels, and evaluating the relationship between neural activity and omissions. Their results report an excitation of neural activity when rhythmic expectancies were violated, with a magnitude relative to the metrical weight of the violation, and related to the type of sound that was omitted causing the violation. In their study, violations on the beat elicited higher activity than violations on less metrically salient positions related to the beat. They also report a less prominent activity when the violations were produced by hi-hats alone, than the combination of hi-hat and kick drum. This finding is revealing, as it is aligned with the results of other monophonic studies which report excitation as a function of the position in the metrical hierarchy (Fitch and Rosenfeld, 2007) but adds another Issue which is the possible effect of the type of instrument that is omitted. Interestingly, the study introduced the effect of the omitted instrument in the overall picture of polyphonic syncopation. These findings suggest that the omission of a hi-hat generates a peak of a lesser magnitude than that of a kick and a hi-hat, thus introducing the frequency dimension, specially the frequency range of the instruments, to a conceptual system so far explained exclusively by means of pulse, meter and confirmation-violation of the acoustic signal. However, their findings cannot be extrapolated to other instruments or combinations of them, as their respective displacements were not evaluated.

In the same context, but starting with a metric for polyphonic syncopation derived from drum arrangements, Witek et al. (2014a) evaluate the relation of polyphonic drum syncopation and motor activities, specifically dance. They conclude that a "medium" dose of syncopation in real-life polyphonic drum patterns (specifically *drum breaks* as mentioned in Chapter 1) magnifies the desire to move and dance to the music. In their study, they use a special computation of syncopation which is based in Longuet-Higgins and Lee's formulation, expanded to fit polyphonic scenarios. Their original formulation of polyphonic syncopation is supported by laboratory observations which are similar to those of Bouwer et al. (2014). According to Bouwer et al., the frequency range of the instruments present on a drum pattern influence how their syncopations are processed in a polyphonic context, specifically attributing more importance to instruments with a predominance of low frequency components (i.e. the kick drum) (Witek et al., 2014b). Their view comes from an ecological perspective in which instruments composed of low spectral components are more disruptive rhythmically: "darker (low frequency) sounds are cross-modally associated with larger and heavier sound sources that are more likely to be close to, or on, the ground (e.g., a bear), while brighter (high frequency) sounds are associated with smaller and lighter sources that may spend more time off the ground (e.g., a bird) (Maurer & Mondloch, 2004)." Aligned with this view, Hove (2014) presents evidence of the higher influence of low frequency instruments in terms of expectancy violation and thus syncopation, observing larger responses to time deviations of the lower pitched stream of two different parallel rhythms. Hove also found stronger auditory-motor synchronization for low-pitch sounds than for high-pitch sounds, an idea also backed experimentally by Burger et al. (2017). These findings suggest that a syncopation sensation induced by a high-pitched sound should cause less disruption to metrical stability than a low-pitch syncopation. Thus syncopation in a polyphonic scenario is influenced by the type of sound that challenges the induced pulse, adding a second dimension to its quantification: one is the place of the metrical disruption and the second is the frequency range of the sound producing it.

It is very significant that polyphonic syncopation has been related to higher level musical concepts as pleasure and desire to move (Witek 2014a). This gives a clue on how complex musical concepts can be broken down into simpler, measurable and straightforward factors which can be observed in a polyphonic drum sequence: pattern, pulse, meter, frequency range, syncopation and desire to move.

2.3 Similarity in Percussive Rhythms

In the previous section it was presented how a *drum break* can be analyzed in order to extract high level rhythmic information from it, which is associated with human sensations such as the degree of syncopation or the desire to move. This rhythmic information can be used to study the connections between different drum patterns. Specially, how the human sensation of two patterns feeling similar could be related to the objective rhythmic information that can be extracted from their musical, musicological or acoustic analysis. To extract objective attributes from stimuli (whatever their modality) and explain with them the similarity sensations between these entities, is an activity in which different processes converge. On one hand, there are cognitive and perceptual studies which seek to define the most important attributes of a stimuli from a human point of view. On the other hand, there are many different formal approaches to similarity between entities, for example, measuring the common and uncommon features, or defining geometric relations among them (Tversky, 2004). This section, therefore, presents the ideas of similarity used in this thesis as well as different approaches taken by researchers to measure rhythmic similarity, both in monophonic and polyphonic scenarios. In general, humans take advantage from establishing connections between entities, which are interpreted as similarity. These connections play a major role in our daily life as they allow us to generate new ideas or discover new relations between concepts, sometimes beyond their superficial appearances.

2.3.1 Notions of similarity

As described by Gärdenfors' (2004) theory of conceptual knowledge, similarity allows humans to compare concepts among them. Similarity serves the purpose of structuring concepts as mental representations based on their descriptive dimensions. Under this model, similarity is the thread that holds together the conceptual structures that we use constantly to process the world we sense and imagine. It has been proposed that similarity is not static but that it is instead malleable and dynamic, as it responds to the perspective by which a mental representation is considered (Ramscar & Hahn, 2001). For example, under certain considerations the concepts *bicycle* and *car* can be similar among them, and completely dissimilar to *chair* if we reflect on them as means of transportation, but instead a bicycle and a chair can be very similar and both very dissimilar from a car if we are weighting them on a Kilogram scale. Concepts have attributes which can be used to classify them in specific categories (i.e. as means of transportation) or to compare them given specific measures (i.e. as in a Kilogram scale).

Hampton (2001) proposes two different types of concepts, one being cultural constructions (i.e. science and history) and other being personal mental representations. Alternatively, Gärdenfors (following Gallistel) classifies concepts as being either scientific or psychophysical (Gädenfors, 2004, page 8). Psychophysical representations imply that “each individual may be using a somewhat different schema for representing the concept, and may defend his or her right to consider it to be correct. The psychological question therefore becomes that determining what are the mental representations of concepts that people use in everyday life” (Hampton, 2001). On the other hand, scientific representations are formed from “the basis of people’s concepts most of the time, and that some individuals, with a lot of training and with the advantage of the cultural transmission of ideas from great thinkers of the past, are able to develop more advanced thinking skills in particular domains.” (Hampton, 2001). In the context of this study it is to be proved how much of the compositional process of a musician is influenced by cultural characteristics, or by acquired taste and experience, and how much is determined by the music rhythm processing of her brain. It is plausible to think that there is an interaction between these two types of concepts (and most likely some others), when a compositional process is at play. Modeling these concepts is determinant for a robust artificial EDM listener and composer, as the one that is being proposed here.

Under certain circumstances, elements defined by specific measurable attributes or dimensional characteristics (as weight, speed, frequency or syncopation strength) can be located in a conceptual space that has as many dimensions as measurable attributes. Each element becomes a point in a common space, a conceptual space, forming relations based on the similarity of their characteristics that allow to elaborate higher level abstractions about that conceptual space. For example, in a fruit space, one could say “a lemon is as similar (or as different) to a tangerine, as an apple is to a pear”. Such statement, whether we agree on it or not, implies quite a number of higher level abstractions that can be traced down to the idiosyncratic use of similarity among fruits, and its intrinsic evolution into a geometrical construction. It means that the similarity between the lemon and the tangerine is comparable with that of the apple and the pear, which means that the magnitude of both distances in a conceptual space is alike. This in turn, can be interpreted as that the characteristics of each fruit, in this common fruit space (as the color, the shape, the sweetness, the acidness, the size, the ruggedness of the skin, etc), determine that the measurable distance between lemon and tangerine is similar as the one between apple and pear. This however does not imply that in every dimension of the fruit space both pairs of fruits are separated by the same values, (i.e. same sweetness distance, same color distance, etc), it just implies that after considering all the available dimensions, their final distances are somehow similar. It is important to emphasize that if such comparison was part of a conversation where someone was commenting on how fruits are alike, such statement would have emerged automatically, without having to consider dimension per dimension in an exhaustive way, and it would be an automatic expression of the geometry of an inner fruit space.

Adopting this framework for reflecting on concepts, adds the possibility to group them and define geometrical boundaries around them, which lead to the construction of categories. Following the example above, one could define that a lemon and tangerine are alike and that they both coincide within the region of the citrus, as the pear and the apple are bounded inside the region of the pomes. In this example, there would be an overlap between the two types of spaces defined by Hampton (2001): a psychophysical space (where one's similarity is originated) and a scientific space where concepts have been studied, measured and classified in an organized way, defining regions as citrus or pomes. Conceptual spaces, as the ones presented above, have been successfully used in different musical domains (e.g. timbre or pitch) to structure the relations between elements.

The way in which this framework replicates human cognitive aspects is to be founded on the idea of similarity as being dynamic, thus allowing to compare a group of concepts from different perspectives which, in its more literal sense, implies a geometrical shift. To illustrate this point, let us recall the fruit space. Under one perspective, if we think of flavor, one can agree that pears and apples, which are sweet, are closer among them than lemons and tangerines which are citric, which makes them closer among them and distant from sweet fruits. However, if we decide to give a higher importance to the color dimension, then green pears and lemons would be more similar among them and tangerines and apples, being orange and red, would be more similar among them than to green fruits. This simple example shows how similarity, and thus the conceptual spaces based on it, are not of a sTable nature, but on the contrary, suppose a dynamic structure rooted on how the descriptive dimensions of elements in the structure are weighted. Similarity has a malleable nature which is tuned by adjusting the importance of the dimensions as we did by shifting from taste to color. "In order to obtain a cleaner and more generally useful set of categories, we may adjust the weights of dimensions, and even construct new dimensions from which to build concepts (Hampton, 2001)." The application of these concepts in music similarity research is relatively recent, (Aucouturier & Pachet, 2002; Logan & Salomon, 2001; Pampalk et al., 2005) but its impact in terms of active users in contemporary digital music reproduction platforms is unprecedented[8].

A definition of similarity must leave space for the different characteristics of two elements and also keep the flexibility to change given the importance assigned to some of those characteristics. Two families of similarity models are reviewed by Keane and Smyth (2001): the first one is the contrast model derived from a model proposed by Tversky (1977), where they argue that, to measure the similarity of two entities, one must group on one hand the features that are common to both and on the other their distinctive features, all weighted independently. This model is characterized as:

$$s(a, b) = \theta f(A \leftrightarrow B) - \alpha f\ (A - B) - \beta f\ (B - A)$$

Where A and B are the set of attributes of entities a and b. The term (A ↔ B) represents the set of attributes common to A and B, (A - B) represents the distinctive features in A and (B - A) the distinctive features in B. θ, α, and β are parameters that reflect the importance of the common and

[8] See for example music streaming systems as Pandora or Spotify.

distinctive attribute sets. The function f is a measure of the salience of the features or sets of features of both entities. Another type of similarity is the alignment model, proposed by Keane and Smyth (2001) where they emphasize on the dynamic aspect of similarity, especially on how the cognitive processes used to apprehend each of the entities influences their similarity, explicitly how similar those processes are. In this model, as in Tversky's, similarity is divided in three groups: the commonalities (the characteristics that match between both concepts), the non-alignable differences (the characteristics of one concept not present in the other) and the alignable differences (both concepts share the same characteristic values, but their values differ). In general there are many mathematical models for computing cognitive similarity (Verguts et al., 2004).

Similarity forms then the basis of our system of concept relations. Thanks to similarity, concepts can be kept entangled into geometrical constructs that are dynamically updated with new experiences and by changing the way we weight the dimensions that describe those concepts. Concepts "play a central role in everyday behavior and action, they permit predictive inference, they are a necessary building block for acquiring and using knowledge of the world. Concepts evolve in order to maximize their general utility value, according to some (as yet unknown) criterion of utility" (Gärdenfors, 2004).

In the context of understanding EDM drum rhythms (and creating smart agents which are able to generate them), the concepts of similarity are fundamental. They allow to establish quantitative relations between drum patterns. They also make possible to understand and visualize the affinity between drum patterns of different genres or among patterns generated by a rhythm-expert agent, or even to compare the patterns generated in different styles. Defining objective metrics to measure similarity relations between rhythms, aligned with human similarity sensations, is crucial in order to work in the domain of musical drum patterns.

2.3.1.1 Dimensional reduction techniques

A typical way to grasp the relationships between concepts with many attributes in an easy-to-visualize two or three-dimensional graph, is the use of dimensional reduction techniques. These are a family of algorithms specialized in converting high dimensional (multi-attribute) spaces into small dimensional spaces, minimizing the difference between the multi-dimensional distance and the low-dimensional distance between the elements of the graph. Commonly, visualization of data is based on low-dimensional structures (2D or 3D), while the information displayed can convey more dimensions by resourceful use of color, form and symbols (Keim, 2002).

In cognitive science, a low dimensional cognitive space is obtained from subjective similarity judgments generated by subjects who compare pairs of instances of a specific domain or type (i.e. fruits, sounds or rhythms). The data generated from such comparisons is processed by dimensional reduction algorithms which output low-dimensional maps representing the similarity relations between the compared instances. These cognitive maps help researchers observe implicit characteristics of how a given domain is processed by humans. The purpose of these maps is to make observations of the relationships between the instances of a domain which are not evident in the pair-wise similarity judgments.

One of the most common dimensional reduction techniques is Principal Component Analysis (PCA). This technique is used when all the elements of a given set have multi dimensional attribute values. PCA consists on finding the multi-dimensional axis (a weighted linear combination of the attributes describing an instance), where the attributes are scattered the most. That axis is called the *principal component* and is composed of weight values for each of the attributes. Additional components can also be added in order to achieve two-dimensional or higher dimensional representations of the elements in the set. The only requisite for the additional components is orthogonality with the other components. So, for example, the search for the secondary principal component is a search around all possible axes which are orthogonal to the main principal component. The secondary component is that in which the attributes are maximally dispersed (Wold et al., 1987).

Other techniques have been developed for the same purpose, such as multidimensional scaling (MDS) (Kruskal, 1964). This technique is based on having all dissimilarity values between each pair of elements of a given set, thus obtaining a dissimilarity matrix. Then, MDS is used on the matrix, specifying the desired dimensionality of the expected resulting space. Finally, the result is a set of coordinates for each element on the set. Unlike PCA, MDS' main goal is to preserve the high dimensional distance among the elements in the space. This methodology is widespread in cognitive sciences and is the foundation of contemporary understanding of many domains such as color (Shepard, 1962), timbre (Grey, 1977), pitch (Krumhansl, 1979) or tactile textures (Hollins, 2000).

2.3.2 Monophonic rhythmic similarity

Having introduced similarity and conceptual spaces as a model in which concepts, such as rhythm patterns, are structured given their characteristics, the idea of rhythmic similarity can be considered. Different authors have studied ways in which similarity sensations between two rhythmic patterns relate to objective properties of the patterns themselves (Paiement et.al., 2007; Cao et. al., 2014; Post & Toussaint, 2011). Experiments are either focused on monophonic or polyphonic rhythms. Some are based on extracting cues from audio recordings, while others rely on the symbolic renditions of the patterns. There are also two main frameworks in which similarity is conceived. One is built upon the ideas revealed by psychoacoustics, cognition and neuroscience, where the sensation of similarity is seen as a consequence of rhythmic processing as exposed above: related with pulse, meter and the violations and reinforcements of the emergent hierarchies. The other framework is inspired by information theory measurements of similarity in data sequences.

An interesting body of work has been created exploring the use of information-inspired metrics for predicting similarity sensations between two rhythms, measuring differences in their inter-onset intervals (IOI), or the time between the start of each note onset and the next. Different metrics such as the edit distance (Post and Toussaint, 2011; Toussaint et al., 2011), many-to-many, one-to-one distances (Toussaint and Man Oh, 2016) or the Hamming distance (Paiement et al., 2007), have been considered to be useful candidates to measure similarity in rhythms. However, no empirical data connects these metrics with theories of rhythm processing in humans, from either a psychoacoustic or a neurological perspective. The interesting properties of these agnostic metrics are their common use in any type of application, disregarded of a cognitive context. They are used for measuring the distances between strings of symbols with no perceptual or musical relation, as they are unaffected by meter, pulse or syncopation whatsoever. Several papers have been devoted to the use of the edit distance in rhythmic similarity contexts from a conceptual and also from a practical point of view, suggesting it captures the similarity elicited when comparing two rhythms (Post and Toussaint, 2011; Toussaint, 2004; Toussaint et al., 2011). But other authors working on different rhythmic metrics, are at odds with the use of the Edit distance for how the metric is defined (Paiement et al., 2007; Cao et al., 2014). The edit distance is popular for measuring distances between strings and it is widely used in linguistics and DNA sequencing analysis (Kim et al., 2013). It is based on measuring the amount of transformations that must be applied to one string to become another, allowing three types of transformations: character substitution (i.e. *abc* for *adc*), insertion (i.e. *ab* for *abe*) and deletion (i.e.*abcd* for *acd*). What has been pointed out by the critics of the edit distance is the use of insertions and deletions which, with a single transformation, can dramatically alter the way a rhythm is perceived (e.g. a reinforcement becoming a syncopation).

For example, a sequence composed of a series of onsets on the pulse on a binary meter, can become a highly syncopated rhythm by an insertion operation (see Figure 2.5). This has been clearly pointed out by Paiement et al. (2007), arguing that rhythm perception heavily depends on the position on which rhythmic events occur and, instead of the edit distance, they opt for the use of a Hamming distance between rhythms. The Hamming distance on its side, is also a rhythm-agnostic metric, but makes no use of insertions or deletions at a low distance cost, as it is the case with the Edit distance. Despite a lack of rhythmic theoretical foundation, the logic behind these agnostic metrics is *syntactical*, as they are concerned with the orders of the onsets and silences within a sequence. It might be the case, however, that the claims of predictive power of these agnostic metrics metrics apply in specific scenarios where memory processing is more relevant than the presence of a pulse and meter, such as in the absence of an induced pulse.

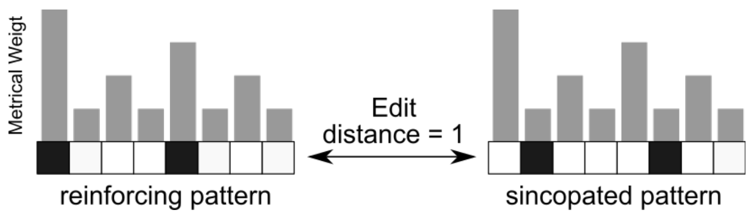

Figure 2.5. Two monophonic patterns at an edit distance 1. The pattern on the left is a pulse-reinforcing pattern while the pattern on the right is a completely syncopated pattern.

The idea that hierarchies and meter are a source of information for our cognitive system, when evaluating the similarity of two percussive patterns, is presented by Johnson-Laird (1991), as he proposes that rhythms can be varied to some extent without perturbing their resemblance if they belong to a certain rhythmic prototype (if they both have similar rhythmic characteristics). He argues that rhythm prototypes are built from families which are assumed to be only dependent on the onsets of a rhythm, forming categories (or families) specific to the phrase between two pulses. Families between pulses can be of three different types, being syncopations (S) the most important one,followed by beat reinforcements (R), and finally phrases which bring nothing (N) to the rhythmic sensation (a rest or a group of notes preceding an onset). A monophonic rhythm can then be reduced into a sequence of families. All possible monophonic rhythms belonging to the same family will, according to Johnson-Laird, maintain a similarity resemblance.

This compact form of devising rhythms and establishing their relationship, was later used by Cao et al. (2014) performing four experiments in order to evaluate what determines the similarity of rhythms, even when timbre, tempo, meter and number of notes within two patterns is held constant.

Cao et al. (2014) proposed a “family theory” whereby similarity between two patterns, which are reproduced with the same monotonic sound, is based on two main factors: one is the presence of an identical sequence of inter-onset intervals (IOI) in both patterns, while the other factor is that the relationships with the meter, the rhythm families, are held constant in both. Their aim is to contrast the cognition-based family theory of rhythm similarity, with the simpler approach of the edit distance rhythm similarity. Their results show how a pattern of onsets is more important than the same family for the closest similarity, and that family has a clear influence on similarity ratings, thus corroborating the general family theory (see Figure 2.6). Exact repetitions of regions (IOIs) and their shiftings showed to be highly influential on the ratings, revealing an effect of surface features over constructs derived from the metric. This might be related to the fact that when the patterns were presented to the subjects in both experiments, there was no explicit pulse (so neither meter) induction. Therefore the effect of the families, which is dependent on metrical weights, a construct of pulse and meter, is not sufficiently present when the similarity assessment takes place.

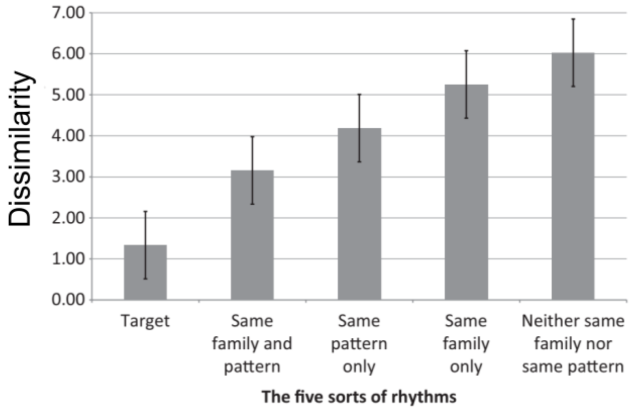

Figure 2.6. Different similarity levels for the controlled pairs of patterns used by Cao et al. (2014) (Figure extracted from that paper)..

As a conclusion, let's summarize that two theories of rhythm similarity in monophonic rhythm have been discussed, the one we refer to as *syntactic* and that considered to be *semantic*. The *syntactic* one groups different measures of similarity based on the transformations and displacements of the musical surface, disregarded of the effect that a rhythm elicits in a subject (i.e. the edit distance). These measures are agnostic so the high perceptual cost of changing a pattern of notes from reinforcement to syncopations can be regarded as a low-cost displacement of a set of notes. A strong rhythmic change could be regarded by these metrics as a small note displacement, thus assigning a high similarity between one rhythm that is reinforcing the pulse and another that is totally syncopated. The other theory, the *semantic* one, proposes how a similarity sensation is based on the notes' position in a pattern in relation to the induced pulse and meter, that is, their syncopation. In this way, different IOI configurations can elicit the same syncopation or reinforcement sensations and thus a sense of similarity. Both of these theories are grounded on well known perceptual mechanisms: on one hand humans excel in counting and finding general patterns from a surface of symbols, for example letters in a word (Meyer & Schvaneveldt, 1971), and on the other, a rhythm elicits a series of cognitive organizations, that define the way in which we process and understand a rhythm. However, empirical discerning among both is still incomplete.

2.3.3 Polyphonic rhythmic similarity

One purpose of this thesis, as it has been presented at the beginning of this chapter, is to understand the relationship between drum pattern variations (typical of EDM compositions) and objective similarity measures. The big picture here is grasping EDM drum production in terms of similarity so that different compositional processes can be explained and recreated using a cognitive perspective. However, as it is presented in the previous section, there is still room for understanding the mechanisms involved in the similarity sensations of two simple monophonic and monotonic rhythmic patterns and, even more, of real polyphonic drum patterns.

One of the main studies carried out with the purpose of understanding the factors involved in polyphonic drum pattern similarity is accomplished by Alf Gabrielsson (1973a; 1973b). He explored how different rhythms of different musical styles, reproduced with the same sounds of a drum machine, were perceived by subjects in terms of similarity. His methodology was to use the multidimensional scaling (MDS) technique to visualize rhythms as points in a conceptual space, based on human ratings of how alike several polyphonic rhythms are considered. Gabrielsson performed different experiments each based on evaluating similarity within different groups of patterns in which he controls parameters as IOI, meter, tempo and instrumentation.

His results offer relevant insights into the range of factors that influence polyphonic rhythm similarity. Each experiment considers a carefully selected set of patterns focusing on a specific aspect. He systematically explores the influence of the instruments' IOIs in similarity, by using a set of patterns all of them sharing a binary meter and the same tempo. On another experiment he focuses on the influence of meter in similarity, by including binary and ternary patterns in the comparison set. A third experiment explores the influence of tempo on similarity and his last experiment uses a set of real music samples to have an insight on the effect of instrumentation on similarity ratings.

His results do not formalize a polyphonic similarity model, but they give thoughtful insight into different factors that are involved in listener's similarity sensation of polyphonic drum patterns. The two main results extracted from Gabrielsson’s research are the notion of rhythm spaces, which emerge as useful structures for visualizing similarity relations between drum patterns, and a clear list of factors that have influenced polyphonic rhythm similarity in his experiments (1973b). These factors are according to himself:

- The meter induced by the sequence.
- The onset density of the patterns.
- The simplicity-complexity of the patterns.
- The syncopations.
- The number of different instruments in a sequence.
- The “movement character” of the rhythms.

Research on monophonic and polyphonic similarity studies reviewed in this section, evidence different advances in the comprehension of this phenomenon. There is experimental research carried out which underlines important factors influencing similarity judgements of rhythms, but in general it still remains an open question for monophonic and, specially, polyphonic scenarios. Monophonic rhythm similarity is more developed, and the models based on syncopation families and identical regions suggest a continuum from the more robust theories of human rhythm processing. This connection of similarity with cognitive and neural processing marks a starting point which could be transited in order to advance towards meaningful metrics.

The connection between the monophonic and polyphonic domains of rhythm similarity is not that strong. It would be ideal that rhythm processing knowledge could provide sufficient knowledge to trace a continuum between what it is known about monophonic similarity sensations and what is known about the polyphonic situation. The theoretical expansion of monophonic syncopation to polyphonic syncopation successfully used in groove scenarios (Witek et al., 2014), is one of the few attempts to close that gap. The acknowledgement of the importance of the predominant frequencies of the different percussive instruments found in a polyphonic drum arrangement, is an advance towards the comprehension of polyphonic similarity. Gabrielsson's studies remain relevant and somehow unique in the polyphonic drum similarity domain, but his results are so open ended that it becomes hard to formalize from them. Connections between Gabrielsson's spaces and similarity factors are still to be developed in order to define polyphonic similarity metrics.

2.3.4 Rhythm spaces

Alf Gabrielsson establishes the formal study of polyphonic rhythm through subject-based dissimilarity studies and the subsequent bi-dimensional visualizations of collections of rhythm patterns (see Figure 2.7). Other rhythm spaces emerged eventually as visualization techniques were used to explore different relations between percussive rhythms.

Desain and Honing have an extensive body of work on modeling human perception of rhythm from a cognitive perspective. In several papers they use a three dimensional space for visualizing rhythms. Each axis of the space represents one of the three inter-onset intervals (IOI) which exist between the four notes of their rhythm. In this informative space, a rhythmic structure is recognized by its position (Desain & Honing, 2003) (Honing, 2002).

Other authors have dealt with rhythm spaces and rhythm similarity in a polyphonic music audio (retrieval) context. Rhythm spaces are implicit in many MIR studies involving rhythm descriptors (Ellis & Arroyo, 2004; Rocamora et al., 2014, Paulus and Klapuri, 2002). Here, spaces are rarely explicitly depicted or used as such, probably because their multi-dimensionality and because the aims lean more towards automatic music classification (Chen & Chen, 1998).

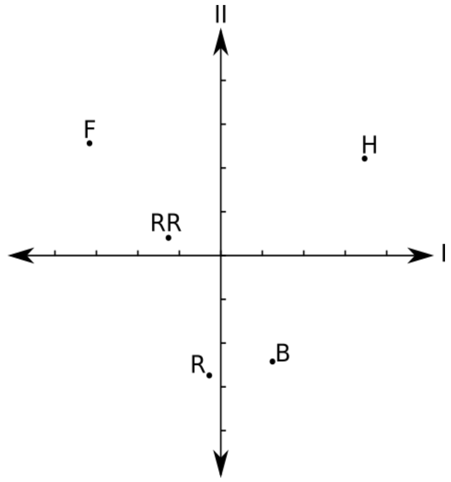

Figure 2.7. A rhythm space by Alf Gabrielsson (1973b). F: foxtrot, RR: rocnk'roll, R: rhumba, B: beguine, H: habanera.

2.4 Style in EDM

2.4.1 Definitions of style

Being art regarded as one of the most sophisticated and complex human activities, the possibility of constructing systems able to generate art pieces is considered as a milestone for Artificial Intelligence and engineering (Boden, 1998). One major concern, in artificial generative art, has been capturing the character of the melodic phrasing of a specific music composer or the essence of a pictoric movement (say Cubism or Impressionism) (Argamon et al., 2010). The goal is to have systems capable of extracting practical knowledge from examples of works of art and then to use that knowledge in the creation of new original art pieces following that same *style*. However, in order to create systems that can detect the essences of a style, and use that for generative purposes, it must be made clear what *style* is, and how it is manifested throughout a collection of art pieces. The concept of style is related to our mental ability to deduce that some items, for example sounds, paintings, gestures, or situations, have something that bounds them together and that makes them different from others and alike among them. Something that makes them belong to the style (Deliège, 2001).

A solid ground on musical style is given by Meyer (1967), one of the main authors addressing this area from a quantitative point of view. His notion of style is partially influenced by that of Information Theory (Shannon, 1948), and his vision has proven fruitful in posterior studies of musical style (Moore, 2001), and on the creation of generative music systems based on style (Temperley, 2007; Conklin & Witten, 1995; Pearce et al., 2005). He writes the following to introduce style:

> *"Style constitutes the universe of discourse within which musical meanings arise. There are many musical styles. They vary from culture to culture, from epoch to epoch within the same culture, and even within the same epoch and culture. This plurality of musical styles results because styles exist not as unchanging physical processes in the world of nature, but as psychological processes ingrained as habits in the perceptions, dispositions, and responses of those who have learned through practice and experience to understand a particular style. What remains constant from style to style are not scales, modes, harmonies, or manners of performance, but the psychology of human mental processes-the ways in which the mind, operating within the context of culturally established norms, selects and organizes the stimuli that are presented to it. For instance, the human mind, striving for stability and completeness, "expects" structural gaps to be filled in. But what constitutes a structural gap will vary from style to style."* (Meyer, 1967, page 19)

According to Meyer, musical style has a probabilistic nature, related to statistical processes as Markov processes (more details in the next section). He suggests that style does not exist in the work itself (the musical pieces or the paintings), but resides in the minds of spectators and artists who have abstracted, first unconsciously and sometimes also consciously (by musical training), the defining or characteristic features of a style. The norms of a style in a subject's mind can be modeled as probabilities which influence the mental behavior involved in the perception and comprehension of the style. One manifestation of these norms in a musical style, is how one sound or a group of sounds, activate expectations on trained listeners that indicate that another sound or group of sounds will be coming at some point in the music continuum. The product of these probability relations, are expectations which are the real materialization of style in a human mind (Meyer, 1967).

In this framework, cultural dynamics play an important role in the establishment and dissemination of a style, as they promote the exposition to certain styles of music, and thus the development of the rules of a style, in the minds of the listeners.

> "Once a musical style has become part of the habit responses of composers, performers, and practiced listeners it may be regarded as a complex system of probabilities. That musical styles are internalized probability systems is demonstrated by the rules of musical grammar and syntax found in textbooks on harmony, counterpoint, and theory in general."(Meyer, 1967)

Meyer's work founded a line of study from which many different types of musicological, psychological (Krumhansl, 2001) and generative music research (Temperley, 2007) took off. From the second half of the 20th century on, these theories of style, aided by statistics and the advent of computers, fueled artists, engineers and researchers from different disciplines to investigate the possible uses of machines to replicate musical style (Argamon et al., 2010). Most of that research has focused on classical music with some papers on popular music and very few dedicated to generative applications of EDM.

It is important to point that musical style research is based in the study and processing of symbolic representations of music. Typically, these studies are approached through musical scores and, for the last decades, using transcriptions in a common digital music format as MIDI (Musical Instrument Digital Interface)[9]. The use of digital music formats is very convenient as it allows to compile and process music as digital information using computers. This allows for fast and reliable computation of multiple statistics and analyses of music and as such it is common practice in contemporary music research. Style studies in music with generative purposes have been approached by different authors (Pachet, 2002, 2003, 2006; Cont et al., 2006; Jacques et al., 2016).

2.4.2 How to study EDM styles

EDM seems to be deeply anchored in conventions that define the different genres and subgenres, emerging from a complex network of factors that include aspects as diverse as marketing trends and social stratification, technological development and studio production techniques, not to mention the different musical roots from diverse ethnicities and geographical idiosyncrasies. EDM styles proliferate as they are developed by the introduction of new musical rules and constraints, mostly in rhythm and timbre. The genesis of new EDM styles is clearly exemplified in the birth of Acid House as a consequence of the special use of the TB-303

[9]MIDI musical format has become a standard since the 1980s when most prominent musical instrument manufacturers joined forces to design a musical protocol that was common to all of them. MIDI allows music scores to be represented in formats that can be processed by machines.

synthesizer in Chicago House music drum tracks[10], or in the establishment of dub techno style by introducing reverberation and delay techniques from Jamaican Dub Music into techno production (Blánquez, 2002). In a broad sense, a study of EDM musical styles should be primarily focused on rhythms and timbre transformations, and secondarily in structural, melodic and harmonic analyses. This perhaps explains why, given the musicological methodologies available, EDM has not been the focus of broad stylistic studies. Another plausible reason is the common idea that some types of EDM music are so simple harmonically, melodically and structurally that musicologists might think “there is not much to study”. In general terms, EDM is founded in the musical dimensions less studied in classical music analysis, namely timbre and rhythm.

However, some authors (Faraldo et al., 2016; Butler, 2006; Collins et al., 2013; Anderson et al., 2013) have researched different aspects of EDM style evidencing, according to Meyer, some of its rules, traits and conditioned expectations. Faraldo (2016) has studied if the idea of key and mode, from a classical music perspective, is still relevant in EDM. Anderson et al. (2013) have proposed a methodology to research EDM styles, using machine learning and directly analyzing the audio signals of songs, focusing on four different styles, namely Breakbeat, Dubstep, House and Drum and Bass. Butler (2006) and Collins et al. (2013) studies of EDM have focused on rhythm and drums and their relation to the formation of a style, providing rich and useful methodologies for further studies. Collins describes EDM styles as ‘’blurred and not easily defined’’ but proposes a style map including commercial Hiphop/Rap (from 1979), Electro (1982-3), Chicago House (1984-9), Detroit techno (1985-9), Acid House (1987-9), Club Techno (from 1989), Jungle (1992-4), Drum and Bass (from 1994) and Garage (both US and UK). Despite these efforts,

[10]A vivid example is acid house pioneer Chip-E talking about how they managed to extract the wobbling sounds of the TB-303 without understanding quite well what was happening, but guided mostly by tweaking the knobs with a “musical sense”. In this case, this experimentation guided by their musical sense materialized in the emergence of a powerful EDM style known as Acid House.

there is still room for stylistic research in EDM, as the repertoire increases constantly and new sub genres are created with the advancement of technology and aesthetic innovations.

Perhaps the main factor influencing the obliteration of EDM in academic research, is the fact that its dissemination and conservation format is audio and not music scores. Audio to symbolic automatic transcription technology (i.e. audio to MIDI) would be indispensable to study EDM fully, as researchers could recover a score from an audio recording. However, as these tools are not technically available yet, EDM style studies have to undergo the complex activity of manual transcription.

Based on the literature reviewed in this section, the idea of drumming style to be developed in further chapters will be defined by the probabilistic analysis of percussive information in symbolic format. As proposed by Meyer (1967), and later embraced and developed by other authors (as Conklin & Witten, 1995 or Temperley; 2007), musical styles are rooted in human minds as habits, dispositions and responses which can be modeled using probabilities. The probabilistic analysis of EDM drum rhythms will be based on the study of drum sequences, specifically in MIDI format.

2.5 EDM Production and Drum Sequencing

Composing music, playing musical instruments and producing audio recordings, three different activities which traditionally have been done by different specialized individuals, are increasingly being fused by the immense potential brought to music production by computers and DAW applications, which allows performing these three activities with a single tool. The division of roles between the music composer, the performer and the audio engineer is being blurred by the current state of technology so that it is very typical to find nowadays a composer who is also an expert in music production, or a performing musician who has taken over composition and music production in a personal studio. To some authors, this can be a sign of an advancement towards a democratization of music production (Goodwin, 2004), and even as a transformation of power relations in studio work (Théberge, 1997). Beyond these interpretations, it is evident how the possibility to do very specialized music work using personal computers, and even porTable devices, has allowed a person previously dedicated to a specific musical role to go beyond its traditional boundaries, learning and taking over new complementary activities.

Most textbooks and tutorials dedicated to EDM production is developed under this premise, as they focus in the introduction of novice or amateur musicians into the multiple and diverse aspects of music composition and production, presenting audio technical work, music composition and interpretation as a continuous activity. Different authors have written about EDM composition (Adamo, 2010, Brown & Griese, 2000; Emmerson, 2013; Hewitt, 2009; Snoman, 2012), all of them covering aspects from setting up an EDM studio, introductory musical theory focused on chords and melodies, notions of timbre manipulation with sound synthesizers, and drum programming according to different EDM styles. It is very significant that drum sequencing sections are approached by explaining how different EDM drumming styles are sequenced. In this sense, rhythm construction is always approached by describing prototypic patterns of an EDM style, and not by explaining the constitutive aspects of rhythm and dance. Even more, there is no description of any aspect of rhythm processing as pulse, or syncopation, to explain the patterns under their framework. All drum sequencing knowledge that can be extracted from these books is valid within the examples extracted, and leaves the reader without theoretical tools to interpret why the rhythms in the examples work for dancing, or how can they be transformed to enhance the drum production process. An original approach on EDM production literature can be found in DeSantis (2015) ,as he presents different dynamics for un-blocking the creative process of producing electronic dance music. As the previously mentioned authors, he also uses piano roll screenshots for representing dum sequences and covers a similar range of topics from technical to aesthetic or music theoretical. The treatment of drum sequencing is again presented without any context of rhythm processing or any rhythm perception theory. The author is, nevertheless, sharp in pointing out the extreme problems of sequencing drum rhythms in creative ways, specifically when not copying a previously existing rhythm. However, he fails to acknowledge any valid framework from which to start overcoming rhythm sequencing creativity. It is important to remark his very unique and useful strategies to make variations from rhythms, which are not found in the other books, again clearly pointing out to another fundamental activity of EDM drum production: the act of making transformations of drum patterns.

Very similar information to the aforementioned literature, but less condensed and perhaps more specialized, is found online in blogs, sites and video channels dedicated to EDM production. These online information presents text, images and videos, describing how to produce electronic music, get started with DAWs, understand and use technologies for audio processing and sound synthesis, make basslines and drum patterns for specific styles and so forth. Specific resources for drum production include transcriptions of famous *drum breaks* extracted from funk and soul records presented in the piano-roll format[11], periodic publications on how to create EDM drum patterns in highly specialized styles[12], archives with examples of drum patterns in prototypic EDM styles[13,14,15,16], a Master's thesis with drum transcriptions of Drum Breaks, prototypic EDM and Afro-Cuban drum patterns and (Hein, 2013), sequencing and sound design of EDM drum patterns[17], or a complete book with a guide on how to compose EDM with three chapters devoted to drums and percussion[18]. The different EDM styles covered online and in

[11]http://funklet.com/
[12]https://www.attackmagazine.com/technique/beat-dissected/
[13]http://subaqueousmusic.com/drum-patterns-for-electronic-music/
[14]http://simonv.com/tutorials/drum_patterns.php
[15]https://mccormick.cx/news/entries/how-to-write-beats.news
[16]http://quadrophone.com/drums/midi-drum-patterns-for-edm/
[17]http://howtomakeelectronicmusic.com/category/tutorials
[18]http://users.skynet.be/shedo/DMR1/Index.htm

printed literature tend to be House, Techno, Breakbeat, Garage, Drum and Bass, Hip Hop, Trance, Chillout, Dubstep, Jungle and Trip hop.

In all EDM production specialized literature, the topic of drum sequencing is always present, as drums are the musical elements that set the rhythmic environment that gets people to dance. Despite the fact that the main dance-functional aspect of drums in EDM is clearly stated by all authors, none of them have made any connections to specialized literature in order to expand on this topic as it is presented in section 2.3.3, (i.e. to expand on how dance is stimulated from drum sounds themselves). In general, this literature lacks of comprehensive explanations on:

- How to create original drum patterns.
- What makes a drum pattern incite people to dance,
- How to transform a pattern maintaining its essential identity.
- How to concatenate drum patterns to keep a continuous flow.

There is an evident disconnection between what is available in the literature to guide EDM producers to carry out the production of dance tracks, and the actual work that has to be done. A breach which is amended by experimentation and trial-and-error by the producers themselves who, at the expense of their time and effort, explore how to deliver proper EDM drum tracks. A similar separation between EDM records and the specialized EDM press is observed by Eshun (1999) as he argues, it is incapable of describing rhythm, keeping it “as an unwriTable, ineffable mystery”, an attitude guided by a “hostility towards analyzing rhythm at all”. This gap, between shallow explanations of EDM rhythm and the actual theory to work in EDM production, is an opportunity for EDM schools, writers, musical media producers and musical software designers. There is a need to address the distance between the rhythms that actually make people dance and the knowledge to understand them, to talk about them and to use their essence as inspiration for programming new rhythms in drum machines.

Some companies have created products for bridging this gap, distributing specialized EDM media building blocks that can be useful throughout the production pipeline. These building blocks tend to be packed with hundreds of MIDI and audio files of drums, basslines, chord progressions, melodies and vocal samples belonging to a specific EDM style (e.g. a Funky House composers' bundle). These elements serve the purpose of saving production time and structuring the foundation of a track with a specific style sparing a composer for time, knowledge or inspiration for a project. These building blocks can be seen as compositional presets, or interchangeable musical units, which usually comprise hundreds of bundled files that can be combined to achieve a sort of stylistic collage. An example of this sort of product, the "Essential Minimal Techno Vol 2"[19] contains 894 Mb of 539 Samples broken down as:

- 366 Rex2[20] Files
- 10 Construction Kits
- 113 Drum Loops
- 16 Bass Loops
- 53 Synth Loops
- 49 Percussion Loops
- 17 Vox Loops
- 46 Fx Top Loops
- 164 Hits and Fx
- A 30 Midi Files

These large quantities of files multiplied by many different styles, end up

[19]https://www.loopmasters.com/genres/40-Techno/products/543-Essential-Minimal-Techno-Vol2

[20]A Rex file is a loop file format supported by different software and hardware samplers and DAWs.

stacked on a producer's hard drive, easily occupying GBs of computer disk. As the collection grows, it tends to be progressively obscured by abundance, as it becomes more and more difficult to navigate a musical collection by name, especially when trying to find patterns that combine, resemble or contrast well with others. Picture a producer searching for an adequate drum loop to layer on top of a percussion loop from the "Essential Minimal Techno Vol 2" collection, and at the same time looking for them to match rhythmically. This would mean having to go through the 49 loops one by one, alphabetically, loading them to the DAW and listening to them evaluating the effect. This is a memory and time-consuming intensive task. Even more if it expands to searching through several style collections. As this media combination technology offers a solution to rhythmic and musical knowledge, it also imposes limits on the production activity, as it implies making musical sense of large amounts of material.

2.6 Generative Music Sequencing

Based on very different ideas from the pre-recorded musical building blocks presented on the previous section, but perhaps with some common intentions in their final goals, scientists and musicians have explored the automatic generation of musical material. The common purpose of these generative music systems is to automate, at least, some portion of the music composition process, which can span from the generation of scores and instrumentation instructions, up to the production of complete musical pieces rendered as audio files. For the design of a generative music system, the composition process is modeled by the creator using any means possible, which can go from pencil and paper (Hedges, 1978) to the use of digital processes. Despite the wide range of possibilities, generative music systems have been mostly developed using computers (Collins, 2008b). For the rest of this thesis, the creation of generative music systems will refer then to the design of computer programs which use algorithms to produce musical material.

Generative music systems can be classified in two groups, according to the music representation they are based upon: symbolic or audio-based. Audio-based representations are a transduction of the raw acoustic signal unrelated to any musical notion, while, on a higher level, symbolic or notation-based are abstract musical representations, as a musical score, indicating the qualities and moments where the specific sounds should be located. Given the current underdevelopment of automatic audio to symbolic transcriptors, the detail and precision of audio based representations cannot be approximated to the results given by symbolic representations. In the end, both operate with symbols, but its creative manipulation depends on the detail of their representations. The actual technological context therefore suggests that using symbolic musical data as the base for the development of a generative music system, is the most straightforward approach.

One of the earliest approaches for devising music systems is noted by Meyer (1967) who, influenced by Shannon (1948) and Weaver's (1953) communication theories, suggest statistical analysis and a Markov process as a model to deal with music as information, in order to extract the traits and rules of a musical style (see section 2.4).

> *"If music is a Markoff process, it would appear that as a musical event (be it a phrase, a theme, or a whole work) unfolds and the probability of a particular conclusion increases, uncertainty, information, and meaning will necessarily decrease. And in a closed physical system where the Markoff process operates this is just what does occur- probability tends to increase."* Meyer, (1967)

A Markov process is a stochastic method where the immediate probability of a variable is determined entirely by the occurrence of most recent variables (Gardiner, 2009). This means that the future state of a variable x_2 can be predicted by a conditional probability $P(x_2, t_2 | x_1 t_1)$ given that $t_2 \geq t_1$, where t is the state of the variable. The next value of x depends on its previous state, and it is expressed in terms of a probability: how probable it is for x to have a value of x_2 given that the past state was x_1. This dependence opens the door for multiple future states (x_2t_2, x_3t_2, x_4t_2 ...) to have different probabilities of occurring given a same previous step. In a musical context this could translate to deciding which note to play in a current state given the note played in the previous state. The probabilities for the Markov process can be derived from the statistical analysis of a corpus of sequences in order to assign probability values that relate a recent condition with different possible variables.

In a musical context, as proposed by Meyer (1956, 1967), these probabilities that relate a context (a recent condition) with the occurrence of a musical event (a variable) are modeled by the probabilities inferred from a particular musical style. Thus, following Meyer, the probabilities extracted from a corpus are a manifestation of the compliance a given musical event with a style. As such, the probabilities extracted from the analysis of a musical corpus, do not inform the underlying musical composition rules of the pieces, but rather a mental agreement (or disagreement) of the occurrence of a musical event within an idea of style. On the next decades new approaches to symbolic style-modeling have emerged. Cope (2004) presented music generative work based in the implementation of music-theoretical rules. Steedman (1984) used formal grammars specifically for jazz harmony. Statistical modeling of style has also been a fruitful approach toward generative music systems (Conklin & Witten, 1995 ; Conklin, 2003; Pearce et al., 2005), especially suited for the automatic generation of music in the styles we are dealing with, about which there is no proper formalized musical theory yet. Some other approaches are focused on the automatic generation of complete EDM pieces (Collins, 2008b; Anderson et al., 2013)

Markov sequence generation has been used in many generative musical applications (Ames, 1989; Brooks et al., 1957; Temperley, 2007; Nierhaus, 2009). More specifically, Markov chains have also been applied in real-time interactive music systems, such as ”M” (Zicarelly, 1989), the

”Continuator” (Pachet, 2002) or ”Omax” (Cont et al., 2010). However, Markov chains and interactive control are two concepts that do not go well together, because a user may not be able to specify additional musical properties wished in the generated material, while preserving Markovian properties and therefore stylistic consistence (Pachet, 2006). Pachet proposes the use of Elementary Markov Constraints (EMC) as a computational solution for obtaining steerable or interactive Markovian sequences. Another downside of a simple Markov processes is that of structure, as the process is useful to grasp the prediction of future states at the note level but it does not necessarily upscale to the motivic or section or structural level of a musical composition (Pachet et al., 2011). Other modeling tools that have been used are genetic algorithms (Johanson and Poli, 1998), neural networks and, more recently, deep learning (Huang & Wu, 2016). General purpose generative systems, devised to output complete music pieces (controlling the rhythmic and also melodic, timbral and structural aspects of a piece) are out of the scope of this thesis, as the focus is the generation of drum pattern sequences. The next section presents systems that deal exclusively with drum pattern generation.

2.6.1 Generative drum sequencing

Independently of the input format, the techniques used for the analysis and synthesis of drum rhythms are diverse, being genetic algorithms (GA), neural networks and stochastic processes the most commonplace.

Using symbolic representations, Burton (1998) system uses a GA to recombine collections of polyphonic drum patterns extracted from drum machines and transcribed manually. As a part of the GESMI project, aimed at generating complete electronic music tracks, Eigenfeldt & Pasquier (2013) uses 1st order Markov chains of 32 steps resulting from the analysis of the drum tracks of 100 transcribed electronic music songs. Tidemann et al. (2009) present a system based on Echo State Networks (ESN), a particular approach of a neural network that is trained in real time by a human MIDI drummer. Once their system is trained, it is set to imitate the sequence that had been used in training. Bernardes et al. (2010) use a GA to create new polyphonic drum patterns, based on the study of a set of MIDI drum loops. The main operations of a GA which provide a variable population of rhythms are crossover and mutation. Crossover is based on a first order Markov chain and mutation on the selection of a step to transform controlled by their metrical weights. Once a population is created, density and complexity are used as user inputs to filter out the output drum patterns (Bernardes et al., 2010).

In audio-based drumming systems, Aucouturier and Pachet (2005) describe a reactive system that adapts to the musical input of a performer on a MIDI keyboard. The generative system is based on the extraction of drum sounds from recordings, and then uses concatenative synthesis to generate rhythms. In the reported example, the mappings between MIDI and drum generation, as well as other generative controls and constraints are defined offline, therefore letting the system to drum along with no real-time control. Collins (2001, 2002) presents a collection of algorithms and techniques for cutting drum loops and reshaping them for their use in EDM composition. Wooler & Brown (2011) describe a fast adaptive system used to create rhythm mosaics resulting from two audio sequences to be cross-faded at the user's will. Cross-fades are not applied to volume but rather to the percussive elements extracted from one track or the other and located in non-disruptive positions. The rules for locating the fragments are based on the Markov analysis of the short rhythms to be cross-faded. These two last examples are interesting due to the creative approach to polyphonic rhythm generation. Most commercial drum programs and plugins available are concerned with sound rendering (synthesis and sampling) and basic sequencing, rather than with intelligent pattern generation or algorithmic composition. We present below, a summary of the most relevant programs we have found connected to our research.

Different Drummer (Technemedia, 2015) and Robotic Drums (Urtubia, 2015) use stochastic methods for generating drums. Both are drum sequencers in which events on a given step are user-controlled by a probability value. Another approach is Stylus RMX (Spectrasonics, 2005), which aims to transform drum patterns rhythmically based on displacing onsets to certain points in the grid. There are two variation parameters: a ”simplify” knob, which reduces the amount of onsets in the loop, and a discrete selection menu called ”variation”, where a fixed amount of variations from the original patterns can be selected. Although not a drum-exclusive application, drum loops can be loaded in order to be transformed. FXpansion’s BFD3[21] created one of the first virtual drummers to be used inside DAWs to emulate acoustic sounding drums as a replacement of a drummer in a recording studio (Figure 2.8).

Electronic artist Cristian Vogel has applied the Euclidean algorithm (Toussaint, 2005) to automatic pattern generation in the Kyma environment (Vogel, 2015), and has used the software to create all the rhythmic elements for his 2014 album ”Polyphonic Beings”. WaveDNA has recently released Liquid Music[22] for Max for Live, which provides building blocks of rhythmic patterns that can be varied and tweaked with unique visual editing tools, such as the ”beatform tumbler” complexity transformer, the ”beatform weaver” combination generator, or the ”groovemover” remixer. Artist James Holden has tackled the difficult notion of groove and the challenges that need to be addressed when interacting with human musicians. Based on Holger Hennig ideas, who examined the effects of synchronization between musicians (Henning, 2014), he has released a free MIDI humanizer (Holden, 2015) which can listen to and respond to musicians in real-time performances.

The generation of drum sequences is perhaps an under-researched topic in the vast panorama of generative music, which has primarily focused in

[21] A review of FXpansion https://www.soundonsound.com/reviews/fxpansion-bfd
[22] https://www.wavedna.com/liquid-rhythm/

melodic and harmonic aspects of music. Even though there are several generative music approaches, the concepts of drumming style in the context of EDM have not been clearly approached. The techniques reviewed for musical sequence generation offer a wide palette of options, but the Markov process approach makes a perfect sense to the problem of rhythm sequencing, given the short length of the sequences (i.e. one or two bars), and their apparent detachment from the musical structure. The straightforward way in which Meyer associates the mental operations of musical style processing with the Markov process, suggests it can be an adequate method for approaching rhythm generation.

Figure 2.8. Screenshot of the BFD3 drum machine.

2.8 Conclusions

As it was demonstrated by Witek et al. (2014a), syncopated arrangements of polyphonic drums, with a crafted balance between reinforcement and denial of the pulse in the low against mid-and-high frequency instruments (Witek et al., 2014b; Hove et al., 2014), are pillars in which the desire to dance is grounded. Years earlier, and without the need of scientific confirmation, early EDM pioneers developed strong intuitions on these same matters, which they mixed with appropriate technological tools that paved the way for new musical genres to appear globally (such as House Music, Techno or Electro). From that moment on, a completely new culture emerged using low cost sound machinery and obsessed with how low percussive frequencies encouraged body movement and with a special openness to practitioners of any musical expertise. This culture manifested itself musically with apparently simple pieces, from a melodic and harmonic point of view, which offered vast richness and depth in the percussive arrangements and timbres used for its composition. These pieces, dance tracks, were designed with a special capacity for carefully mutating rhythmically, so that endless physically engaging musical structures were its default outcome.

In order to understand how an EDM dance track works, allowing to keep the desire to dance active and staying away from boredom, despite its emphasis on a specific pulse and meter, careful observation of the variations occurring at its core has to be done. This is even more necessary, if the objective, as it is in this work, is to design new music machines capable of reproducing such drum variations. It is clear that an evaluation tool is needed, some sort of measuring device that could give answers on which changes occur and of which magnitude. The assumption we make is that if we have a proper tool for understanding the inner mutation of EDM dance tracks, we might be able to reverse-engineer the tracks and be able to create new variations based on those observations. The topic then becomes a combination of similarity and polyphonic drum pattern processing principles. With such special combination of knowledge, we aim at devising the rhythmic measuring tape that will allow us to construct similarity analysis of EDM drum patterns, and building generative systems around it.

As far as we could find, polyphonic drum similarity was first addressed by Gabrielsson (1973b), who combined a multidimensional analysis of similarity judgements with well studied perceptual and cognitive phenomena of pulse acquisition, meter entrainment and pulse reinforcement and denial. As Gabrielsson's early explorations point, the sensation of similarity between drum patterns is influenced by different factors that obviously include the instrumental IOI organization, but also others that go beyond it, and which are constructed by the mediation of the IOI surface of the pattern and our rhythm processing system. In this sense, when the combination of the syntaxis (the order of onsets and silences in time) and semantics (the rhythmic interpretaion we extract from the musical surfaces) are said to be responsible for our similarity sensations of rhythm, the monophonic similarity explorations of Cao et al. (2014) and Johnson-Laird (1991) emerge as related theories that can point out the way to go when defining a polyphonic similarity metric. Gabrielsson´s polyphonic experiments explore many aspects that influence rhythm, but the patterns he uses as stimuli belong to a broad diversity of dance rhythm cultures, perhaps a wider territory than EDM, and his experiments include tempo and meter variations which are not at the core of EDM. Perhaps an EDM context with similar methodology as the the one used by Gabrielsson might pave the way to a higher understanding of polyphonic rhythmic processing, specially focusing on similarity when the tempo is not a variable factor, and pulse and meter induction are always present (as it happens on an EDM dance track). These same ideas can be projected onto the similarity studies carried out by Cao et al. (2014) with their family theory of rhythm. In their similarity experiments pulse induction for the subjects is not controlled, so it is unclear how the factors related to rhythm families (dependent on pulse, meter and syncopation) are really influencing similarity sensations. Perhaps, rhythm families (rhythm *semantics*) are more relevant for judging similarity in the presence of a pulse and thus more relevant for our context. As such, these two scenarios related to monophonic and polyphonic sensations of similarity are our departure points for contextualized experimentation in EDM, specially focused in devising a metric that will help us analyze musical information.

There is a general coincidence from different scientific points of view on how rhythm is processed by humans but only few research efforts have managed to extend knowledge of rhythm processing to that of similarity. Namely Johnson-Laird (1991), Cao et al. (2014), Witek et al. (2014a) and Gabrielsson (1973b) have planted the first seeds for a structured comprehension of similarity sensations rooted in psychoacoustic rhythmic knowledge and sharing the concepts of similarity as devised by cognitive science. These first attempts show some routes which are open to debate and experimentation.

A diversity of algorithmic techniques have been used for the generation of musical sequences being the most recurrent Markov process, as genetic algorithms and neural networks. All these techniques imply different technological approaches to the analysis and generation of music. Although the rhythmic aspect of music and percussion in general are the least explored dimensions and materials in generative music, there are some examples to be considered for the development of an EDM generative system. From the range of techniques in generative music literature, the use of Markov models in different implementations as those proposed by Pachet, or the multiple viewpoint systems, suggest a simple yet solid starting point for the generation of EDM drum sequences. Memory limitations, inherent in Markov process sequence modeling, are perhaps not so relevant when targeting the generation of relatively short drum patterns as opposed to the creation of complete musical compositions as *dance tracks* or *drum tracks*.

The next two chapters present experiments and activities resulting from the present state of the art. Chapter 3 presents monophonic and polyphonic similarity experiments, which will be carried out in order to gain an insight into rhythmic cognition. These experiments are derived from the results discussed in sections 2.2 and 2.3. These results will be used as guidelines towards the creation of predictive similarity models for drum patterns. Chapter 4 is focused on the generative aspect of drum rhythms, based on the notions of style, generative music sequencing and conceptual spaces, reviewed in sections 2.3, 2.4 and 2.6. In Chapter 4, two original technologies will be introduced. One dealing with automatic generation of drum patterns in style and the other exploring the use of low-dimensional spaces to control rhythm generation.

3. TOWARDS METRICS OF RHYTHM SIMILARITY

3.1 Introduction

This chapter presents a series of experiments carried out to expand the current comprehension of the mechanisms involved in assessing the similarity between two drum rhythms. There are two transversal questions to this whole process namely: what elements does a listener grasp from a pair of rhythmic patterns to define how similar they feel? Can those elements be formalized in an algorithm in order to create a similarity metric that is simple to compute? It is assumed that, as it happens in other domains as color (Shepard, 1964) or tonality (Krumhansl, 2001), there must be objective measures that can be extracted from rhythms themselves, which can be used as inputs to algorithms that can predict subjective similarity ratings, and with them be able to construct conceptual spaces (see section 2.3.1). The final objective is thus to be able to determine how a group of drum rhythms, as the ones typically used in EDM production, can be organized automatically by similarity.

As it has been presented in section 2.3.2 there are some advances in models and algorithms for similarity prediction between monophonic rhythms. The method proposed by Cao et al. (2014) is based on identical regions (IR) and syncopation families (SF). An IR is defined as a sequence of onsets and silences that is repeated in both patterns being compared, possibly shifted in location (see Figure 3.1). SF, on the other side, are defined as different states of the inter-pulse fragments of a pattern: they can be syncopated, pulse reinforcements or nothing. After performing two similarity experiments, they conclude that the similarity sensation reported by subjects judging monophonic patterns is correlated with both IR and SF. From a cognitive perspective, the effect of IR and SF trigger two different processing mechanisms referred here as *syntactic* and *semantic*. IR are based on comparing what has been regarded as the musical surface, the acoustic sequence of onsets and silences. The syntactic analogy proposed to describe this cognitive mechanism is based exclusively in the order of the elements in the sequence of the rhythm. On the other hand, the SF theory relies on more elaborate cognitive mechanisms, as the superimposition of a pulse and a meter over the acoustic signal, and the arousal of syncopation or pulse reinforcement sensations (see section 2.2 for an introduction). This mechanism is referred to as *semantic*, based on the possible rhythmic significances of the drum sequence.

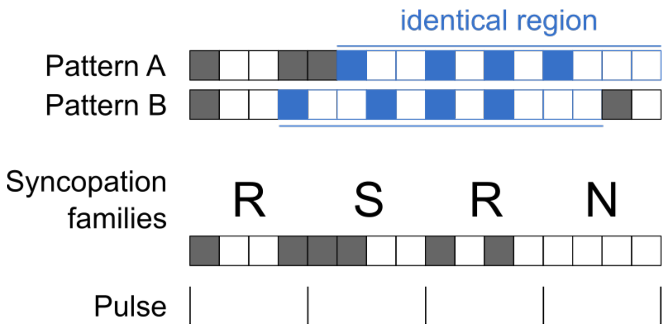

Figure 3.1. Graphic explanation of identical regions and syncopation families. Grey squares are onsets and white squares are silences. Identical region in red. Syncopation families are assigned to each intra-pulse sub-pattern, R: reinforcement, S: syncopation, N: nothing.

There are two questions regarding the results presented by Cao et al. that will be expanded with new experiments in this chapter:

- It is remarkable that the IRs present in both patterns being compared, do not need to occupy the same position in both patterns. Their experiments show that IRs can be displaced in time without affecting similarity. But there is no systematic exploration of the effect of the time shifts in similarity judgments.
- The influence of pulse induction is not taken into account in their experiments. The monophonic pattern stimuli are presented to the subjects without any previous rhythmic context. As this thesis is rooted in EDM, where the pulse is fundamental, it is crucial to understand how a concurrently induced pulse affects similarity ratings, and to interpret these ratings via IR and SF.

New experiments similar to the ones Cao et al. will be carried out entraining subjects to a pulse sensation before evaluating similarity, and designing the stimuli by controlling the shift of the IRs. By designing and carrying out a new experiment, the applicability of the IR and SF model for measuring rhythm similarity, will be better understood.

Expanding to a polyphonic perspective of rhythm, the experiments of Alf Gabrielsson (1973a, 1973b) mark a solid precedent in the process of understanding rhythm similarity in polyphonic drum rhythms (see section 2.3.3). His results present different low-dimensional rhythm spaces based on subjects' similarity ratings, using different sets of drum patterns extracted from a drum machine, as stimuli. Some of his reported experiments adjust the patterns with a constant meter and tempo before exposing them to subjects. This condition recreates perfectly the context of EDM studied in this thesis, where the pulse and the meter are generally constant factors within a *dance track* or even in a session of several mixed *dance tracks*. One outcome of Gabrielsson's polyphonic similarity paper (1973b) is a list of rhythmic factors, extracted as a summary from all his experiments, which impact subject's similarity sensations and thus the resulting rhythm spaces. These factors are:

- The meter induced by the sequence.
- The onset density of the patterns.
- The simplicity-complexity of the patterns.
- The syncopations.
- The number of different instruments in a sequence.
- The "movement character" of the rhythms.

Although these are qualitative factors not explicitly encoded as compuTable data, they present a reasonable ground from where to start building a model for polyphonic drum rhythm similarity.

There are two specific ideas and conditions in Gabrielsson's work which are important for the development of this thesis and will be expanded carefully in this chapter. First, the rhythms used in his Experiment 1, where the meter and pulse are kept constant, are presets from the drum machine he uses. Second, as mentioned above, the factors Gabrielsson suggests influence similarity are qualitative, but for them to work on a similarity algorithm, they must be formalized through different equations obtained as objective measures from the patterns. These factors are formalized in this thesis as new rhythmic descriptors, adapting contemporary knowledge on monophonic and polyphonic processing of rhythm, as proposed by Witek (2014b), Hove (2014) and Buger et al. (2017).

A new experiment replicating Gabrielsson's methodology was carried out using contemporary EDM rhythms and obtaining a new rhythm space. The correspondence between both experiments will be analyzed, examining if the factors that influenced Gabrielsson's results can also be observed in the new EDM space. This correspondence will be evaluated using the new rhythmic descriptors proposed.

The following sections of this chapter will present experiments carried out with the aim to formalize and evaluate similarity metrics for EDM drum rhythms. In section 3.2, one experiment using monophonic rhythms will be presented. In this experiment the focus will be evaluating the effect of inducing or not the beat before judging similarity. It will also shed light on the size of the shift in the identical region (IR) in two patterns being compared. The results of this experiment will be used in section 3.3 to implement two monophonic similarity metrics and to expand them into tentative polyphonic similarity metrics.

In section 3.4 novel symbolic rhythmic descriptors, built upon aforementioned theories of rhythm cognition and perception (reviewed in sections 2.2.2.2 and 2.3.3), are presented. These descriptors are used in section 3.5 as means to define objective methodologies to predict polyphonic similarity, using Alf Gabrielsson's rhythm spaces as a target. Given that the previous experiments lead to the construction of polyphonic similarity methodologies, a final experiment is presented in section 3.6, which is carried out to evaluate the different polyphonic similarity prediction methodologies. This final exercise will allow the selection of the most robust procedure to measure rhythm similarity regarded as the one that has the best fit with Gabrielsson's and the new EDM rhythm spaces.

Section 3.7 summarizes the results presented throughout this chapter. In this section the contributions of the different experiments to the state of the art of rhythm similarity will be considered. The end of this section presents a discussion on possible follow up activities that could further extend of the work presented here.

3.2 Experiment 1: Monophonic Similarity and Syncopation

This experiment is designed to expand on the experiment presented by Cao et al. (2014) on rhythmic similarity, who report that similarity can be explained by combining two different cognitive mechanisms: one based on the inter onset interval (IOI) of the patterns, which is referred here as a *syntactic* experience of the rhythm, and thus based solely on the sequence of characters. The other mechanism is based on the "significance" of the rhythm, which is based on the pulse, the meter and the syncopations or pulse reinforcements. This later mechanism is referred to as *semantic* because it relies on how a rhythm is interpreted by its relation to cognitive constructs. In their paper it is proposed that both mechanisms are active when assessing similarity: the *syntactic* is evidenced when similarity is affected by the two patterns having identical regions (IR), even if they are not placed on the same position. The *semantic* mechanism is based on a segmentation of the patterns at each pulse, and the labeling of each intra-pulse pattern into syncopation families (SF) as: syncopated (S), beat reinforcing (R) or nothing (N) (Figure 3.1).

Cao et al. report a first experiment where the participants' task is to listen to one pair of rhythms and then to another pair, and to judge which pair is of a greater similarity. The pairs of patterns to be judged are designed to have the same amount of onsets and to have IRs. Their results show that patterns are most similar if their IR contain the same pattern of onsets, quite similar if they are in the same SF, and least similar if they do not contain an IR of onsets and have different SF. These results suggest the importance of both mechanisms in similarity judgements.

In a second experiment they ask for the same pairwise comparison, using same-length patterns. However, patterns are all resulting from controlling the IOI pattern and the family. To construct the patterns they start from a ‘target’ pattern and make four variations on it, changing the rhythm family and inserting an identical region (IR) in both patterns which is then shifted in time. The target pattern is then compared to the four different variations: one pattern with same SF and IR, another with same SF but no IR, another with different SF same IR, and the last pattern with different SF and no IR. Their results show how both cognitive mechanisms the *syntactic* and the *semantic* are active when assessing similarity, as both factors influence similarity sensations reported by subjects (Figure 3.2).

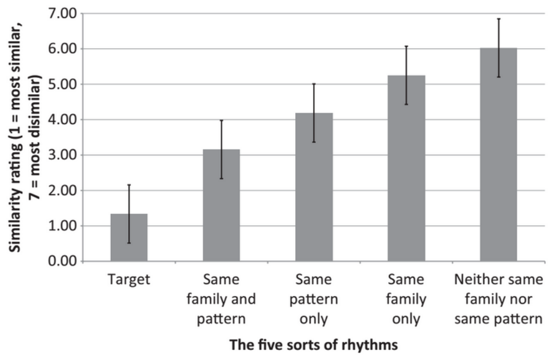

Figure 3.2 Monophonic similarity factors and their relation to perceived similarity by the subjects. Extracted from Cao et al. (2014) experiment 2.

There are two questions regarding the experiments presented by Cao et al. (2014). One is that they do not make specific remarks on how the patterns were reproduced to the subjects so it is assumed that the patterns were presented without any previous pulse-inducing acoustic stimuli. Although a pulse sensation is progressively acquired as a rhythmic pattern evolves in time, the arousal of a pulse and a meter arrives (if the pattern is pulse inducing) after five to ten onsets have passed (Honing, 2012). Given that the patterns in their experiment are of one bar length it is then unclear whether the pulse sensation is present during the similarity judgment process. I suspect the absence of a clear pulse affects the significance of the *semantic* mechanism when assessing the similarity between two patterns, as the foundation of this mechanism is precisely the presence of a pulse sensation. Therefore, hypothetically, if a pulse sensation is present during the similarity comparison of two patterns, the *semantic* mechanism can be active and can thus become more relevant than when the pulse is not induced. This in turn might affect the way in which similarity is processed, as both cognitive mechanisms at play could have different relevance in the judgment. In short, it is possible that the pre-induction of the pulse defines a stronger cognitive guide for the interpretation of a pattern based on the *semantic* mechanism. The importance of considering a pulse-induced context, is that it resembles the situation of music creation, specifically EDM, where rhythm patterns and musical variations are produced within a strong pulse-inducing framework.

The second question arising from Cao et al. (2014) experiments is a lack of detail on the anatomy of the IRs they use. Important matters such as the length and the starting point of the IRs, and mainly the shift in metrical position from pattern to pattern. Although they report an influence of the IR in similarity judgements (using shifted IRs) the design of their patterns is not systematic in regards of the shift. Therefore it is pertinent to evaluate how the shift and the starting point of the IR might have any impact on similarity.

This experiment has then two main objectives:

1. To test the relation between the *syntactic* and *semantic* mechanisms when evaluating similarity. This will be evaluated by analyzing the effect of inducing or not the beat when comparing a same pair of patterns, and then, using two different metrics to explore the results. One metric is the Edit Distance (ED), a distance disengaged from pulse, but specialized in comparing displacements and transformations in patterns (see section 2.3.2). The other metric is the Syncopation Distance (SD), a metric based on Johnson-Laird's rhythm family theory (1991) which is supported on the presence of a pulse, a meter and the reinforcement or challenges to the pulse as syncopations. The ED metric is completely related with the *syntactic* mechanism while the SD is rooted in the *semantic* mechanism (see section 2.3.3).

2. To understand the importance of the shift and the origin of the IRs when evaluating similarity between two patterns, both when the pulse is induced and when it is not.

The experiment is carried out in two stages, one where similarity ratings are given to pairs of rhythms without any pulse induction, and a second phase where the same pairs of rhythms are rated with a pulse induction.

3.2.1 Methods

3.2.1.1 Participants

Twenty-one subjects (19 males, 2 females) were recruited among the Music Technology Group (MTG) staff and UPF pool of Master students to participate as subjects in this experiment. All of them reported musical experience of more than 5 years at least as amateur music performers. Two of the subjects had been educated in non western musical traditions. The subjects were invited to participate freely in the experiment and no reward was offered for their participation.

3.2.1.2 Material

Objective Distance Metrics

Two objective similarity metrics are used to measure the distances between the patterns presented to subjects in the experiments. These metrics are compared with the results of the similarity judgements to observe possible correlations.

The edit distance (ED) is a measure of the transformations one pattern of characters must undertake to become another one. There are three transformations allowed: swapping characters (changing abc for adc), inserting a character (changing ac for abc), removing a character (changing xyz for yz). The edit distance is the sum of these three different transformations made to one sequence of characters in order to become another sequence. As we are dealing here with monotimbral and monophonic patterns composed only of onsets and silences, the patterns are represented by sequences of ones and zeroes, where the character ‘1’ represents an onset and ‘0’ represents a silence. The edit distance between two monophonic patterns is agnostic of any pulse and meter context.

The syncopation distance (SD) is based on the acknowledgment of the pulse and meter induced by a monophonic pattern, and the subsequent syncopations or reinforcements found in the pattern. It is derived from the syncopation family theory, formulated by Johnson-Laird (1991) and later used by Cao et al. (2014) in their experiment. The procedure to compute the SD starts by segmenting a pattern at each pulse. Then, each intra-pulse sub-pattern is classified in different syncopation categories according to its relation with the pulse, either a reinforcement, a challenge or none. The classification presented here is a variation of Johnson-Laird's method, in which beats are clustered in three broad categories: *syncopation*, *reinforcement* or *nothing*, depending if the elements of the beat are a reinforcement, a challenge, or have no interaction with the pulse.

Group	Family	Syncopation Value	intra-pulse sub-patterns
1	R3	3	1010_ 1010x
2	R2	2	1000_1000x 1001x 1011x
3	R1	1	0010_ 0010x 0110_ 0110x 1110_ 1110x
4	N	0	0000_ 0000x 1111x 0011x 0001x 0111x
5	S1	-1	0100_ 0100x 1100_ 1100x 0101x 1101x
6	S2	-2	0001_ 0011_ 0111_ 1111_
7	S3	-3	0101_ 1101_
8	RS	0	1001_ 1011_

Table 3.1. Eight syncopation states for intra-pulse sub patterns used to compute the Syncopation Distance (SD) Metric. The first column is the ID of the syncopation. The third column is the syncopation value; column 4 presents the patterns that belong to each group. Symbols '_' or 'x' indicate a silence or an onset at the beginning of the next beat.

Syncopations are expanded into three possible categories according to their syncopation value (groups 5 to 7, Table 3.1). The reinforcement category is also split in three groups (groups 1 to 3, Table 3.1), according to their syncopation value. Additionally, a new category, where a syncopation and a reinforcement are both present (group 8, Table 3.1), is created. In total there are 8 different categories in which every within-pulse sub-pattern can be classified: three types of reinforcement, three types of syncopations, a ‘nothing’ category and a reinforcement-syncopation category. Expanding the groups in which a beat can be classified from the original three (only syncopations, reinforcements and nothing), to eight in our model, offers more detail on the role of each intra-pulse segment and helps differentiate between different syncopations or different reinforcements (see Table 3.1).

The procedure to classify each intra-pulse sub-pattern is based in computing its syncopation value using the metrical salience profile 2 0 1 0. This profile is derived from Lerdahl and Jackendoff’s GTTM (Lerdahl and Jackendoff, 1984) in which weights are proportional to the duration of the note each accent represents: an accent on a whole note has a higher weight than an accent on a half note, which is higher than an accent on a quarter note, and so forth (see section 2.2). In our beat profile, the first onset that is coincident with the pulse, has a higher weight (2) than the third 16^{th} note (1).

It is important to note that an onset on the fourth step of a sub-pattern generates a syncopation only if the first step of the next beat is a silence. Therefore, to calculate the appropriate syncopation values for every sub pattern, the first step of the following sub-pattern has to be considered. The syncopation value for each sub pattern is the sum of each onset’s metrical weight whenever it is preceding a silence.

Stimuli

Nine one-bar patterns were designed as bases creating variations from them by shifting an identical region (IR). Four variations per base were created, obtaining a total of 36 patterns used as stimuli. The variation patterns were created so that a small fragment of the base pattern, the IR, was displaced 1 to 4 1/16th note steps. When performing this shift, both base and variation patterns contain the same IR but are located at a certain distance from the original position.

Group	pattern A / pattern B

A	**1 0 1 0 1 1** 1 0 1 0 0 0 1 0 1 0 / 1 **1 0 1 0 1 1** 0 1 0 0 0 1 0 1 0 **1 0 1 0 1 1** 1 0 1 0 0 0 1 0 1 0 / 1 0 **1 0 1 0 1 1** 1 0 0 0 1 0 1 0 **1 0 1 0 1 1** 1 0 1 0 0 0 1 0 1 0 / 1 0 0 **1 0 1 0 1 1** 0 0 0 1 0 1 0 **1 0 1 0 1 1** 1 0 1 0 0 0 1 0 1 0 / 1 0 1 0 **1 0 1 0 1 1** 0 0 1 0 1 0
B	**1 0 0 1 0 1 1** 0 0 0 1 0 1 0 0 0 / 1 **1 0 0 1 0 1** 1 0 0 1 0 1 0 0 0 **1 0 0 1 0 1** 1 0 0 0 1 0 1 0 0 0 / 1 0 **1 0 0 1 0 1** 1 0 1 0 1 0 0 0 **1 0 0 1 0 1** 1 0 0 0 1 0 1 0 0 0 / 1 0 0 **1 0 0 1 0 1** 1 1 0 1 0 0 0 **1 0 0 1 0 1** 1 0 0 0 1 0 1 0 0 0 / 1 0 0 0 **1 0 0 1 0 1** 1 0 1 0 0 0
C	1 **1 1 0 1 0 1** 0 1 0 0 0 1 0 0 0 / 1 0 **1 1 0 1 0 1** 0 1 0 0 1 0 0 0 1 **1 1 0 1 0 1** 0 1 0 0 0 1 0 0 0 / 1 0 0 **1 1 0 1 0 1** 0 0 0 1 0 0 0 1 **1 1 0 1 0 1** 0 1 0 0 0 1 0 0 0 / 1 1 1 0 **1 1 0 1 0 1** 0 0 1 0 0 0 1 **1 1 0 1 0 1** 0 1 0 0 0 1 0 0 0 / 1 1 1 0 0 **1 1 0 1 0 1** 0 1 0 0 0
D	1 **1 0 1 0 1 1** 0 0 0 1 0 1 0 0 0 / 1 0 **1 0 1 0 1 1** 0 0 1 0 1 0 0 0 1 **1 0 1 0 1 1** 0 0 0 1 0 1 0 0 0 / 1 1 0 **1 0 1 0 1 1** 0 1 0 1 0 0 0

	1**101011**000101000/1100**101011**001000
	1**101011**000101000/11010**101011**01000
E	10**101011**01010000/100**101011**0010000
	10**101011**01010000/1000**101011**010000
	10**101011**01010000/10100**101011**10000
	10**101011**01010000/101010**101011**0000
F	101**100101**0010000/1010**100101**010000
	101**100101**0010000/10100**100101**10000
	101**100101**0010000/101000**100101**0000
	101**100101**0010000/1010000**100101**000
G	110**101001**0001000/1100**101001**001000
	110**101001**0001000/11000**101001**01000
	110**101001**0001000/110100**101001**1000
	110**101001**0001000/1101000**101001**000
H	10100**101011**01000/101000**101011**1000
	10100**101011**01000/1010010**101011**000
	10100**101011**01000/10100100**101011**00
	10100**101011**01000/101001000**101011**0
I	10010**110101**01000/100100**110101**1000
	10010**110101**01000/1001010**110101**000
	10010**110101**01000/10010100**110101**00
	10010**110101**01000/100101101**110101**0
Control	1001011010101000/1001011011101000

Table 3.2. The 37 patterns used in the experiment grouped by base pattern. The identical region is highlighted in bold.

The original position of the IR was also controlled, so that each group had an IR selected from steps 1, 2, 3, 4 and 6. The size of all the IR is 6 steps, measured in 1/16th note lengths from the first onset to the last. There are 3 or 4 onsets present on each IR. A 37th pair, consisting of two identical patterns, was added for controlling the consistency in the answers. Rhythms are reproduced with a clave sound sampled from the Roland TR-727 drum machine with no dynamic changes. The symbolic representation of the patterns is binary, where a 1 indicates an onset and 0 indicates a silence. Therefore the patterns used throughout this work are coded as 16 digit sequences of zeroes and ones (see Table 3.2) of one-bar length.

3.2.1.3 Procedure

This experiment was carried out in two stages separated by a week. Both stages had the same subjects and the same rhythmic stimuli. The only difference is that on the second stage a pulse was induced before and during the presentation of the rhythm stimuli. Besides this, every other aspect of the experiment was kept. On the second stage of the experiment, a kick drum was played four times on the start of every beat at a tempo of 120 beats per minute, then the kick drum and one of the patterns of the pair were played simultaneously, then just the kick drum again four times and finally the kick drum simultaneously with the remaining pattern, as schematized in Figure 3.3 (bottom).

The system used to carry out the experiment was implemented in Pure Data Extended. It consists of two play buttons to reproduce each rhythm of the pair. After listening to a pair of patterns, subjects rate dissimilarity on a 7-step Likert scale. Levels 0, 2, 4 and 6 of the scale were labeled as ``The same '', ``quite similar'', ``not very similar'', ``not similar at all''. Levels 1,3 and 5 were not labeled.

In the first stage (patterns presented without metrical context), the 37 rhythm pairs were presented in a subject-specific randomized order and without any possibility to listen to them more than once. The tempo was set constant to 120 BPM. On the interface each pattern was played by pressing its corresponding button (labeled as pattern A and pattern B) which was disabled after clicked (see Figure 3.3 top). Once both rhythms were played, the Likert scale was enabled for subjects rating. Once one pair of rhythms was rated, a ‘next’ button to go to the following pair was enabled. When the ‘next’ button was pushed, a 4-seconds pause started and after it the next pair was loaded and the play buttons became active again. This procedure was repeated until all the 37 pairs were ranked.

In the second stage, the 37 rhythm pairs were presented again in a subject-specific randomized order. Both stages of the experiment use the same interface, but on the beat induced stage the play button reproduced the whole sequence of kick, kick + rhythm A, kick, kick + rhythm B (see Figure 3.3 bottom). The tempo was set to 120 BPM and the kick drum was played four times on-the-pulse before the first rhythm was presented. In addition, the gap between the first and the second rhythm was filled with four times on-the-pulse kick pattern.

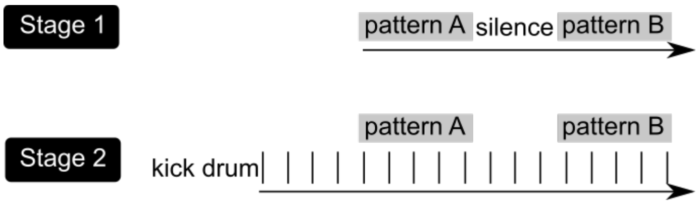

Figure 3.3. Representations of how two patterns A and B are presented in both stages of the experiment. Without pulse induction (top) and with pulse induction (bottom).

3.2.2 Results

Once all the subjects finished the experiment, the mode of the similarity ratings of each pair is used as the representative subjective similarity value in both stages. A general view of the data shows a clear difference between the similarity results obtained for the same pairs of rhythms depending whether they are presented within a rhythmic context or not (Figure 3.4). The between-subject similarity obtained for all pairs in both experimental stages is not convergent. In some cases it is the same (pairs 2, 12, 24, 25, 26 29, 30, 31, 36) in some cases highly contradictory (pairs 3, 4, 9, 13, 17, 18, 22, 23, 27, 35) but generally in disagreement (75% of the pairs). This strongly suggests the same pairs of rhythms are rated differently depending on the presence or absence of a rhythmic context.

To get a better picture of the results of stage 1, pairs with more than 50% of the results scattered over 2 marks were discarded as inconsistent between subjects (see Figure 3.4). Only 16% of the pairs were removed, namely pairs 1, 3, 4, 19, 23 and 35. As with the similarity ratings of stage 1, the analysis of the similarity judgements on stage 2 (when the pulse is induced) is based on the mode of the ratings for every stimuli pair. The same was done for the results obtained in stage 2. The stimuli pairs are analyzed in search for the most consistent inter-subject ratings. Ratings with 50% of the results spread out three or more perceptual scale values are removed, namely pairs 23, 24, 26, 27, 28, 35, being the 16% of the original set.

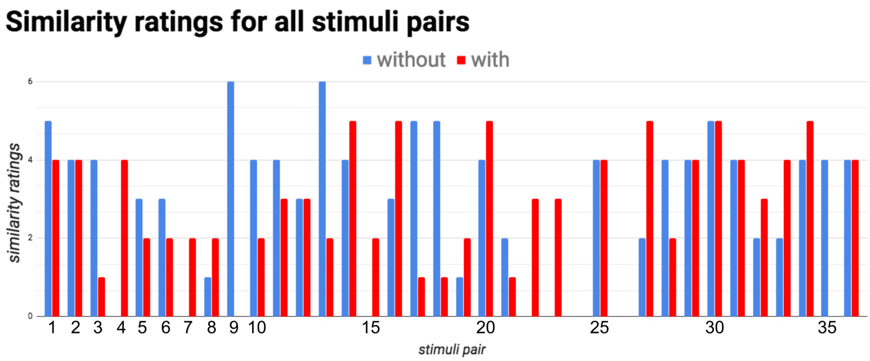

Figure 3.4 Similarity ratings for all stimuli pairs. Results with rhythmic context dark gray, without rhythmic context light gray.

Each pair of stimulus, plotted by its similarity ratings and the shift of the IR, is organized by groups from ‘a’ to ‘i’ in Figure 3.5. There is a trend that suggests that an increase in the shift reduces the similarity rating when the pulse is not induced (Friedman chi-squared = 23.878, df = 4, p-value = 8.45e-05) with significant Spearman rank order correlations of all pattern groups (a: -0.97, b: -0.87, c: -0.95, d: -0.87, e: -0.87, f: -0.89, g: -0.46, h: -0.82, i: -0.22. P-Values a: 0.0048, b: 0.0539, c: 0.0138, d: 0.0539, e: 0.0539, f: 0.0405, g: 0.4338, h: 0.0886, i: 0.7177) (Figure 3.5, top). Low p-values and high negative Spearman rank order correlations, suggest a negative correspondence between shift and similarity, so that the farther the IR is shifted, the lower the similarity rating. These results suggest a relation between the IR and similarity ratings when the rhythms are presented to the subjects without pulse induction. It also expands the features of the IRs such as the size and shift, complementing the results of Cao et al. (2014).

Figure 3.5 Similarity rating vs shift without the presence of a rhythmical context discriminated by groups, from a to i. 6:the same, 0:not similar at all.

Every stimulus pair is compared with the shift of the IR from one pattern to the other, and with the similarity rating obtained when a the pulse was induced (Figure 3.5 bottom). These results show very high p-values (a: -0.67 p-value 0.2152, b: -0.71 p-value = 0.1817, c: -0.05 p-value = 0.9347, d: -0.32 p-value = 0.6042, e: -0.05 p-value = 0.9347, f: -0.67 p-value = 0.2189, g: -0.50 p-value = 0.3910, h: -0.82 p-value = 0.0886, i: -0.67 p-value = 0.2189), which strongly suggest that no aspect of the IR is relevant to assess rhythmic similarity when a pulse is induced.

Possible relations between the two objective measures and the similarity results obtained in both stages, are presented on Table 3.3. The ED is correlated with the results only when the pulse is not induced (rho = 0.52343, p-value = 0.00136) but in the presence of a pulse it presents no significant correlation (rho = 0.25491, p-value = 0.133). This suggests that the onset comparing mechanisms are active in the absence of a pulse, thus influencing similarity sensations. On the other hand, the Syncopation Distance (SD) presents a significant correlation (rho = 0.46, p-value = 0.0098) with the similarity ratings when the pulse is induced, but it shows no significant correlation (rho = 0.0273, p-value = 0.256) with the subjective ratings, when the pulse is not induced. This suggests that the symbol counting of the edit distance, which is associated to the *syntactic* cognitive mechanism, is predominant when the pulse is not induced, but not when the pulse is induced.

Metric	Without Pulse Induction	With Pulse Induction
Edit Distance	**Rho 0.52343, p-value 0.00136**	Rho 0.25491, p-value 0.133
Syncopation Distance	Rho 0.027372, p-value 0.256	**Rho 0.46, p-value 0.0098**

Table 3.3 Spearman Rank correlation values for each objective metric and the similarity ratings without pulse induction

3.2.3 Discussion

The influence of shift in similarity ratings in both experimental stages differs in tendency. While in stage 1 (no pulse induction) shift seems to have an inverse correspondence with similarity, for most of the groups on stage 2 (with pulse induction) no direct relation with the shift is appreciated. Presumably, the emergence of IRs and their shift as a relevant factor for rhythmic similarity, only in the case where there is no pre-induced pulse, could be related to the *syntactic* perceptual mechanism triggered when no metrical cues are offered to decipher a musical sequence in terms of its rhythmical properties. The workings of this so-called *syntactic* mechanism could be analogous to comparing the similarity between two words by looking at the letters and their order and not by the meaning of the words. It seems to be clear that a shallow similarity computation may happen based on superficial features (positions of onsets and silences) and in the absence of a rhythm context.

On the other hand, a more abstract and layered mechanism, a *semantic* one, operates when a metric context is at hand to process the patterns.

This previous observation is aligned with the fact that the amount of notes needed for a beat to be induced is from 5 to 10 (Desain & Honing, 1999). This could lead to conclude that in stage 1, a sense of beat was acquired just as every phrase was ending, and therefore no metrical structure was ever induced during this experiment. Nevertheless as 36 pairs of rhythms, all at 120 BPM were listened during stage 1, a reminiscent notion of the tempo could be accumulated after each exposition and influenced further comparisons. This observation is out of the analysis and all results of stage 1 are treated as non beat inducing. Another factor that is left out, is the possible use of memory to recall a rhythm that just finished with the late acquired meter, so the rhythm is evoked with a meter although such meter was not originally present.

Groups that have lower correlation between the shift and the similarity ratings are groups in which the IR had an origin closer to the start of the rhythm. The farthest the origin, the least correlation between shift and similarity ratings (see Figure 3.6). Spearman correlations for different origins are origin 1: -1.0000, origin 2: -0.8721, origin 3: -0.9000, origin 4: -0.6669. Their respective pairwise two-sided p-values are 0.0001, 0.0539, 0.0374, 0.2189. Although all Spearman correlations are high, their significance decreases progressively as the origin of the IR increases.

The same analysis of the influence of the IR's origin and shift with similarity in stage 2, yields the following Spearman correlation values: origin1: rho = -0.6708204 p-value = 0.2152, origin 2: rho = -0.3162278 p-value = 0.6042, origin 3: rho = 0.00000001 p-value = 1.0000, origin 4: rho = -0.97467943 p-value = 0.0048. It is revealing that none of the p-values accounts for significance, again showing a disengagement between inducing a meter and the relevance of having an IR in two patterns. This can lead to the conclusion that a mechanism based on IRs is not relevant when an induced pulse is present.

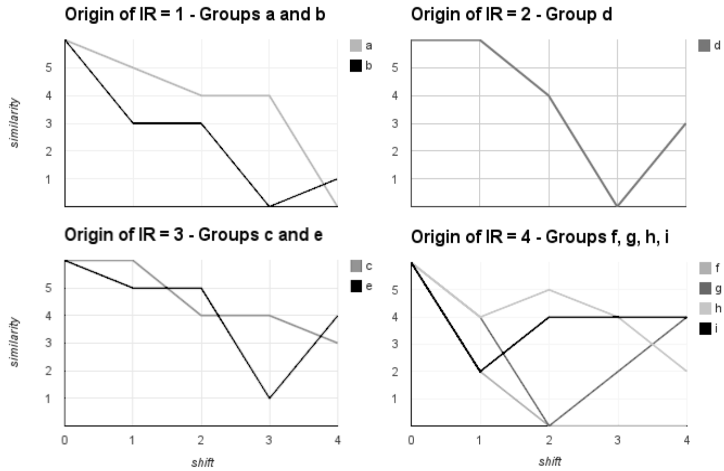

Figure 3.6 Relationship between origin and shift for stage 1.

As a summary, five observations regarding both stages can be presented.

- Similarity ratings of patterns change depending on the presence or absence of a pulse which metrically coincides with the onsets of the patterns being measured.

- In the absence of a pulse, a mechanism based on searching identical regions (IR) of one pattern into the other one is predominant, over coincidences and syncopation, for giving a similarity rating. This mechanism is analogous to a *syntactic* analysis of a sequence of characters.

- Similarity ratings without a rhythmical context are inversely related with the shift in steps of the IR from one pattern to the other.

- The power of an IR mechanism for predicting similarity decreases as the IR moves away from the start of the rhythm.

- In the presence of a pulse, a mechanism based on syncopation is more relevant for predicting human similarity ratings. This mechanism is analogous to a *semantic* analysis of a character sequence given that a rhythmical "meaning", related to the presence of a pulse and a meter, is used to process the similarity between two rhythms.

- The syncopation distance (SD) is a valid metric to predict similarity in monophonic patterns when the pulse is induced.

3.3 Using Beat-Induced Similarity Ratings to Define Similarity Metrics

3.3.1 Introduction

In section 3.2, Experiment 1 presented how similarity ratings between two patterns differ when the pulse is induced and when it is not. It seems that two different mechanisms are in charge of the judgements depending on the the presence or absence of a strong rhythmical context inducing the pulse. A syncopation-based metric, such as the Syncopation Distance (SD), is significantly aligned with human similarity ratings when the pulse is induced, whereas the presence of identical regions (IR) are more important for similarity when the pulse is not induced. When the pulse is not induced, the influence of the identical regions in similarity is higher when the identical region is at the beginning of the pattern. This suggests that there is an asymmetry in the mechanisms involved in judging similarity, weighted towards the start of the patterns. That is, it seems that the first portion of a pattern has higher influence on the subjective similarity ratings.

All the previous conclusions are used in this section to try to create metrics for describing subjective similarity ratings, focusing in the results obtained in the presence of a pulse-inducing stimulus. As mentioned in the previous section, EDM music, and thus its composition and production, is carried out in the presence of strong pulse-inducing rhythms, therefore the metrics of interest for this thesis must be based on pulse-induction scenarios. The SD metric is significantly correlated with the similarity ratings given by subjects. As this metric is based on segmenting patterns at each pulse, it seems that the intra-pulse segments are independent and meaningful fragments of a rhythm capable of transporting important cognitive information. This probable independence of each sub-pattern in relation to subjective similarity judgments, is going to be explored in this section.

The concept of *awareness* is introduced, conceived as weight factors applied to each intra-pulse pattern when computing a distance metric. Conceptually, *awareness* weights emphasize or moderate the importance of each sub section of the pattern on the final distance value.

3.3.2 Expanded metrics for monophonic similarity

Here, two new metrics are proposed. One is the expansion of the SD metric, presented in the previous section, with the additional weighting for the similarity values of each sub pattern. The other metric, is based on measuring the coincidences between each sub pattern of two monophonic rhythms, while also using the awareness weighting for the different sub-patterns. The weightings for each metric will be deduced from the experimental results on the previous section: a multiple linear regression will be computed between the subjective similarity judgements (X) and the similarity obtained between each intra-pulse sub pattern (Y1, Y2, Y3, Y4).

3.3.2.1 Syncopation and awareness distance (SAD)

The Syncopation and Awareness Distance is based on splitting patterns into sub patterns at every pulse (in this case each intra pulse pattern has four steps or four digits), and assigning each sub pattern to a syncopation category, depending on its relationship to the pulse (an expanded explanation is in section 3.2.1.2, see Figure 3.7). For computing the distance between two patterns, first they are both converted to a sequence of syncopation families, assigning a syncopation category to each intra pulse pattern according to Table 3.1. Each sequence of onsets and silences is thus converted to a sequence of families. Then, the coincidence between the family sequence of each pattern is evaluated. Finally, each coincidence value is weighted with a value which is deduced from the similarity ratings of Experiment 1 (see equation in bottom left of Figure 3.7). This metric is expressing the relationship between the different syncopations and reinforcements found in a pattern emphasized (or deemphasized) according to the results of Experiment 1.

The SAD metric is based on comparing if syncopation groups are coincidental between different patterns. This means that a change from one family to any other family is penalized by our algorithm despite if the change is between syncopations (groups 5 to 7 in Table 1) or between reinforcements (groups 1 to 3 in Table 1), or if it is a change from a syncopation to a reinforcement group, or if it is a syncopation of the nothing group (or vice versa).

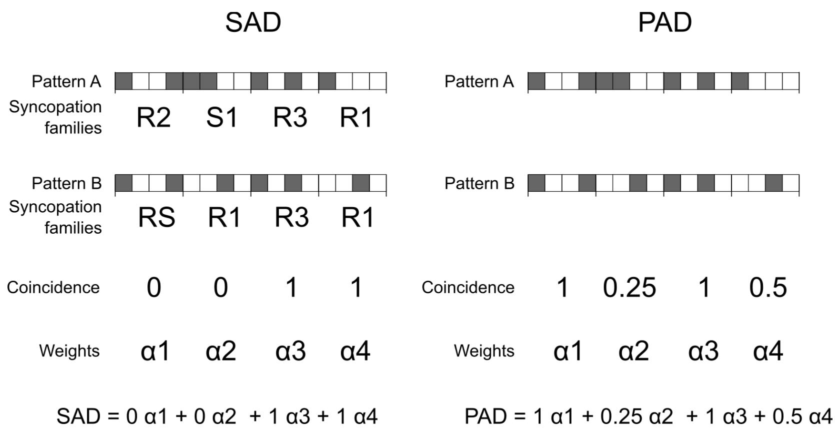

Figure 3.7. How to compute SAD and PAD distances from two monophonic patterns.

3.3.2.2 Pattern coincidence and awareness metric (PAD)

This metric is based on comparing the inter onset sequences between two patterns. As with the SAD metric, the first step consists on segmenting the patterns at each pulse. But for the PAD metric, what is compared is the actual sequence of onsets and silences in each position. The coincidences for each sub pattern are computed as the percentage of onsets and silences located in the same position. As in the SAD, the coincidence of each intra beat sub pattern is weighted according to the results of Experiment 1 (see equation in Figure 3.7 bottom right). This metric expresses the relationship of having identical patterns at different intra-pulse sections taking into account the awareness as extracted from subject ratings of Experiment 1.

3.3.2.3 Deducing the weights for PAD and SAD

Here we use the similarity judgements reported by the subjects in Experiment 1, when the pulse was induced, to find the weights for our newly introduced metrics PAD and SAD. To calculate the weights of PAD, a linear regression between the coincidence result of each beat and the similarity ratings is computed. The normalized weights obtained for beats 1 to 4 are 1, 0.27, 0.22 and 0.16 respectively. To calculate SAD, a multiple linear regression between each beats' coincidence and similarity ratings was carried out. It generated the following normalized weights for beats 1 to 4: 1, 0.075, 0.14 and 0.12 respectively. These values are used as indications of the *awareness* for each beat. The resulting awareness profiles of both PAD and SAD metrics have a similar behaviour (see Figure 3.8). In both cases the importance of the first beat is almost 5 times larger than that of the other beats.

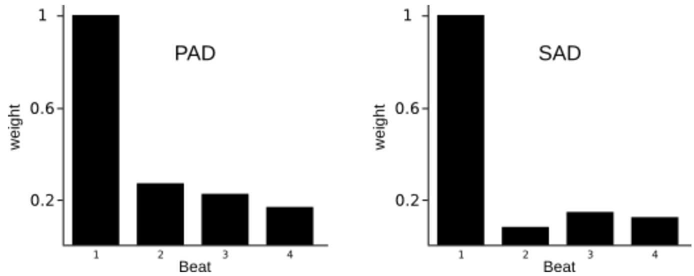

Figure 3.8 Awareness profiles of the PAD and SAD distances that generated best correlations with rhythm similarity ratings.

This *awareness* difference between the four intra-pulse sub patterns, suggests that a difference in the first beat has a higher impact on the similarity sensations than in the rest of the beats. It reflects the importance of the first beats' syncopation families (SAD) and pattern coincidence (PAD), when assessing the similarity sensations between two monophonic patterns.

Using the obtained weight profiles, the PAD distance has a Spearman Rank correlation value of 0.76 (p-value < 0.001) with the similarity judgements while the SAD distance has a Spearman Rank correlation value of 0.81 (p-value < 0.001) (Figure 3.9).

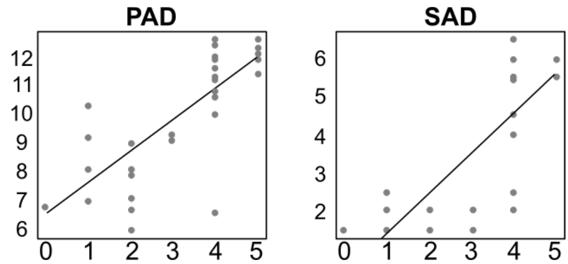

Figure 3.9 PAD and SAD predictions correlated with similarity ratings. X axis: similarity ratings, Y axis PAD and SAD predictions from left to right.

3.3.2.4 Example to compute the metrics

Here, an illustrative example to understand PAD and SAD is presented. The two first beats of a given pattern A, have the following onset/silence configuration 1001 0110, and another pattern B has 1100 0010. Their respective syncopation groups are 8, 3 and 5, 3. Analyzing the syncopation coincidence for the first beat of patterns A and B, we get that 1001 (when the next beats starts with a 0) belongs to family 8 and 1100 belongs to family 5 (see Table 3.1). Clearly 8 is different from 5. At the second beat, 0110 and 0010 both belong to group 3, thus coincidence is 1. The SAD metric consists on weighting each coincidence based on the profiles presented above, so each coincidence will be multiplied by a weight: (0 x 1), (1 x 0.075) and then summed, 0+0.0075= 0.0075.

On the other hand, the pattern coincidence for PAD is computed by looking at the percentage of coincident onsets and silences on the same beat of each pattern. Their coincidence values would be 2/4 = 0.5 because there are 2 out of four notes coincident between 1001_ and 1100 for the first beat. For the second beat, there are 3 coincidences between 0110 and 0010 so the coincidence value is 3/4 = 0.75. The coincidences for each sub pattern are then weighted with the PAD profile presented above: 0.5 x 1 = 0.5 and 0.75 x 0.27 = 0.2025. The PAD value is the sum of the weighted similarity of both beats: 0.5 + 0.2025 = 0.7025.

3.3.4 Discussion and conclusions

Both SAD and PAD have high correlation values with human similarity ratings ($\rho=0.81$, $p<0.001$ and 0.76, $p<0.001$ respectively). This validates the idea of each inter-pulse pattern having a different importance when beat induced subjects try to evaluate the similarity of two monophonic rhythms. These results also show how the first beat is the most important for predicting similarity when using both metrics, followed by the third, the fourth and the second in the case of SAD; and second, third and fourth in the case of PAD.

The linear regression between a syncopation-based metric and the subject-based similarity ratings shows how different inter-pulse weights, as the *awareness*, maximize the correlations between objective predictions and ratings. This manifests how syncopation is a strong predictor for similarity in a monophonic format and also suggests how inter-pulse patterns are important units of analysis in rhythmic processing.

3.4 Symbolic Descriptors for Polyphonic Drum Similarity

As this thesis aims to develop compositional tools for EDM, especially dealing with drum rhythms, there is a clear need to understand how polyphony affects the processing of musical rhythms and thus the notion of similarity between two of them. What has been learned in the previous experiment, in a monophonic context, will be expanded and combined with other studies on polyphonic drum similarity. In order to understand the mechanisms underlying human processing of polyphonic rhythm, and to be able to elaborate models that simulate that processing automatically, three different sources of knowledge are revised and integrated into one main research methodology: the experiments of Gabrielsson (1973b), contemporary experiments on polyphonic processing of rhythms (Witek, 2014a, 2014b; Hove, 2014; Burger et al. 2017) and the results from the two previous sections 3.2 and 3.3.

First, the experiments of Alf Gabrielsson (1973a, 1973b) are a strong precedent in establishing experimental procedures and providing results in the subject of polyphonic rhythm similarity (find a review in section 2.3.3). One of the main contributions is the presentation of his similarity experiments as *rhythm spaces* and a list of factors that, he concludes, influence similarity judgements. These factors are:

- The meter induced by the sequence.
- The onset density of the patterns.
- The simplicity-complexity of the patterns.
- The syncopations.
- The number of different instruments in a sequence.
- The “movement character” of the rhythms.

A second source are the experimental results by Hove (2014), Bouwer et al. (2014), Witek et al. (2014a) and Burger et al. (2017), which present advances in the comprehension of how humans process polyphonic rhythms. In these studies, the importance of the main frequency of the different instruments of a polyphonic drum pattern is reported to influence listener’s rhythm processing. Depending on the predominant frequency of a drum sound, it might have a stronger power for disturbing or for confirming the meter of a polyphonic pattern. All these studies conclude that the instruments with the lowest frequency (e.g. the kick drum) have a higher impact in the establishment of a pulse or to disturb it (e.g. a syncopation), than high-frequency instruments (i.e. the hi-hats). This view is backed experimentally by Witek et al. (2014b), who devise a polyphonic syncopation metric based on three different instrument ranges: low, mid and high, represented by the kick drum, the snare and the hi-hat respectively. This metric is used to study the impact of syncopation in the desire to dance.

The third source is the experiment presented in the previous section, where two distance metrics are developed from the results of Experiment 1 (section 3.2). The relevance of these metrics is grounded in three ideas:

- The syncopations present in a monophonic rhythm are a source of differentiation, which influences subjects when processing the resemblance between two rhythmic patterns. This applies when a pulse is induced while listening to the patterns being compared.

- The coincidence of the onsets and silences between two patterns is another source of differentiation related with the way subjects process the similarity between two rhythmic patterns.

- Different sections of a monophonic rhythmic pattern have different importance in the process of establishing a subjective similarity. Experimental results suggest that the first section of two patterns being compared is the most important when establishing a similarity sensation. This might reflect the so-called "primacy" effect in tasks that involve short-term memory retention of information in which first items are best remembered (Tulving & Clark, 2000).

Incorporating these sources of knowledge into a unified perspective on polyphonic rhythm similarity, some conclusions are drawn to lead the exploration of new methods for establishing relationships between EDM drum patterns. There is a clear relevance of syncopation and meter in the processing of monophonic and polyphonic patterns, that is transversal to the three different sources presented above. A differentiation between three different frequency ranges for percussive sounds (low, mid and high) is fundamental, as the most energetic frequency bands of a sound affect the way in which the rhythm it produces is processed. This seems intuitively related to human perception as a mechanism to provide distinction, while avoiding frequencial overlapping between instruments. From Gabrielsson's fundamental factors influencing the similarity of two rhythms we can keep the density, or the amount of onsets in a pattern, along with the number of different instruments in a pattern, and the "character" of a rhythm. This wide combination of factors is used here to define a new comprehensive set of descriptors which can be extracted from symbolic rhythmic sequences. These descriptors are designed in order to capture the different qualitative factors mentioned, in simple and straightforward algorithms.

Our focus on symbolic patterns resides in the need to discard the effect of timbre in subjective similarity measurements. As it has been presented in section 3.2, and also coinciding with Gabrielsson's methodology, the use of consistent timbres when rendering percussive patterns allows for the subjects to focus only on rhythm when undertaking experiments. Otherwise, the effect of timbre affects the similarity sensation between a couple of patterns distorting the way in which the resemblance is measured. Thus, these descriptors differ from others used in automatic rhythm classification research (for example Gouyon et al., 2004) as the ones that will be presented here are i) based on notions of human rhythmic processing, and ii) not based on audio signal analysis.

3.4.1 Symbolic drum pattern descriptors

Here we will adapt some ideas reviewed in section 2.3.3 where the factors influencing polyphonic similarity sensations are related to metrical weight and also to the frequency range of the instruments involved in the polyphonic arrangement. Consequently, we will take advantage of the typical acoustics of percussion sounds, a mapping from the General MIDI Level 1 Percussion Key Map[23] (GMPKM) to three instrument categories (low, mid and high) is defined (see Table 3.4). It is based on the typical spectral center of the sound (i.e. a low tom belongs to low frequency and a snare to the mid frequency instruments). This mapping allows a drum pattern compliant with the GMPKM to be converted from an arbitrary number of parallel instrument patterns into three streams of monophonic percussive patterns, namely *low*, *mid* and *high*.

General MIDI note	Name	Category
35	Acoustic Bass Drum	low
36	Bass Drum 1	low
37	Side Stick	mid
38	Acoustic Snare	mid
39	Hand Clap	mid
40	Electric Snare	mid
41	Low Floor Tom	low
42	Closed Hi Hat	high
43	High Floor Tom	mid
44	Pedal Hi-Hat	high
45	Low Tom	low
46	Open Hi-Hat	high
47	Low-Mid Tom	low
48	Hi-Mid Tom	high
49	Crash Cymbal 1	high
50	High Tom	mid
51	Ride Cymbal 1	high
52	Chinese Cymbal	high
53	Ride Bell	high
54	Tambourine	high
55	Splash Cymbal	high
56	Cowbell	high
57	Crash Cymbal 2	high
58	Vibraslap	high
59	Ride Cymbal 2	high
60	Hi Bongo	high

[23]The General MIDI standard has a list of 46 percussive instruments which are mapped one-to-one to a specific note. This is used to indicate what sort of sound will be heard when that note number is selected on a General MIDI synthesizer. https://www.midi.org/specifications/item/gm-level-1-sound-set.

61	Low Bongo	mid
62	Mute Hi Conga	high
63	Open Hi Conga	high
64	Low Conga	low
65	High Timbale	mid
66	Low Timbale	low
67	High Agogo	high
68	Low Agogo	mid
69	Cabasa	high
70	Maracas	high
71	Short Whistle	high
72	Long Whistle	high
73	Short Guiro	high
74	Long Guiro	high
75	Claves	high
76	Hi Wood Block	high
77	Low Wood Block	low
78	Mute Cuica	mid
79	Open Cuica	mid
80	Mute Triangle	high
81	Open Triangle	high

Table 3.4 General MIDI Level 1 Percussion Key Map instruments, note number, name and category.

This procedure of using only three streams is an adaptation of the methodology used by Witek et al. (2014-2) backed in experiments by Hove (2014), Bouwer et al. (2014), Witek et al. (2014-1) and Burger et al. (2017). It also resonates with the Auditory Scene Analysis theory (Bergman, 1990) in which the multiple and concurrent data generated during the parallel analysis processes in the cochlea and the auditory nerve are simplified into a small number of auditory streams.

Once a symbolic drum pattern is converted into a combination of three band-wise patterns, they are analyzed according to the different factors pointed out by Gabrielsson, to influence similarity at a polyphonic level: syncopations, densities, number of instruments, the meter, the simplicity-complexity of the patterns, and the movement character of the rhythms. The crossover between the three instrumental levels and the sources of information are presented in Table 3.5. The different equations for the descriptors are presented below.

The computation of these descriptors, assumes a symbolic and polyphonic drum pattern in which the percussive instruments are compliant with the

GMPKM, and which has a minimum time resolution of 1/16th note. In order to compute the descriptors, a polyphonic pattern is converted to a triad of symbolic monophonic percussive streams (low, mid and high) using the mapping presented in Table 3.4. The descriptors are quantified as presented in the following subsections.

<table>
<tr><th>Frequency range</th><th colspan="2">syncopations</th><th colspan="3">densities</th><th>number of instruments</th><th>complexity</th></tr>
<tr><td>high</td><td>hisync</td><td rowspan="3">polysync</td><td>hiD</td><td>hiness</td><td rowspan="3">StepD</td><td rowspan="3">NOI</td><td>hisyness</td></tr>
<tr><td>mid</td><td>midsync</td><td>midD</td><td>midness</td><td>midsyness</td></tr>
<tr><td>low</td><td>losync</td><td>loD</td><td>lowness</td><td>losyness</td></tr>
</table>

Table 3.5. List of the different descriptors used, where the concepts of syncopation, frequency range and density (the amount of onsets per time unit) and complexity are combined to define quantifiable measures. [24]

3.4.1.1 Number of instruments (NOI)

This is the simplest metric to compute as it is just the amount of different instruments present in the symbolic polyphonic pattern.

3.4.1.2 hisync, midsync, and losync

The syncopations are quantified following Longuet-Higgins and Lee (1984) (section 2.2.2.1), a method based on a nested metric profile similar to the one presented by Lerdahl and Jackendoff in the GTTM (1985). For each stream, when an onset is followed by a silence its metrical value is extracted. The sum of all the metrical values extracted is the total syncopation value for each stream. These values are reported as the *hisync*, *midsync* and *losync* respectively.

3.4.1.3 Polysync

Polyphonic syncopation is computed with the method proposed by Witek

[24]Code available here: https://github.com/danielgomezmarin/rhythmtoolbox.

et al. (2014b). A single value from the complete polyphonic pattern is extracted combining metrical weight (see section 2.2.2) and their newly introduced instrumental weight. Their algorithm to compute polyphonic syncopation assigns a weight inversely proportional to the instrumental frequency range (low-frequency range has higher weight than mid-frequency and high-frequency). The algorithm is fully documented in the Supporting Information section of the paper.[25]

3.4.1.4 hiD, midD, loD

Sum of onsets for each different instrument group, divided by the total number of steps in the pattern.

3.4.1.5 losyness, midsyness, hisyness

Quotient of the syncopation value and the sum of onsets for each instrument group.

3.4.1.6 stepD

Sum of the steps in the pattern which contain at least one onset, divided by the total amount of steps.

3.4.1.7 lowness, midness, hiness

Share of the total density of patterns that belongs to each of the different instrument categories. Computed as the quotient between the densities per instrument category and the total density.

3.5 Experiment 2: Objective Similarity in Alf Gabrielsson's Rhythm Spaces

Having established a set of polyphonic drum descriptors, the next step is

[25] Link to Witek et. al. Polyphonic syncopation algorithm: https://doi.org/10.1371/journal.pone.0094446.s012

to test their performance in different real-life musical scenarios. As has been presented in the conclusions of chapter 2, Gabrielsson published a study on polyphonic drum rhythms similarity which will be the starting point for this experiment (Gabrielsson, 1973b). Specially, the results of Gabrielsson's Experiments 1 and 2 (GE1 and GE2) will be analyzed, given the peculiarities of the rhythms selected for his experiment: they are reproduced with the same synthetic timbres of a drum machine using the same tempo. These factors dismiss the intrusion of tempo and timbre in the listener's evaluation of similarity. They are also profitable as they resemble the same natural conditions of compositional work in EDM, where the tempo of a *dance track* or even a DJ session concatenating several *dance tracks* are kept at a constant tempo (Collins et al., 2013).

The patterns used by Gabrielsson are factory presets of the Ace Tone Rhythm Ace FR-3 drum machine[26] which were recorded to magnetic tape. The patterns used in GE1 and GE2 were foxtrot, rockn'roll, rhumba, beguine, habanera and waltz[27]. In the procedure reported for GE1 subjects listened to triads of rhythms, and then selected which pair is the most similar (see Table 3.6). The subjects were ten male and 6 female musicians defined as subjects who had performed music for at least four years, most of them as amateur musicians. The subjects had mixed characteristics regarding their experience with different kinds of music and musical instruments. The median value for the number of years performing music was eight years.

Expt. 1

	F	H	RR	R
H	67			
RR	107	81		
R	35	37	94	
B	27	66	93	188

Table 3.6 Similarity counts for each pair of rhythms in GE1 as reported in

[26] https://www.gearogs.com/gear/8813-Ace-Tone-Rhythm-Ace-FR-3
[27] Video of the FR-3: https://www.youtube.com/watch?v=BHiwVcQkKP4

Gabrielsson's original paper.

Converting Table 3.6 with the pair similarity counts into a dissimilarity matrix, a Multi Dimensional Scaling (MDS) algorithm is used with it to obtain a bidimensional space where all five patterns are located according to their reported distance. A similar procedure is followed fro GE2, being the only difference the additional use of waltz pattern in the rating procedure. The spaces obtained by Gabrielsson are graphic representations of the dissimilarity matrix (Figure 3.10). Gabrielsson's interpretations of the axes spanning the spaces are educated guesses without strong grounding on empirical data, therefore in our experiment it is sought to approach them with the help of our rhythm descriptors presented above.

In Gabrielsson's paper the patterns from the FR-3 drum machine are transcribed to symbolic musical notation. The proposed polyphonic descriptors will be extracted from these transcribed patterns obtaining a descriptor vector for each pattern. These vectors will be used to approach the positions of each pattern in the rhythm space resulting from GE1 (Figure 3.10). A Lasso regression (Tibshirani, 1996) will be used to discriminate which descriptors are sufficient (and how important they are) to predict the position of each pattern in the 2D space. Lasso regression is a method typically used for variable selection, to enhance the predictability and interpretability of a model, by reducing the number of variables needed to get accurate predictions. Thus, the resulting set of descriptors should capture the essence of subjects' ratings revealed through the structure of the spaces. The hypothesis is that such a set of descriptors can be good predictors of Gabrielsson's spaces. If this is the case, then we could further ask if they can be generalized to predict different spaces with patterns from other drumming styles.

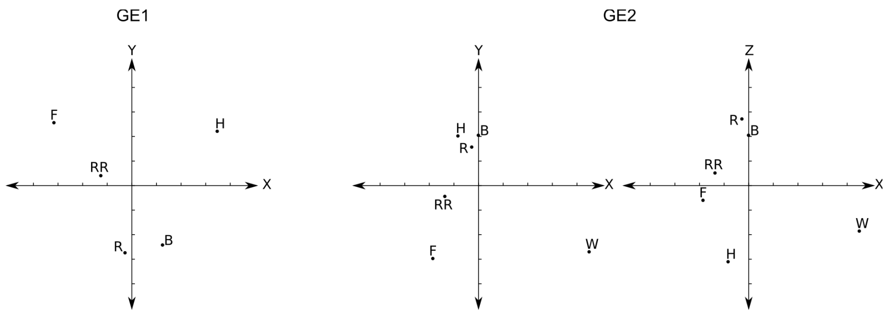

Figure 3.10. Gabrielsson's spaces from Experiments 1 and 2 (GE1 and GE2). F: Foxtrot, RR: Rockn'roll, R: Rhumba, B: Beguine, H: Habanera, W: waltz.

3.5.1 Methods

3.5.1.1 Materials

The resulting spaces from Gabrielsson are used as a source from where the coordinates of each pattern are extracted. These coordinates are presented in Table 3.7.

Pattern	GE1 coordinates	GE2 coordinates
foxtrot	-0.640, 0.55	-0.4, -0.58, -0.1
rocknroll	-0.270, 0.09	-0.3, -0.1, 0.1
rhumba	-0.060, -0.55	-0.09, 0.3, 0.55
beguine	0.240, -0.49	0.0, 0.4, 0.41
habanera	0.700, 0.45	-0.18, 0.41, -0.61
waltz	-	0.85,-0.51,-0.36

Table 3.7 Patterns used in GE1 and GE2 and their coordinates.

3.5.1.2 Procedure

Each pattern used in GE1 and GE2 is transcribed to MIDI format and all descriptors are extracted from them. Then a Multi Task Lasso regression is used (alpha 0.03) using the positions of the patterns in each space as a target (X) and the matrix of descriptors as variables (Y1, Y2, Y3…). The

Lasso regression returns a subset of variables and weights that maximize correlation with the positions of the patterns in each space.

3.5.2 Results and discussion

The output of the Lasso analysis shows that using this set of descriptors [midD, hiD, hiness, lowsync, hisyness] (see Table 3.5) the results of both axes of GE1 are perfectly linearly correlated. For E1 and x axis the weights -0.660, -0.86, -0.068, -0.266, 0.118 respectively present a Spearman correlation of 0.999 (p-value < 0.005). For E1 and y axis the weights respectively present a Spearman correlation of 0.999 (p-value < 0.005). For the space resulting from E2, Lasso analysis shows the descriptor set [midD, hiness, lowsync, midsync, hisync, losyness, hisyness] yields perfect correlations with its three axes. For E2 and x axis the weights are 0.785, -0.073,0.242, 0.574, -0.709, 0.044, 0.506 respectively and present a Spearman correlation of 0.999 (p-value < 0.005). For E2 y axis the weights are 0.333, -0.07, 0.21, -0.031, 1.005 0.002, -1.411 and preents Spearman correlation of 0.942 (p-value < 0.005). For the z axis the weights are 0.313, -0.032, -1.12, 0.104, 0.81, -0.052, -0.723 and presents a Spearman correlation of 0.999 (p-value < 0.005).

The sets of descriptors and weights, perfectly describe the spaces reported in GE1 and GE2 and thus might be related with the overall polyphonic rhythm similarity sensations from which the space was created. It could be fair to presume that, if the space captures a similarity sensation, these descriptors play a role in our perception of polyphonic similarity that could go beyond its particular application in GE1.

	Syncopation	Sync/Density	Density	Density%
High	*	* *	*	* *
Mid	*		* *	
Low	* *	*		

Table 3.8 Relevant descriptors for the prediction of both axes of GE1 space. The rows are the three instrumental categories and the columns the types of descriptors that were relevant. The asterisk represents its use in either of the axes.

From the summary presented in Table 3.8, Syncopation is clearly a main factor for differentiating the patterns at the three different instrumental levels. Syncopation on the low instrumental group is used to predict both spaces. The quotient between the syncopation and the density is found relevant in the high and low frequency instrumental group. For both spaces the high density quotient is found relevant. The density of the high and mid category of instruments is relevant, both spaces coinciding in the importance of density in the mid instrumental level. The density percentage (instrumental density divided by total density) is found relevant for both spaces in the high instrumental category. Descriptors in the three instrumental categories are used to predict GE1 suggesting that they are useful for the human discrimination process of polyphonic rhythms. This, validates the approach of mapping instruments to categories in the symbolic domain as discussed in section 3.4.1.

The highlighted importance of the low syncopation, given its usefulness to predict both GE1 and GE2 spaces, is a confirmation of the argumentation exposed at the introduction of this section. The importance of low frequency instruments in the definition of a syncopation sensation in a polyphonic context is discussed by different authors (Witek et.al. 2014b; Hove, 2014; Burger et. al., 2017; Bouwer et. al., 2014) and has been corroborated in this experiment. This fact is also aligned with Gabrielsson's results as he argues syncopations are one important driver for discrimination of rhythms in polyphonic contexts.

The density of the patterns, another of the factors described by Gabrielsson to influence similarity, is definitive for predicting the positions of the patterns in both axes. Density, discriminated by instrumental groups in the form of midD and highD (density column in Table 3.8) and hiness (density% column in Table 3.8) are relevant discriminators of similarity. It is important to note that the low frequency densities, lowD and lowness are not present in the set of relevant descriptors.

Although the number of different instruments in a sequence is another factor proposed by Gabrielsson to influence similarity and it is one of the descriptors computed, it has no relevance for the prediction of space GE1.

3.6 Experiment 3: Generalizing Polyphonic Descriptors for EDM

In the previous experiment (section 3.5) small sets of descriptors are found to quantitatively describe and predict the results (similarity ratings, perceptual space computed from them) of Gabrielsson's GE1 and GE2 (1973b). The question here is how general these features are. In other words, are these features fitted to the particularities of the rhythms used by Gabrielsson or would they work when other, quite different, rhythm patterns are rated? From a computational point of view this means that, if we have a new set of drum patterns and their location on a human-based rhythm space, by extracting only the descriptors defined on the previous experiment, and by using a dimensional reduction technique, the rhythm space can be predicted with some accuracy. Two very well known dimensional reduction algorithms are used, namely PCA and MDS. PCA finds a principal vector in the descriptors space in which the values of all descriptors are maximally dispersed and then additional orthogonal vectors are found to conform the predicted rhythm space. The other alternative, MDS, is based on a dissimilarity matrix of the patterns, which is computed as the euclidean distance between the descriptor vectors of each pattern (see section 2.3.1.1 for a comprehensive review).

For this experiment a new EDM rhythm space is created, based on subject ratings following Gabrielsson's methodology: selecting a collection of EDM patterns and presenting pairwise combinations to subjects who report their similarity, and then, with those answers, creating a new rhythm space. An EDM drum rhythm collection is compiled specifically for this experiment.

3.6.1 Methods

3.6.1.1 Participants

A total of 36 subjects participated in the survey, 5 females and 31 males, all had experience in music production or musical training.

3.6.1.2 Material

In order to get a subject-based rhythm space, a set of rhythm patterns is needed, so we turned to the EDM production literature and collected drum patterns explicitly associated to a certain EDM style (section 2.5.2). All patterns are 16 steps long, each step lasting for a 16th note. A total of 75 different patterns were collected, 70% of them belonged to the most prominent styles, House Music (28%), Breakbeat (26%), and Techno (16%) and the rest 30% belonged to Garage, Drum n' Bass, Hip-Hop, Trance, Chillout, Dubstep, Jungle and Trip-Hop.

With the whole 75 pattern collection we created a preliminary rhythm space. First we extracted the complete list of aforementioned descriptors and then using PCA they were visualized in a bi-dimensional space. This preliminary space was divided in nine equal-size rectangular areas and then one pattern from each area was selected. In this way, the list of 75 patterns is reduced to 9 patterns (see Table 3.9), each one intended to be representative enough of the variability of the included categories. The 9 patterns (from Techno, House and Breakbeat styles) selected for the experiment are rendered to audio (in order to be played in the rating experiment) using single shot samples from the Roland 707, 808 and 909 drum machines which are instruments typically used in the styles we are focusing on. All selected patterns use instruments included in this set: Low Conga, Bass Drum, Side Stick, Maracas, Hand Clap, Snare, Closed Hi-Hat, Low Tom and Open Hi-Hat (Figure 3.11).

	Left	Center	Right
Top	Techno grinding analogue	techno industrial	techno hardcore
Center	deep house	dirty house	deep tech house

Bottom	break synthetic subs	funk break	break funky drummer

Table 3.9 Patterns selected from the sub regions of the preliminary space. Left, center, right and top, center bottom represent the subdivisions of the space as explained in the text.

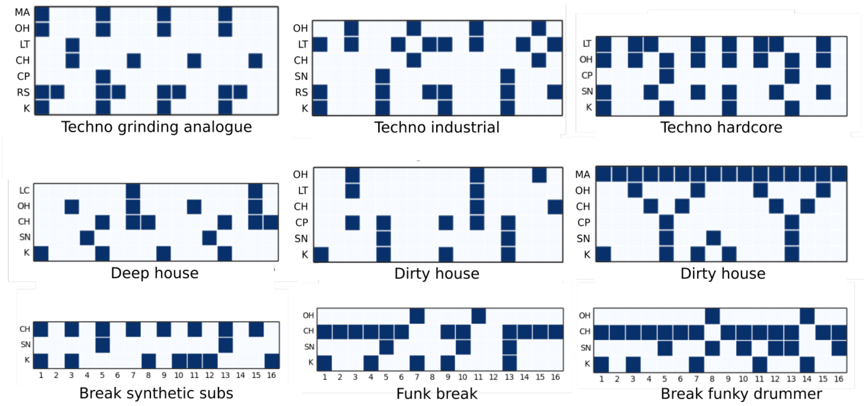

Figure 3.11 Nine EDM patterns selected out of the 75 pattern list. Each pattern was selected for being the farthest of its region. K:kick drum, RS: rimshot, SN: snare, CP: clap, CH: closed hi-hat, OH: open hi-hat, LT: low tom, HT: high tom, LC: low conga, MA: maracas.

3.6.1.3 Procedure

A computer program is prepared to carry out the experiment. Before the subjects start the experiment, several patterns are presented to expose the timbre range of the percussive sounds used, and also examples of "identical", "similar" and "completely different" pairs of patterns are provided as reference. To evaluate all combinations between the 9 patterns, subjects rate the existing 36 possible pattern pairs in a triangular 9 element matrix (avoiding comparing a pattern with itself or repeating any pair). Additionally, 4 randomly selected pairs are presented twice for controlling the consistency of each subject's ratings, so, in total, subjects rate 40 pairs of patterns. The pairs are presented in a random order preventing the same pattern to be in consecutive pairs. Before a pair is reproduced, the order of its two patterns is also randomized so the same pair is presented in the two possible arrangements (i.e., a-b or b-a) to different subjects. Subjects listen to the same pair as many times as they need and the similarity value is reported in a Likert scale with a range from 1 to 10, where 1 means the pair is completely dissimilar and 10 means the pair has the topmost similarity (i.e., the pair contains equal patterns). When the subjects complete the experiment, they answer some questions about themselves: age, gender, years of musical training, years of musical performance training, years of percussive musical performance training, hours per week spent attentively listening to music, experience in electronic music production, experience in electronic drum programming, number times listened to the pairs before answering. Finally, the possibility to leave a comment on the experiment is provided.

3.6.2 Results

In order to simplify the analysis, the 10 point scale is mapped to a 5 point scale where each range of the new scale groups two values of the original scale (1 groups the results of ratings 1 and 2, 2 groups ratings 3 and 4 and so on). Three subjects rated different pairs as being "exactly the same", and therefore these subjects were discarded from the experiment, because there were no identical pairs (i.e., we considered the subjects were not properly attending to the task). The control pairs were used to perceive distortion in the ratings of the same pairs, and the average of the maximum difference of all subjects when rating the same pair is 1.8 units which is a 36% of maximum variation. In order to approximate our analysis to that of Gabrielsson, we create a subgroup of subjects compliant with the musical background reported in his experiments, which is "amateur musicians who had performed music for at least 4 years". A subset of our General Group composed of 18 subjects with at least 4 years of musical training was defined and we will refer to it as the Musicians group. For the General group, the inter quartile range (a measure of statistical dispersion) mean is 1.81 units and 1.48 units for the Musicians group suggesting more agreement in the Musicians sub-group. Pair (1, 3) presents a slight bimodal behaviour for the General Group which is reduced in the Musicians Sub group.

The observed means for each assessed pair present slight differences when both groups are compared using the median rating values for each pair. Only 9 pairs out of 36 differ in median value from one group to another: 6 pairs present changes in a degree of 2 units, and 3 pairs present changes in a degree of 1. Pairs that involve rhythm Deep House do not change between groups and the pairs that involve rhythm 6 Deep Tech House have 4 changes between groups. The difference between the spaces generated by the Musicians group and the general group were not too big to be considered diverted from the General group so we operate with all results.

An MDS is applied to the obtained dissimilarity matrix generating a bi-dimensional space (Figure 3.12). We can observe in the obtained space that the three genres from where the rhythm patterns were extracted span across three distinct regions of the space. Breakbeat patterns are located in the positive region of the X axis, while Techno and House patterns are located from the zero to the negative portion of the X axis. The X-negative quadrants of the space contain, in the Y-positive region the Techno patterns, and in the Y-negative region the House patterns. In EDM, rhythm and timbre are the most salient musical characteristics to define styles (Butler, 2001), so it is relevant for EDM drum patterns to carry important stylistic/similarity information. This stylistic information comes through, in the resulting subject-based EDM space, as patterns of he same style end up located in specific and independent regions.

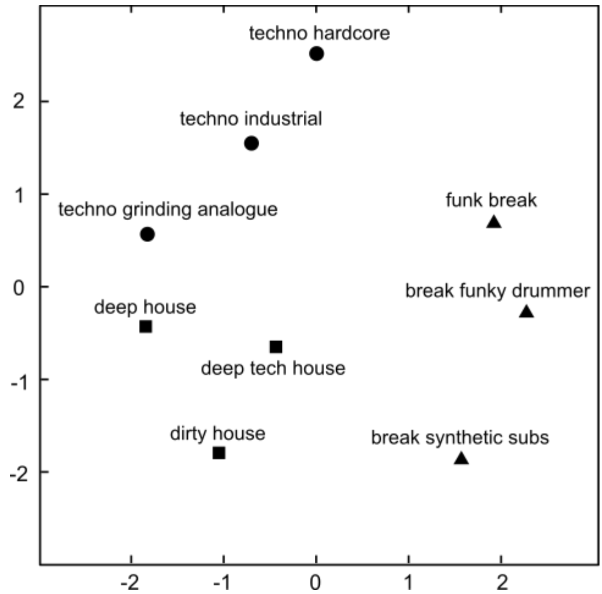

Figure 3.12. Bi dimensional space obtained by using MDS on the dissimilarity matrix of subject ratings.

We extract the two descriptor sets found in the previous experiment (by adjusting E1 and E2) from the patterns in the EDM space. The descriptor set for E1 is [midD, hiD, hiness, lowsync, hisyness] and for E2 is [midD, hiness, lowsync, midsync, hisync, losyness, hisyness]. Using those descriptor values we compute PCA and MDS to evaluate if the locations of the EDM patterns can be predicted with any of the sets. Table 3.10 presents the correlations between the predicted space and the resulting EDM space.

Using the E1 set of descriptors for analyzing the patterns of the EDM experiment and then applying MDS to those results (E1 set MDS), we observe correlations of $\rho=0.67$ (p-value < 0.05) and $\rho=-0.78$ (p-value < 0.05) for x and y axis respectively. The other two combinations of sets and dimensional reduction techniques that are borderline correlated with the EDM space are, E1 set PCA and E2 set MDS but none of them have statistical significant correlations for any of the axes. The method of using the E1 set and then MDS captures the distance sensations reported by the subjects both in Gabrielsson's experiment 1 and in our EDM experiment.

	X	Y
E1 set PCA	$\rho = 0.62$ p=0.076	$\rho = 0.68$ p=0.042
E1 set MDS	$\rho = 0.67$ p=0.049	$\rho = 0.78$ p=0.012
E2 set PCA	$\rho = 0.52$ p=0.154	$\rho = 0.45$ p=0.224
E2 set MDS	$\rho = 0.63$ p=0.067	$\rho = 0.683$ p=0.042
E1+E2 set PCA	$\rho = 0.466$ p=0.205	$\rho = 0.683$ p=0.042
E1+E2 set MDS	$\rho = 0.383$ p=0.308	$\rho = 0.5$ p=0.17

Table 3.10 Spearman Rank correlations between each EDM axis and the prediction using the descriptor sets from GE1 rendered using PCA and MDS.

3.6.2.1 From EDM to GE1

The inverse process, making a Multi-Output Lasso analysis to extract a set of EDM well-correlated descriptors, and then finding how they can predict the GE1 space was also explored. In this case, the resulting descriptor set that better correlates ($\rho = 1$, p-value < 0.05) with the EDM axes was different. However the predictions towards the GE1 space, either by using PCA or MDS with the EDM set, were not statistically significant. This is expected as Gabrielsson's patterns are much different among them, representing six different musical styles, and thus they cover a large perceptual/musical space, while the patterns used in the EDM experiment are variations of three styles which cover a small-scale rhythm space. In this case, the macro-scale descriptors can predict the small-scale space but not the opposite (Figure 3.13).

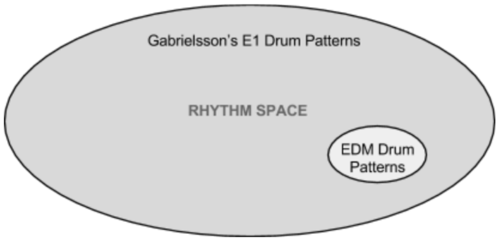

Figure 3.13. Comparison between Gabrielsson's space and the EDM space.

3.6.2.2 Relevant set of descriptors

The descriptors derived from the Multitask Lasso analysis that has the best fit with GE1 space and our EDM experiment using MDS are [midD, hiD, hiness, lowsync, hisyness]. These descriptors cover all frequency ranges in which the drum patterns are segmented (low, mid and high). The only low frequency descriptor present is lowsync which is expected given the crucial importance of the syncopation of the low frequencies in the overall syncopation sensation of a drum pattern, as proposed by Hove et al. (2014). The mid frequency range descriptor midD represents the normalized onset density of the mid frequency. The high frequency descriptors are hiD, hiness and hisyness, all related with the density and the syncopation of the instruments mapped to the high frequency category.

3.6.3 Discussion

By defining a broad set of descriptors and using them to fit symbolic rhythms as defined in Gabrielsson's spaces (1973b), we discovered descriptors that allow to construct very general rhythm spaces that were reported long ago, but without such descriptors-based analysis. Then, we have seen that these descriptors, based on main concepts of rhythm cognition, allow to construct spaces constrained to EDM, which even have a stylistic significance. Consequently, we could use those descriptors in systems that present, visualize and manipulate pattern collections. This way, users could have 2D representations close to their mental representations, which could be exploited for search, selection and invention tasks.

Although the reported experiments were not designed for classification, it is revealing that the concept of EDM style comes through in the space resulting from our second experiment. It can be seen as a demonstration of how musical concepts as House, Techno and Breakbeat emerge based on listening to a handful of instances, each occupying a specific region of a conceptual space. As 61% of the Musicians group reported having experience in EDM production, the distribution by styles can also represent the effect of a pre-existing knowledge about EDM, affecting how patterns are perceived and their similarity judged.

The drum patterns used in Gabrielsson’s experiment 1 come from different dance music cultures, namely Western, Afro Cuban and South American. The patterns in the EDM set belong to three EDM subgenres: Techno, House and Breakbeat. Patterns in E1 contain more cultural and rhythmic diversity than the EDM drum patterns used in experiment 4. This suggests that, conceptually, the space denoted by Gabrielsson’s patterns spans over a wider region than the EDM patterns of our EDM experiment (see Figure 3.13). A further interesting experiment, beyond the scope of this thesis, would consist of adding some EDM patterns to those used in Gabrielsson’s experiment in order to see how far EDM ones are from the rest, and how close between themself they appear, in such a big picture

Our results show how using a reduced set of descriptors, namely [midD, hiD, hiness, lowsync, hisyness], computing Euclidean distance between the descriptor vectors and then using MDS, a perceptual rhythm space composed of EDM patterns is reproduced with significant correlation values. Although this is a significant advancement towards a system capable of automatic drum pattern organization, further experiments must be carried out to evaluate its robustness in larger datasets.

3.7 Conclusions

Our studies on rhythm similarity offer new evidences on the workings of rhythm-related cognitive musical processes, related to rhythm in which pulse, meter syncopation and frequency range play a major role both in monophonic and polyphonic scenarios. The results obtained in the experiments strongly suggest how, in the event of evaluating the similarity between two different monophonic patterns, two different mechanisms are at play. The *syntactic* one, which is focused in comparing the order of the onsets and silences of both monophonic patterns being assessed, and which is used when the patterns are listened to in the absence of pulse induction. A second mechanism, a *semantic* one, is found to be triggered when two monophonic patterns are assessed by similarity in the presence of a pulse context. This finding is backed by experimental results that present how the similarity between two monophonic patterns is judged differently in the presence and absence of a pulse-inducing sequence. Without pulse induction the similarity results correlate with the Edit Distance, an infomation-based string similarity metric based in counting how many characters are needed to transform one sequence of onsets and silences into another. The results obtained in a pulse-induced scenario, on the contrary, show no significant correlation with the Edit Distance, but they do correlate significantly with a metric based on measuring the syncopation in the different patterns. These two opposed correlations suggest an explanation to the influence of the pulse in the judgments: when there is no pulse to be used as a reference to understand the onsets of the patterns, then comparing the onset and silent events of two patterns is a useful mechanism to assess their similarity. However, when a pulse sensation is induced, a layer of cognitive mechanisms is triggered to make sense of a monophonic pattern: the pattern and the pulse are fed back onto each other to deduce a meter and, with it, hierarchies are aroused in the listener's minds which are finally used to weight the expectancy of each note in the pattern measured as syncopation (Longuet-Higgins and Lee, 1984; Palmer and Krumhansl, 1990). This triggered process builds up a cognitive framework which is used in a similarity comparison in pulse-induced scenarios. The ultimate layer in that cognitive framework is syncopation, which appears as a crucial factor for assessing similarity in pulse-induced scenarios. This *semantic* mechanism, whereby monophonic patterns are processed and compared by their rhythmic "meaning", is highly relevant for this thesis as it is concerned with EDM music creation which is rooted in eliciting a pulse sensation.

The importance of note hierarchies in the *semantic* mechanism, which are the source for syncopation sensations, can perhaps be extrapolated to a higher structural level in a pattern. When the subjective similarity judgements in a pulse-entrained scenario were correlated independently for each intra-pulse sub pattern, the first portion of the pattern showed the highest importance towards the reported similarity sensation. That is, the similarity of the first beat of two monophonic patterns is the most important fragment to define the overall similarity in a pattern. This effect, named *awareness* in the sections above, can be an extrapolation of the metrical hierarchies evidenced theoretically by Lerdahl and Jackendoff (1985) and experimentally by Palmer and Krumhansl (1990) in which the expectancy of a note at the beginning of a one-bar musical phrase is higher than the presence of any other note within the one-bar sequence. The awareness, could then be interpreted as the level of attention imposed over the different intra-pulse sections of a monophonic sequence. Although the concept of *awareness* proved to be useful for predicting similarity sensations, further research, which is outside the scope of this thesis, is needed to understand this phenomenon.

As in the assessment of the similarity between monophonic patterns in pulse-entrained scenarios, the concept of syncopation, along with note density and frequency range of the onsets found in a polyphonic rhythm pattern, are fundamental to rhythm similarity prediction. The experiments presented in this chapter show how five descriptors, rooted in the aforementioned concepts, are recurrent for accurately predicting similarity sensations among polyphonic patterns. When predicting classic drum rhythm spaces as those created by Gabrielsson (1973b) and when predicting a contemporary EDM drum rhythm space, created for this thesis, five descriptors [lowsync, midD, hiD, hiness, hisyness] are recurrently found to be responsible for the accuracy of such predictions. These descriptors happen to be in tune with scientific knowledge in the area of polyphonic drum rhythm processing. The five descriptors correspond to three different frequency regions (low, mid and high), representing three different streams of audio, in which drum patterns can be processed by humans auditory scene analysis. The *lowsync* descriptor measures the amount syncopation found in the low frequency region of instruments. The *midD* descriptor measures the amount of onsets performed with sounds belonging to the mid frequency range in a polyphonic drum pattern. The hiD measures the amount of onsets belonging to the high frequency range, the *hiness* descriptor measures the amount of onsets belonging to the high frequency range and the *hisyness* informs about the normalized syncopation in the high frequency range. These effective descriptors are based on simple yet strong principles of rhythm cognition, which makes them straightforward to compute and useful for similarity prediction.

The approach we have found to accurately predict drum rhythm spaces devised by human ratings is based on these use of the five symbolic descriptors, along with the Multidimensional Scaling (MDS) technique. The simplicity of this method from both a conceptual and a computational point of view, suggests an elegant algorithm for automating the creation of polyphonic drum rhythm spaces. With this methodology, which experimentally presents the best fits with human rhythm spaces, drum pattern collections can be processed in order to create 2D maps suited for the exploration and retrieval of drum patterns, which will be addressed in Chapter 4.

4. GENERATIVE TOOLS FOR EDM RHYTHM

4.1 Introduction

This chapter presents original work focused in three main activities of EDM drum production discussed in chapters 1 and 2, namely the comprehension of EDM styles, the generation of variations of EDM drum patterns and the organization of symbolic drum patterns. These three activities are approached with novel tools designed to expand what is currently offered to EDM producers in actual DAWs and plug-in musical software, upgrading these tools with intelligent behavior aiming at a better practice of EDM production.

One fundamental concept for the development of these applications is the fact that acoustic and even virtual (i.e. drum pad based performances) drumming are very physical activities, while sequencing drums is a non-real-time activity that lacks of demanding synchronic motor involvement. The passivity in the analytical process of sequencing drum patterns, which precisely are the musical elements intended to induce movement in music and, most of all, physical motion in listeners, suggests a contradiction between the stillness of the musician, and the traditional physicality of drumming, and the motor sensation induced in the dancers. From this perspective, the use of continuous and gestural control in the process of real-time rhythm sequencing is set as a goal for the software tools researched during this thesis.

We have developed two software applications, which are presented throughout this chapter: *DrDrums*, an intelligent drum machine based around the concept of drumming style and variation, and *RhythmSpace* a drum rhythm space automatic arranger with generative capabilities. Both applications are presented with a detailed description of their inner workings, focusing on the ideas and algorithms that enable their smart functionality and continuous interaction. These two applications are a logical evolution from the ideas presented on chapter 2 and the experimental results discussed on chapter 3.

4.1.1 Fundamentals of DrDrums

DrDrums is a smart drum machine with generative and variation capabilities, aimed at stylistic EDM drum production. DrDrums addresses some of the issues discussed on chapters 1 and 2, which EDM producers face when composing drum rhythms that are then expanded to a dance track:

- EDM is very stylistic, in the sense that rhythmic patterns of different styles have very distinct features which support much of the musical personality of a style. Thus, a producer must be very knowledgeable of the EDM style being produced, specially the inner workings of drum arrangements and the reasons why the elements induce movement. However, production guides and literature are reduced to monolithic examples of drum patterns without explanation of why they work rhythmically in relation to dance. Comprehension and reproducibility of a drumming style is thus a must for an EDM producer.

- Starting up an EDM *drum track* project with an empty page can lead to a scary situation, unless the producer has prefigured a very clear idea of the music she wants to create, or unless she is a highly skilled composer. Having explicit machine-aided musical intuitions about the work to be accomplished, might be an elegant way to avoid the distress of the empty page, or to get a lead on a possible direction to evolve a drum pattern.

- Performing variations on a drum pattern once it has been established as a working seed for a *drum track* is a fundamental activity for an EDM producer. Typically it is performed note-by-note, adding one onset at a time, and some times even as a trial-and-error activity. Here, again, music literature and tutorials fail to address techniques and methodologies useful for transforming polyphonic percussive patterns while preserving the deep identity of the pattern being transformed.

These three issues, which are mostly EDM theory voids, as well as the mentioned need for a dynamic sequencing gestures, are addressed in the design of the DrDrums application. In the following sections of this chapter, the processes used to devise this application are thoroughly explained. Finally, two different subject-based evaluations, one qualitative and another quantitative, are carried out to assess the features of DrDrums, specially the ones related to style replication.

4.1.2 Fundamentals of RhythmSpace

The second application presented in this chapter is RhythmSpace, a graphical tool for organizing, visualizing and retrieving drum patterns based on rhythmic similarity. RhythmSpace is grounded on the research presented on chapter 3 which explores objective metrics for predicting polyphonic rhythm similarity. A major feature of RhythmSpace is its generative functionality, that goes beyond retrieving existing patterns in the collection as it can also generate new ones based on drum interpolation. RhythmSpace addresses directly the need for a file browser with music cognition capabilities that takes care of organizing a large amount of drum files in an intuitive and musical way that musicians can relate to, as it was exposed in section 2.5.

4.1.3 Technologies of DrDrums and Rhythm Space

In order to design and build the DrDrums and RhythmSpace applications, discrete bits of technology had to be developed to solve diverse aspects of drum generation, transformation and organization. The approach taken to address all of these aspects was quite similar, and based on applying research results on rhythm perception and cognition previously presented in Chapters 2 and 3. The specific technological developments discussed throughout this chapter and which constitute the basic elements for DrDrums and RhythmSpace are namely: an algorithm to increase or decrease the onsets of a rhythm pattern maintaining the main onsets, called agnostic density transformer (ADT); a drumming style extractor based on analyzing MIDI drum patterns; a drum pattern generator based on using the extracted drumming styles; a system for automatically organizing polyphonic drum patterns in a 2D space by similarity; and finally, several algorithms for drum pattern interpolation.

In the following sections the concepts behind the agnostic density transformer (ADT) (section 4.2), and those of style (section 4.3) and variation in EDM drum pattens (section 4.4) will be presented. Their implementations will be presented and evaluated within the DrDrums and RhythmSpace applications (sections 4.5, 4.6 and 4.7).

4.2 Agnostic Density Transformer

Making variations of a drum pattern while preserving its identity is an EDM producer key activity, that has been discussed throughout this thesis, specially on Chapters 1 and 2. Given its importance, researching how to progressively transform a pattern in a musically meaningful and controlled manner, is one of the main goals of this work. As discussed on previous chapters, syncopations are a key factor influencing similarity judgments in percussive patterns (for a review see sections 2.6, 3.2 and 3.3), as also are the amount of onsets in a pattern, also known as the density (see Gabrielsson, 1973 discussed in section 3.7). Following the definition of syncopation given by Longuet-Higgins and Lee (1984), presented in Chapter 3, it can be seen that syncopation is independent of the amount of onsets on a pattern. That is, they are two independent variables affecting different properties of a rhythm pattern: syncopation (defining the salience of the notes in a pattern against a pulse) and density (defining the total amount of notes present).

Syncopated notes in a pattern are least metrically expected and, as such, they are the ones that define the salient points of a rhythm. They are definitive for determining its character, in the sense that syncopations are the ones that deviate from the more expected, reinforcing notes. As such, altering the types of syncopations in a percussive pattern is a major change, an idea thoroughly discussed in sections 3.2 and 3.3, where the influence of syncopation in similarity sensations is evaluated. Changing the syncopations between two patterns, makes up for high perceptual differences, altering the character of a pattern; but density can be affected without disturbing the main syncopated, thus salient, notes in a pattern. For an EDM producer, this type of rhythm variation, which affects the fullness of the pattern while maintaining the essential notes of a drum pattern, is a useful composition tool.

We design, then, an agnostic density transformer, which is by definition independent of any idea of style or any external reference besides the induced pulse and meter of a rhythm, thus 'agnostic', is designed. The agnostic density transformer (ADT) is based on the principle of acknowledging the syncopated notes as the most salient events in a percussive pattern. The main concept is to conceive an algorithm that minimizes the cognitive impact of adding or subtracting onsets of a given input pattern.

Based on metrical weight profiles, as the ones proposed by Lerdahl and Jackendoff (1985) or Palmer and Krumhansl (1990) (see Figure 4.1), the salience of notes and silences in a drum pattern can be established. With these salience profiles, syncopation values can be measured for each of the notes present in a pattern, defining a degree in which each note defies the pulse and, as such, is more salient. Additionally, from the experimental results presented in section 3.2, regarding monophonic rhythm similarity, it is concluded that some of the inter-pulse sections within a percussive pattern are more important than others. Namely, in a monophonic one-bar pattern with minimum resolution of 1/16th note, the changes excerpted at the first quarter of a pattern are much more similarity disturbing than changes after the other pulses. That is, if two percussive monophonic patterns differ only in the region within the 1st and 2nd pulse, they can be considered more different than if the dissimilarity was between any of the other pulses. In general, transformations in different regions within a pattern have a different impact in similarity, being the region comprehended between the 1st and 2nd pulse the most important, then the 3rd and 4th, then the region after the 4th pulse, and finally the region between the 2nd and 3rd pulse (see Figure 4.2).

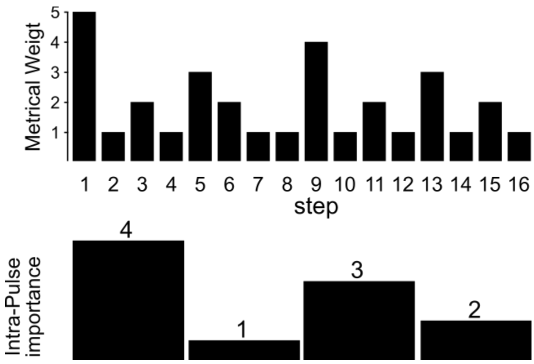

Figure 4.1. Theoretical metrical weights as defined by Lerdahl & Jackendoff (top) and importance of intra-pulse regions as presented in Section 3.3 (bottom).

Combining both the metrical salience levels of the pattern, onsets with the region in which they are located within the pulses, an *importance* value can be assigned to each of the notes of a pattern. This importance value can be used to decide which of the notes can be affected, while minimizing the impact of the transformation; to decide, in other words, where to add or remove a note, where it is less noticeable for a listener.

The idea of the ADT algorithm is to number each of the steps of a pattern with an importance value based, on analyzing the notes and their metrical salience. After the analysis, the density of the pattern can be controlled by assigning onsets to steps which are equal or lower than the user-controlled density value. Therefore, when the desired density is equal to the density of the original pattern, the same pattern will be retrieved; when the desired density is equal to 1, the single most important note of the original pattern will remain (that is the note with the highest syncopation in the most salient region); and when density is set to just one value below the total steps (when all steps but one are filled with onsets) that silence is going to be the one that generates the most important syncopation of the original notes of the pattern. The current implementation of the algorithm is based on one-bar length drum patterns in 4/4 measure, with a minimum time resolution of 1/16th note. The description of the algorithm is as follows:

1. Adjust a drum pattern to the 1/16th note minimum step resolution, one bar length and the type of meter. Use the metrical weight profile by Lerdahl and Jackendoff (1985) to assign the salience to each note and calculate its syncopation value.

2. Segment the events of the monophonic drum pattern between *relevant* and *irrelevant* onsets and *relevant* and *irrelevant* silences. Relevant onsets are those that are preceding a silence. Irrelevant notes are those 1/16th notes which precede another note. Relevant silences are those preceded by a note and irrelevant silences are those preceded by another silence.

3. Assign importance values to the steps, where each of the relevant notes are located, based on their syncopation values and the region awareness profiles (see Figure 4.1). Starting with the note with the highest syncopation value located in the most salient region, and descend by syncopation value and region value.

4. Continue assigning importance values to the steps where the irrelevant notes are located, given the salience of their region.

5. Now continue assigning importance values to the steps where irrelevant silences are located, given the salience of their region.

6. Finish assigning importance values to the steps where the relevant silences are located based on the importance value of the notes preceding them. The relevant silence with the lowest importance value must be the one located right after the relevant note with the lowest importance value.

7. Now that all the steps have been numbered, control the density of the pattern by placing notes in steps which have equal or lower importance value than the desired density.

4.2.1 Agnostic density transformation example

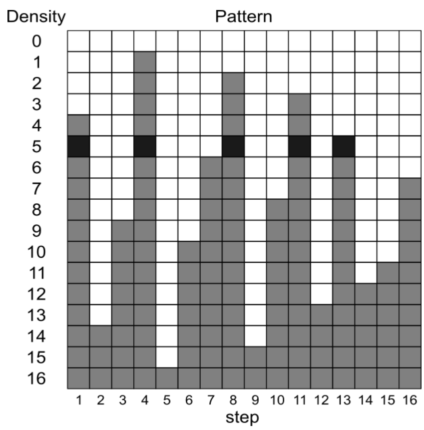

Figure 4.2. Evolution of a Rhumba clave pattern (black) when the ADT algorithm adds and removes onsets[28].

[28] See a demo of the ADT here: http://bit.ly/2EuWJSB

An example of the ADT in action is presented in Figure 4.2. Here we can see how the Rhumba clave pattern, with an original density of 5 onsets, is transformed by adding and subtracting onsets using the ADT algorithm. The onsets with the lowest syncopation value and in the least important pulse are subtracted. Onsets at steps 13, 11, 8, 1 and 4 are removed, leaving the onset with the highest syncopation value in the most important intra-pulse region (region I) . To increase the density, onsets are added to the *irrelevant* silence steps, which are those silent steps that are right before an onset. The first irrelevant silences are filled following an inverse intra-pulse importance order, that is, the least important silences in the least important intra-pulse regions are filled. Onsets are added progressively at steps 7, 16, 10, 3, 6, 15. When density 11 is reached, the only silent steps are relevant-silence steps, that is, silences located after any of the notes of the original pattern. When density is increased after this point, the sensation of the original pattern will be progressively lost. The steps progressively filled are those on the least important intra-pulse region and with the least important syncopation value. The final steps that are filled are 14, 12, 2, 9 and 5.

4.2.2 Considerations of real-time interaction with density

The ADT by itself can be seen as a proof of concept application to confirm seminal ideas such as syncopations being the most important notes of a pattern, and also that some regions of a pattern are more important than others when excerpting changes. Also, the fact that an input pattern can be transformed in real time adds an interactive appeal to its use, specially when used, in polyphonic drum patterns, to change the density of different instruments in real time. In DrDrums this ADT is used as a post processing tool for transforming the densities of the different instruments in the generated rhythms.

One important feature of the ADT is the dynamic real-time addition and subtraction of notes from a monophonic pattern, which opens the door to a new gestural control of rhythm sequencing. The possibility of tweaking a knob and listen to a progressive real time rhythm density transformation, expands the motility involved in drum sequencing, historically reduced to discrete clicks, towards a dynamic and action-based activity. In this sense it closes the gap between the traditional quietness of sequencing and the movement induced by drumming. The ADT offers the possibility for EDM rhythm performers and producers to amplify a gesture into the perceptible dynamics of drumming, as a transducer of the continuous movement of a slider towards the multiple hits of a drum.

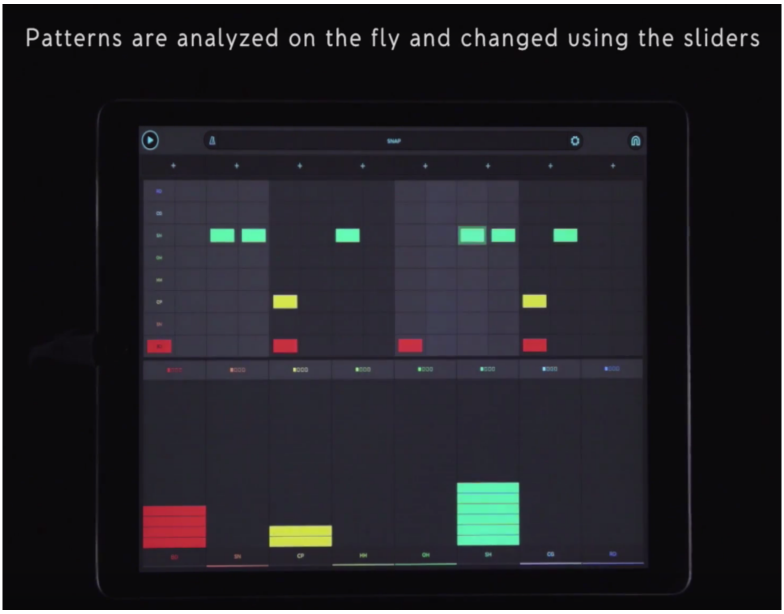

Figure 4.3. Snapshot of the introductory video of the Snap drum machine. The lower panel features vertical sliders used to input the density value of the drum machine's different percussive instruments.

This ADT algorithm has already been successfully[29] implemented commercially in the Snap drum machine developed by the ReacTable company[30], one of the partners in the Giant Steps project that framed the development of this thesis. Snap features a classic Piano Roll Interface as discussed in Chapters 1 and 2 (sections 1.2 and 2.5). In the Snap drum machine the density control is the main feature which makes the difference from any other piano-roll-based drum machine (see Figure 4.3).

4.3 Drumming style extractor

As it was presented in section 2.4, one of the particular aspects of EDM drum sequencing is that it heavily relies on style as the focus of its activity, meaning that producers recreate specific features of an established style, as the foundation of an original dance track. Inspired by the definitions of style presented by Meyer (1967), suggesting that a style is an impression of the mind after being exposed to a particular set of examples which have common features, a line of original work was devoted here to understanding how to deal with EDM drumming styles, specially aiming at reproducing a drumming style with a generative tool. The material to have as reference in order to define a drumming style is a collection of drum patterns in symbolic format which are expected to have some common features. These reference patterns can either belong to a same EDM style, a single track or be produced by a specific musician.

After reviewing the work by Ames (1989), Eigenfeldt & Pasquier (2013), Conklin and Witten (1995), Pearce & Wiggins (2004), Pearce, Conklin Wiggins (2004) and Pachet (2002, 2003, 2006) in which probability distributions and stochastic processes (such as the Markov process) are used in musical generative imitation tasks, and contrasting these works with the genetic algorithm methods presented by Burton (1998) and Bernardes et

[29] Reviews of the Snap drum machine can be found here:
http://www.thethreeofive.com/reacTables-snap-is-a-Tablet-drum-machine-for-quick-pattern-improvisation/
[30] http://www.reacTable.com/snap

al. (2010), Chapter 2 concludes with a decision to explore statistical-analysis-based systems, for studying EDM drumming with generative aims. The general goal is to extract a formal entity, the *style*, from a set of symbolic drum pattern examples. This entity must allow to apply the features present in studied patterns into new generated patterns. In simple terms, to extract a stylistic structure from a collection of patterns which can be used to endlessly generate new similar patterns.

Given the nature of the analysis used for extracting the styles, the resulting style entity is a probabilistic space. This type of stylistic representation is ideal, as it can be processed mathematically in profitable ways, for generative purposes. In the following sections the extraction of styles from symbolic patterns will be presented and, once a style is extracted, some transformations are going to be explored. Specially, we will focus in the possibility to combine two different styles to obtain a hybrid one, and also in the transformation of a style obtaining its most common features.

4.3.1 Pre-processing the patterns

The first stage for extracting a style from a group of symbolic drum patterns consists of a homogenization phase, where all patterns are cut to a defined length, quantised to a common minimum time resolution, and all its elements are mapped to a specific set of instruments. This preprocessing stage is done for narrowing down the features of the patterns used, as it is simpler to work with them for research purposes. However, the methodology presented throughout this chapter can be expanded and modified to work with different pattern lengths, time resolutions, meters (for example ternary) and number of instruments, without affecting the applicability of the concepts here presented.

This first stage consists of selecting the patterns in symbolic format, MIDI for simplicity and universality purposes, and to cut them making sure they have a fixed size of one bar. Then, the time resolution is adjusted to 1/16th note duration (a semiquaver) with a meter of 4/4, obtaining 4 1/16th notes in between each pulse and getting a total of 4 beats in each pattern. If any of the patterns analyzed are longer than one bar, or even if they are a complete drum track of an EDM song, they are adjusted to this format. Hence, it is ensured that all patterns can be described as a matrix with 16 columns, each column representing the minimum time resolution of 1/16th notes. The rows of the matrix are the different instruments present in the patterns, and the onset of a given instrument at a given step is represented by a 1 whereas a silence at a given step is represented by a 0.

After the patterns and its notes are time-formated to fit the 16 columns of a matrix, the rows are to be processed. Each MIDI note (row) present in the file being preprocessed, represents a percussive instrument following the General MIDI Level 1 Percussion Key Map[31] (GMPM). This map is a standard for assigning a certain type of sound to a MIDI note so that when the MIDI pattern is reproduced, an audio engine can render each note with an appropriate sound. Again, for the simplicity of the research in this thesis, the number of MIDI instruments/rows is reduced to eight using a specific customized mapping which allows to convert from the 46 percussive instruments available GMPM to 8 instruments, but this principle could be extended to more instruments. The eight instruments used in the style extractor are kick drum, snare, closed hi-hat, open hi-hat, rim shot, clap, low and high congas. These instruments represent a consensus between the main elements of an acoustic drum kit (kick, snare and hi-hat), with some elements constantly present in EDM such as as claps, rimshot and congas. The final representation of a drum pattern is a matrix of 8 rows and 16 columns filled with ones and zeroes (see Figure 4.4).

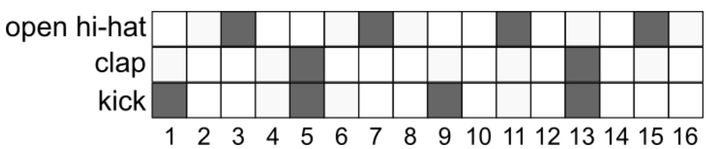

Figure 4.4. A piano roll representation of a typical House pattern.

31 https://www.midi.org/specifications/item/gm-level-1-sound-set

4.3.2 Analyzing the patterns

Once the patterns have been preprocessed, they are ready for analysis. The first stage for analyzing the patterns is to determine the vertical coincidence of events at every step and assign a single value to it (see Table 4.1). This step is taken in order to obtain a condensed representation of a polyphonic pattern rather than having scattered monophonic sequences. Given that there are 8 instruments which can have an onset or a silence at every step, there are $2^8 = 256$ possible combinations of instruments at every step. These combinations go from silence (all instruments are zero) to a full hit of the eight instruments (all instruments are 1). Now, any possible combination of patterns at a given step is turned into a number, representing a vertical combination of events. With this process, every pattern is converted to a sequence of 16 events. In order to assign the numeric representation to the vertical combinations of instruments' onsets and silences, a binary to decimal conversion is applied (see Table 4.1). It is important to note that the value that represents the vertical coincidence of onsets has nothing to do with any property of the onsets and it is only used for nomenclature purposes.

instrument	Onset Combination						
k	0	1	0	1	0		1
s	0	0	1	1	0		1
ch	0	0	0	0	1		1
oh	0	0	0	0	0		1
cp	0	0	0	0	0	...	1
rs	0	0	0	0	0		1
lc	0	0	0	0	0		1
hc	0	0	0	0	0		1
code	0	1	2	3	4	...	255
Event	silence	k	s	k, sn	ch		all

Table 4.1. Example of vertical pattern combinations and the numerical representation used.

Having a set of patterns of a given style converted to single-event sequences, these are batch processed by step, defining the chains of events of different orders (or lengths) that precede an event at every step. For example, if an order O = 2 analysis is being done for a given pattern P_1 at step S, where there is an occurring event e_{p1s}, then the previous O events are grouped and assigned as precedents saying that e_{p1s-2}, e_{p1s-1} precede e_{p1s}. This is performed for every pattern collection at each of its steps accumulating all events e_{pNs} with their precedents e_{pNs-2}, e_{pNs-1}. Same precedents are grouped and their following events counted. If at step S the precedent (e_{p1s-2}, e_{p1s-1}) is common in patterns p1, p2 and p3, all three following steps e_{p1s}, e_{p2s} *and* e_{p3s} are grouped and their repetitions counted. For example, if following steps for precedent (e_{p1s-2}, e_{p1s-1}) in patterns p1 and p2 are the same ($e_{p1s} = e_{p2s}$), then the count of precedents for (e_{p1s-2}, e_{p1s-1}) is [2 e_{p1s} ,1(e_{p3s})]. The counts for each of the events are finally normalized, ending with a general result for a given step S and a precedent (e_{p1s-2}, e_{p1s-1}) to be [0.66 e_{p1s} , 0.33(e_{p3s})]. This means that given a precedent (e_{p1s-2}, e_{p1s-1}) there is a probability of 0.66 that the next events are e_{p1s}, and a probability of 0.33 that the next events are (e_{p3s}) (see Figure 4.5).

Analysis of order 2 and step 4

Patterns

step	1	2	3	4	5	6	7	...	16
pattern A	*1*	*8*	*0*	*1*	*4*	*0*	*5*	...	*2*
pattern B	*4*	*0*	*0*	*3*	*4*	*3*	*9*	...	*6*
pattern C	*1*	*7*	*1*	*1*	*4*	*2*	*6*	...	*4*
pattern D	*1*	*8*	*0*	*3*	*0*	*0*	*6*	...	*4*
pattern E	*7*	*8*	*0*	*1*	*9*	*0*	*5*	...	*4*

Resulting matrix

	probability of events at step 4				
previous events	*1*	*2*	*3*	...	*256*
0 0	*0*	*0*	*1*		*0*
...					
7 1	*1*	*0*	*0*		*0*
...					
8 0	*0.66*	*0*	*0.33*		*0*

Figure 4.5. Example of generating a matrix [2, 4] (order, step) based on five different patterns A to E.

As every pattern in a collection is step-analyzed with a given order O, the result is a matrix with the possible states present at every step grouped with the probability of each precedent events. The matrix for step *S* and order *O* is in the form:

precedent events	**event *e1***	**event *e2***	**…**	***event en***
$e1_{S-1}, e1_{S-2}, \ldots, e1_{S-O}$	$p(e1_S \mid e1_{S-1}, e1_{S-2}, \ldots, e1_{S-O})$	$p(e2_S \mid e1_{S-1}, e1_{S-2}, \ldots, e1_{S-O})$	…	$p(en_S \mid e1_{S-1}, e1_{S-2}, \ldots, e1_{S-O})$
$e1_{S-1}, e1_{S-2}, \ldots, e2_{S-O}$	$p(e1_S \mid e1_{S-1}, e1_{S-2}, \ldots, e2_{S-O})$	$p(e2_S \mid e1_{S-1}, e1_{S-2}, \ldots, e2_{S-O})$	…	$p(en_S \mid e1_{S-1}, e1_{S-2}, \ldots, e2_{S-O})$
…	…	…	…	…
$en_{S-1}, en_{S-2}, \ldots, en_{S-O}$	$p(e1_S \mid e_{S-1}, e_{S-2}, \ldots, e_{S-O})$	$p(e2_S \mid e1_{S-1}, e1_{S-2}, \ldots, e2_{S-O})$	…	$p(en_S \mid e1_{S-1}, e1_{S-2}, \ldots, e2_{S-O})$

Where all possible combinations of precedent list of events $[(e1_{S-1}, e1_{S-2}, \ldots, e1_{S-O}), (en_{S-1}, en_{S-2}, \ldots, en_{S-O})]$ are used as rows, and all possible future events ($e1, e2, \ldots, en$) are considered as columns. The cells carry probability values of event *en* happening after a given list of precedent events.

Each order has a different size of matrix for all its steps, as order 1 has

256 possible events preceded by 256 possible events at each step, configuring a probability space composed of sixteen (one per step) 256 x 256 matrices. But for order 2 there are 256 possible events with $256^2 = 65.536$ possible preceding states, which dramatically increases the size of the matrix for each step. In general, the sizes of the *i* dimension of the matrix (the possible preceding sates) increases exponentially with the order, following the rule $(possible\ states)^{order}$. As the possible states are fixed for the analysis at 256 possible combinations, then the size of the *j* dimension of the matrix is $(number\ of\ instruments)^2$. Generalizing, for any collection of patterns, the precedence analysis generates a number of matrices equal to the maximum used order, times the number of steps in the analysis. The size of the matrix generated at each order and step is $(((Number\ of\ instruments)^2)^{order}, (Number\ of\ instruments)^2)$. This complete set of matrices then becomes the representation of a drumming style, which can be used in a further step to generate new patterns in that specific style, and also to make adjustments in real time to the style (such as emphasizing certain probabilities or combining it with other styles). All these processes are presented in Section 4.4.

The purpose of each matrix is thus to provide a set of probabilities to select a candidate event at a specific step, given an order and certain precedent events, just as in a Markov Process. The complete set of step matrices can be used to generate a sequence of events at every step that when concatenated, constructs a new drumming pattern.

For the purpose of this thesis, a python script was developed to process collections of MIDI patterns and generate every [*O S*] matrices needed. This script is included in the more general tool called *RhythmToolBox*, which is a python library of scripts to process drum patterns in every way that is presented throughout this thesis[32]. In the case of style generation, a single function called *makestyle* is used to generate new patterns based on those in a collection.

[32] www.github.com/danielgomezmarin/rhythmtoolbox

4.3.3 Generalizing stylistic music knowledge extraction

In the previous section, the concept of a "style structure" is introduced, based on analyzing sets of MIDI files. Here, a more general view, which can be used as the basis for generating the style knowledge for any type of symbolic musical information, is presented. This section developes one of the outcomes of supervising a Master's student in his attempt to create an intelligent bass line generator for EDM (Calopa, 2016). The melodic analysis required in this Master's project led to the construction of a general methodology, such that it could be used in the analysis process of general symbolic musical pieces.

An addition to the methodology presented in the previous section, is the possibility to extract different parallel information in order to achieve a more detailed insight into the characteristics of a musical sequence. Diverse musical descriptors, for example the coincident instruments played, the notes, their duration, or the vertical density of simultaneous polyphonic events, can be taken into account in this general model. In this model, every new possible value for each descriptor at a precise moment in time can be stored. This allows, in a further generative stage, to come up with new musical events resulting from the combination of different descriptors. This multi layer system allows the generation of new material based on the inner relations of different aspects of a musical piece (i.e. combining previous information of rhythm, pitch and chord to select the pitch and duration new note).

In general (see Figure 4.6), the layers of different descriptors (*D1, D2 … Dn*), are composed of a timeframe *TF* (which is the period of time in which these descriptors are to be measured), a step *S* of that timeframe, an algorithm on how to compute each descriptor *D* (e.g. "density = sum the amount of instruments in this timeframe"), the order *O* (which determines the length of the influence of past events, in the selection of a new event), and finally a series of matrices where all that data is stored.

Each MIDI file is decomposed into a large set of matrices of different orders, which are computed at a given step for each descriptor. The important feature about this method for extracting and storing data is the possibility to aggregate several corresponding matrices of different files (i.e. same order, step, timeframe and descriptor, coming from different files), thus creating matrices with more complex and dense data, while keeping them normalized and ready to be used by the generative algorithms.

Our concept of style could thus be further defined as "a collection of matrices, appropriately indexed by order, step, timeframe and descriptor, which have been created by adding information at every timeframe, for all files in a collection". So far, this definition of *style* is very inclusive and makes no distinction of the musical symbolic files which are analyzed. Therefore, what stylistically defines a style is not questioned or filtered by the system, but is more the result of the mashup of every matrix generated by the input data. This flexible and useful definition of style allows to create combinations of very different information sources (albeit all MIDI for the moment), which can be useful for further generative stages.

This method generalizes to process parallel sequences of a different number of descriptors. It can also be used to process music sequences with different timeframes in order to achieve different temporal resolutions. One aspect that has to be considered when using this method of analysis is the fact that Markov sequences are not a strong method for processing structural aspects of musical sequences in the long run (Pachet, 2011).

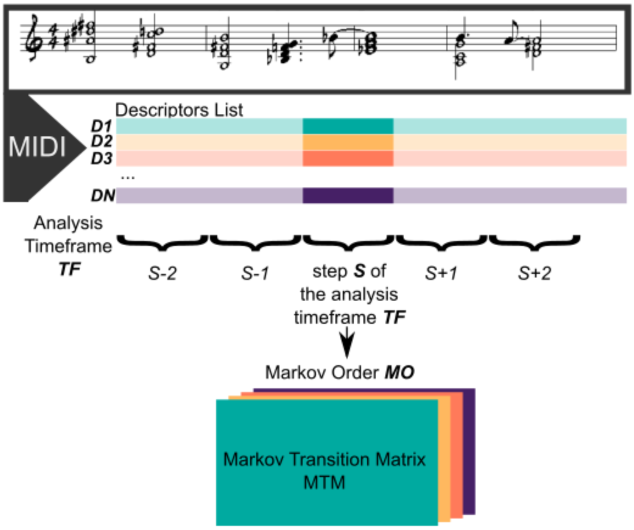

Figure 4.6. Schematic representation of how a MIDI file is analysed with different descriptors (*D1, D2, DN*) at each given time-frame *TF*. Then at a step *S* of *TF* a *MTM* is computed for each descriptor and for each Markov order *MO*.

4.4 Drum Pattern Generator

This section presents how to take advantage of a style representation, extracted using the methodologies presented above, to generate new drum patterns in specific styles. This abstract entity of *drumming style* will be processed and manipulated to allow EDM drum production overcome the reduced style definitions found in literature (Brown and Griese, 2000; Emmerson, 1988; Hewitt, 2009; Snoman, 2012; Adamo, 2010) and also to relieve the lack of compositional tools available for EDM *drum track* composers (see sections 2.4 and 3.5).

4.4.1 Using style information to generate a Pattern

Once we are able to transform a collection of drum patterns into statistical knowledge using the style extractor, the possibility to generate new patterns based on that knowledge becomes feasible. When traveling through the probabilistic space of the style, new patterns emerge as possible statistical combinations existing in the space. The surface of the probabilistic space determines the path that any pattern of the learned style can take, as well as regions the patterns never cross. This translates in the probability that some combinations of drum hits (events) can be played at a given step and finally be used in the generation of a new pattern.

In order to generate a new pattern based on the *drumming style* extracted, some specific steps are followed:

1. Define an order O in which the generation is going to be carried out and the length L of the patterns to be generated. The length L must be equal or smaller than the length used in the drumming style extraction phase.

2. Create a random value list RV, containing $L+1$ random numbers in the range from 0 to 1.

3. Select the matrix [$S1$ O] for order O and step 1, using the matrix structure of the style.

4. Create a list with every row (list of precedent drum events) which has at least one nonzero probability of preceding any drum event, using the selected matrix [$S1$ O]. This list represents all the possible preceding lists of events that anticipate a drum event at step 1.

5. Select one element from the list of possible precedent drum events created in the previous step, using the last element of the list of random numbers RV_{L+1}. This means to select one row which has at least one nonzero probability.

6. Use the row selected on the previous step and normalize all probability values to obtain a probability distribution. Use the first element of the random value list RV_1 and the normalized probability distribution to select one event en_1 that will be the first event of the new drum pattern.

7. Create a new list of precedent events, once en_1 and its list of precedent events is obtained. Remove from the list of precedent events the oldest event $e_{(S\text{-}O)}$ and add at the beginning the event en_1 (the generated for step 1) to obtain a new precedent list of events.

8. Use this new list as precedent in the matrix for order *O* in step 2 to select a row which contains future events and their probabilities.

9. Normalize the probabilities of all the nonzero events in the selected row to obtain a probability distribution. Use the second element of the list of random values RV_2 and the normalized probability distribution to select one event en_2 that will be the second event of the new drum pattern, using the discrete probability density function (DPDF) (see Appendix B).

10. Once the event for step 2 is selected, repeat steps 7, 8 and 9 to select events for the subsequent steps.

11. Output the list of events $e_1, e_2, e_3, \ldots e_L$ as a new drum pattern.

Following these previous steps, given an order *O* and a pattern length *L*, any extracted style knowledge can be used to generate new patterns following the rules of precedence embedded in the analyzed patterns in a very straightforward way. The next sub-sections present some additional refinements, applied to this process in order to develop some stylistic concepts for the generation.

4.4.2 Style interaction concepts

At the heart of this generative system lays a process of selecting possible drumming events mediated by the past events, a probability distribution, and a random value. As will be presented below, high-level musical concepts as stylistic combinations of learned styles, commonness/oddness of the drum patterns generated, and pattern variations, are implemented and controlled by modifying this stochastic event selection process. The first concept, *style combination*, comprises a series of manipulations to extracted styles that lead to the generation of new hybrid patterns, based on previously nonexistent styles which can be created in real time. The second concept, *commonness* is based on the studies by Andersen & Knees (2016) who suggest the idea of steering a generative musical process into familiar or unfamiliar results, favoring the convergence towards the most prototypical patterns of a style or, on the contrary, the emergence of surprising results. The last concept, patten *variation*, takes advantage from the manipulation of the random values used to select an event from a probability distribution.

4.4.2.1 Style combinations

As presented on section 4.4.1, once a style is extracted and a generation process is carried out, following the procedures described, the probability of any event en_{SO} to occur (at a given step S and using an order O and a given a list of precedent events) is represented by a probability distribution. This probability distribution is equivalent to a row in the step-order [S O] matrix. If two different drum pattern collections are previously analyzed and a style knowledge is derived from each of them, at every step S and at a given order O, each style has its own probability distributions that determine the probability of every possible event en_{SO} to occur after a list of precedent events. In a given step S, the two probability distributions can be interpolated to obtain a resulting one that combines both styles. The interpolation is a weighted sum:

$$P(interpolation) = [P(style\ A) \times i] + [P(style\ B) \times (1 - i)]$$

Where i is the interpolation value between 0 and 1. *P(style A)* and *P(styleB)* are normalized probability distributions where each possible event is associated with a probability. Each probability in *P(style A)* is multiplied by the interpolation value i and added with the corresponding probability of *P(styleB)* multiplied by *1-i*. This effect can be seen in Figure 4.7. Where, at a given step, the two styles are mixed 50% each.

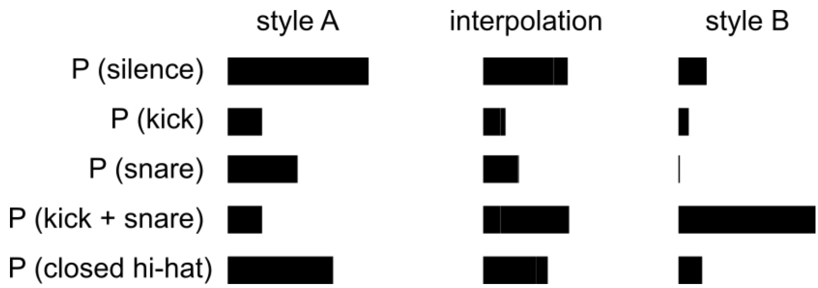

Figure 4.7. Example of an interpolation between style A and style B, at 50% each. The bars represent the probability of each event occurring at a given step.

The practical use of this system is to combine styles dynamically at every step, so that the pattern generation process is based on a precise mix of probability distributions. If the interpolation value *i* is kept constant throughout the generation of a whole pattern, the probability distributions at every step will be conformed of constant proportions of style A and style B. This, in turn, ensures that the whole generated pattern responds to a specific combination of styles A and B. The generation of patterns that resemble two styles in a specific amount opens the door to new and exciting possibilities in drum production practice. One practical example can be that a producer loads his own style and combines it with an extrinsic style creating new patterns by fusing both styles interactively. Conceptually, this process can be seen as amplifying the creative possibilities of an EDM composer. The system allows to extend a musicians' compositional practice, evidenced in his own style, through the addition of another style.

4.4.2.2 Commonness

Commonness is a concept based on the idea of style, as the repeated appearance of similar features within a collection of objects defined by many different features (Gärdenfors, 2004; Meyer, 1956). Based on this idea, if a feature at a certain value is consistent throughout a collection of patterns of the same style, it can be said to be relevant for the appraisal of such a style. This behavior was observed experimentally in Chapter 3 section 6.2, where drum patterns of different musical styles have very similar values in both axes of the EDM space, delimiting a particular region for Techno, House and Breakbeat. In that example, the values of the two axes are within certain boundaries for each style present in the collection. In this sense, if a list of objects in a collection is analyzed computing different meaningful features, the values that are most repeatedly observed in each feature are regarded as the most common. Following this idea, the most common feature value of a collection is the one that is present in most of the objects. Commonness is then related to the number of appearances of a feature value in a collection of objects.

Extrapolating this idea to the process of selecting drum events at a given step (see section 4.4), once the preceding events are defined (thus a row from the probability matrix is selected), the probability of each possible event en_{SO} to occur is directly related to the number of times event en_{SO} occurred after the preceding events in the analyzed patterns. That means that the higher the probability of event en_{SO}, the more common it is within the style. This idea opens the door to shaping the probability distributions of a given list of preceding events (a row in a probability matrix) in order to strengthen (or weaken) the probabilities of the most probable values and to weaken (or strengthen) the least ones. If the most common events are strengthened at the expense of the least common, it favors *commonness*. On the contrary, to weaken the most common events at the expense of increasing the least common in order favors *uncommonness*.

Stylistically, commonness could be translated as reproducing the most recurrent elements of the drumming found in the collection of patterns. On the other hand, uncommonness is the opposite: searching for the least recurrent drumming events found in the style, somehow as exploring the event combinations that were rarely seen in the style. In practical terms, these operations enable interesting explorations of style for pattern generation, either by sharpening a style or by blurring it.

Two algorithms of "commonness" are implemented. Sigmoid commonness is based on a sigmoid transfer function, where probability distributions are ordered by their probability values and multiplied by a sigmoid transfer function. The second algorithm, Power commonness, is based on a power function applied to all probability values of a distribution.

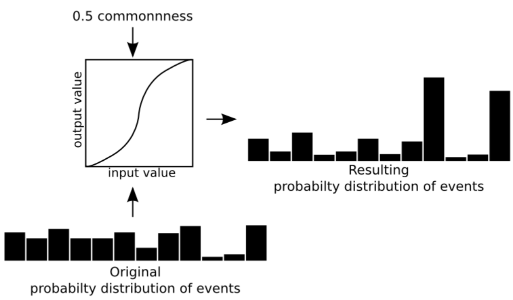

Figure 4.8. Algorithm for sigmoid commonness. The original probability distribution is input (below) processed by the sigmoid function which generates a new probability distribution (right). The sigmoid function is transformed by an input commonness value (top).

Sigmoid commonness

The Sigmoid ”commonness/uncommonness” control is implemented using a sigmoid transfer function to alter the strengths of the probability values at a given step (see Figure 4.8). First, the probability distribution of a given step is decreasingly sorted according to their probability values. Subsequently the reordered Table is multiplied by the sigmoid function, whose skewness and slope are controlled by the user. This algorithm allows to sharpen or flatten these probabilities (via the skewness control), but also to invert them (with the slope control).

Power commonness

Power commonness is based on a power increase algorithm where all probability values are raised to the power of commonness (see Figure 4.9). When commonness is positive the highest probabilities have a bigger increase than he lowest (see Figure 4.9, right). When commonness is negative, the probability distribution is inverted thus making the least probable events to gradually become more probable (see Figure 4.9 left). The uncommonness region has a particular behavior as the probabilities decrease progressively until commonness is 0 (see Figure 4.9 left), where all probabilities are equalized, making all events that had nonzero probabilities to be of the same magnitude. When commonness is 0 the tendency of the generation system to steer towards any specific region is lost, as all probable paths are equally probable. When the commonness reaches a point below 0, inverted commonness starts to occur as probability values are inverted, thus making less probable values higher and high probable values smaller.

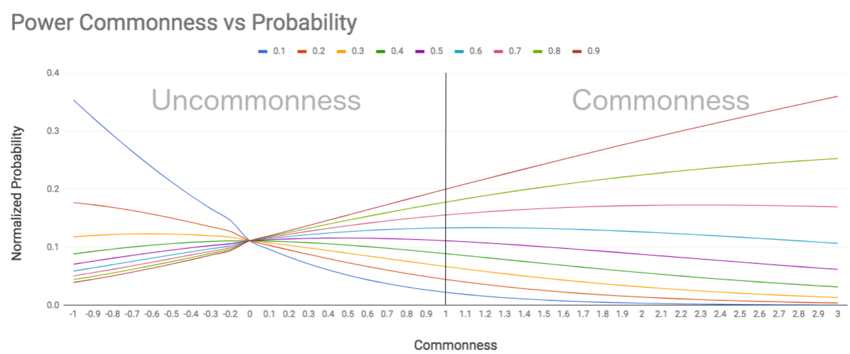

Figure 4.9. Effect of Power Commonness (x axis) in the probability value (y axis) for different original probability values.

4.4.2.3 Inducing variation

The idea of rhythm pattern variation has been presented throughout this thesis as one of the fundamental needs of an EDM producer yet it is one of the most complicated from a musical point of view. Concepts or instructions to create EDM rhythm variations, based on applied music theory, are scarce and insufficient (as discussed in sections 1.2 and 2.5), leaving a lot of room for idiosyncrasy and trial-and-error experimentation.

The analysis and generation system presented in this chapter is based on an event-selection process mediated by a discrete probability density function (DPDF). This function takes as input a probability distribution and a random number, and outputs a single event. When this process is repeated at every step of a sequence, the concatenation of selected events composes a new drum pattern. In this process there are only three elements at play, the probability distribution, the random number and the pattern generated. Each of these elements can be manipulated to introduce variation to the drum patterns:

- Post-processing the polyphonic pattern generated. This idea has been described in section 4.2 where an Agnostic Density Transformer (ADT) is presented as a tool that can add or subtract onsets from the monophonic patterns. If an ADT is added to each of the monophonic patterns that compose a polyphonic pattern, then a powerful transformation system can be achieved.

- Transforming the probability distributions. As it was presented in sections (4.4.2.1 and 4.4.2.2), by modifying the matrices resulting from the analysis of a collection of patterns, the generation process might be steered to specific probabilistic regions, emphasizing diverse aspects of the style, as commonness/oddness and stylistic combinations.

- Transforming the random number input to the DPDF. If the probability distribution is fixed and the DPDF is used to select an event from it, the change in the used random value affects the output event. This process will be explored throughout this section as a source of variation.

First, lets recount how a change in a given RV_S value (the random value at step S) affects the selection of the event en_S if there is more than one nonzero value in the probability distribution, that is, if there is more than one event *en* to be selected at a given step. As an example, using the DPDF with the probability distribution presented above, and different random values, the output is as follows:

event	*e1*	*e2*	*e3*	*e4*	...	*eN-1*	*eN*	
probability	0.2	0	0.1	0.5		0.2	0	$\Sigma=1$

e1 is output if the random value is between 0 and 0.2
e3 is output if the random value is between 0.2 and 0.3 (0.2+0.1)
e4 is output if the random value is between 0.3 and 0.8 (0.2+0.2+0.5)
eN-1 is output if the random value is between 0.8 and 1
From this example two conclusions can be drawn, given that the probability distribution has more than one nonzero value. First, all the different events *en* in a probability distribution can be output from the DPDF as the random value RV_S changes progressively from 0 to 1. Second, although a progressive change in RV_S implies a change in event en_S, the magnitude of the change in RV_S is not related to the change in *en*. So changing RV_S can be a source of variation at the step level but it does not have a linear relation with ΔRV_S.

Zooming out, from a pattern generation perspective, if the *RV* list is altered, then a change in the output pattern can be expected. The only condition needed being that the probability distributions should have more than one nonzero value. This perspective makes possible pattern variations as a consequence for inducing variation to a pattern as a consequence of altering the complete *RV* list but, as mentioned above, this mechanism has a side-effect, a degree of unpredictability.

Lets use an example. In Figure 4.10 four probability distributions are presented, each of them containing three different events *e1*, *e2* and *e3*.

	Step 1 e1	Step 1 e2	Step 1 e3		Step 2 e1	Step 2 e2	Step 2 e3		Step 3 e1	Step 3 e2	Step 3 e3		Step 4 e1	Step 4 e2	Step 4 e3
e1	0	0	0	e1	1	0	0	e1	0	0.1	0.9	e1	0	0	1
e2	0.8	0.1	0.1	e2	0	0.5	0.5	e2	1	0	0	e2	0.2	0.5	0.3
e3	0.6	0	0.4	e3	1	0	0	e3	0.3	0	0.7	e3	0.6	0	0.3

	Step 1	Step 2	Step 3	Step 4	
RV	0.2	0.1	0.3	0.5	0.7
patt	e1	e1	e3	e1	e3
RV'	1	0.9	0	0.3	0.6
patt'	e3	e1	e2	e2	e3

Figure 4.10. Example of matrices, *RV* list and output pattern.

In the leftmost matrix, corresponding to step 1, the nonzero rows are those preceded by *e2* and *e3*. Following the methodology presented in 4.3, the last value of the *RV* list (RV_N) is used to select the past of the first step. As RV_N is 0.7, which is equal or greater than 0.5, the past is *e3* and the probability distribution is [*e1* 0.6, *e3* 0.4]. RV_1 is 0.2 so the step selected from matrix [1 1] is *e1* (highlighted in blue). As RV'_1 is 1 then the output event in step 1 by the DPDF is *e3* (highlighted in pink).

In matrix [1 2] (order 1 step 2) as the previous state given by RV_1 is e2, the probability distribution is [*e1* 1]. Being e1 the only possible output despite the value of RV_2. In the case of using RV'_2 the probability distribution is also [*e1* 1], being e1 the only possible output despite the value of RV_2. In step 2, despite the change from RV to RV', the output event is the same for both random lists.

In matrix [1 3] the previous event given RV and RV' is *e1*, so the probability distribution is [*e2* 0.1, *e3* 0.9]. RV_3 is 0.3 and RV'_3 is 0.0 so the outputs are *e3* and *e2* respectively.

From this example, two conclusions can be drawn. First, the propagation of variation along a generated sequence of events is not completely defined by a change in the *RV* lists. In the previous example, although *RV* and *RV'* are different, there is a convergence to the same event in step 2. Second, the magnitude of the change is not linearly related with the variation of the pattern. As seen in step 2, a big change of value from $RV_2 = 0.1$ to $RV'_2 = 0.9$ does not imply a change in the output. In step *e3*, on the other side, although the change between RV_3 and RV'_3 is small (from 0.2 to 0 in a unitary scale), the output step is variated.

In order to predict the effect of changing the *RV* list in the output sequence, the relation between the RV_S value and the probability distribution must be analyzed. There is an interaction between the change in the RV_S value, and the probability distribution mediated by the number of probable events to occur. That is, the higher the amount of events, the higher the probability to induce change given a variation in the RV_S value; and also, as the DPDF is a cumulative probability function, the higher the probability value of certain event, the least probable it is to induce variations given a change in RV_S. Finally, the size of the change in the RVS value and the number of events is also inversely related. The more events in the probability distribution the smaller the change in the RV_S value needed to induce a variation. A proposed equation to predict the probability of change at a given step can then be written in terms of the probability distribution, the change in *RVs*, and the number of events in the distribution.

$$P(change) = flatness\ (PD) * \Delta RVs/E$$

Where the *flatness* of the probability distribution is the ratio between the geometric mean to the arithmetic mean of the probability distribution, also known as the Wiener entropy. *E* is the sum of the probable events within the probability distribution and ΔRVs is the change in the random value.

Interpreting the flatness value is intuitive, as it represents the evenness of the probability values of all the events. The more even the probabilities, the more probable it is to move to another event given a change in RV_S. The probability of inducing a variation given a change in the RV_S value can be step by step-by-step measured on the probability distributions.

The variation-induction system presented in this section relies only on the possible drum events occurring within the style at a each step. The system works by transforming the RV list progressively so that the RV_S value at each step is transformed in a magnitude relative to the amount of transformation. This, eventually, causes the event selected by the RV_S value and the DPDF function (see Appendix B) to change as explained above.

To implement this algorithm, a simple approach is taken. The RV list is treated as a vector and is progressively shifted in magnitude using a value proportional to a *variation* amount, and then wrapped between 0 and 1. A continuous wrapped sweep of the RV outputs all possible values of the random value at every step S, forcing the DPDF function to output each of its probable values at every step. This algorithm is implemented and used as part of the *DrDrums* system as it will be presented later.

4.4.2.4 Inducing variation by timbre grouping

Another method for inducing variation in a drum sequence, based on the system presented in this Chapter is explained theoretically, although it is not implemented. As presented in section 4.3.2, there are 8 possible single drum events used in this model (kick drum, snare, clap, rimshot, closed hi-hat, open hi-hat, low conga, hi-conga), plus all the combinations among them for a total of 256 possible event combinations (including silence). These 256 events are indexed with a number which doesn't follow any perceptual order or represent any magnitude. The number is just a binary to decimal conversion where the binary numbers indicate which of the 8 instruments are played. Given this agnostic order, replacing event e_n for e_{n+1} does not imply a smaller change (in any perceptual dimension) than changing from e_n to e_{n+10}. In this sub-section new possible ways for inducing variation are theoretically developed, but not implemented.

From an acoustic perspective, and regardless of the style, each of the 256 possible events e_n are made up from drum onset combinations which can be analyzed given their acoustic properties, specially timbre and frequency range. An analysis using these perceptual dimensions can establish similarity relations between the 265 different events which could be exploited in a variation algorithm. For example, the combo event [kick, closed hi-hat] could be related in similarity to a single kick or a single closed hi-hat, as both instruments are present in the combo. Even more, the kick drum (a typical low frequency instrument) could be replaced by a low conga (another low frequency instrument) creating a similarity relation with the kick drum and closed hi-hat combo. In this way, similarity relations between the 256 different events could be established, defining the closeness of one acoustic event with the remaining 255 events, at least at an instrument-label-based level .

Once established, these label-based similarities could be used to replace events of a pattern in a controlled manner, for example replacing one event, or a group of them, with the most (or least) similar ones.

4.4.3 The order effect

One fundamental characteristic of using probabilistic spaces with generative aims is that the diversity of events that can occur at a any step is inversely proportional to the order used to generate the sequence. So, as the order goes higher, the generation becomes more deterministic, up to a point in which generated patterns become just an exact reproduction of the ones that have been learned.

All the concepts and transformations described so far in this chapter are applicable regardless the Markov order used to analyze the dataset. Theoretically, with any sequence of events that exist within the Markov analysis, at least one event is followed. However, as the size of the Markov analysis increases, the restrictions for the next event to occur become more strict, leaving fewer choices of events to select in every step. This means that the higher the Markov order, the least number of different events for the next event are, thus eliminating the possible branches that could be the source for variation. At high Markov orders, the complete rhythms of the database start to be recalled, thus reducing the originality of the rhythms generated by the system.

4.4.4 Zero frequency states

A typical problem with orders higher than zero, and when the probabilities are manipulated as explained above, is the breakdown of the generative system due to the unavailability of candidates. This is called a Zero Frequency State (ZFS) which means that, at a given step, given a list of preceding events (of either length or order equal or higher than one) there are no possible events to follow. The generative system encounters a state that does not exist in the database and it loses the ability to guess "what events can comes next".

A solution to this problem of ZFS is to temporarily reduce the Markov order, changing the matrices used for the generation from [*S O*] to [*S O-1*] to check if, with a smaller list of past events (which makes the finding of candidates for the future event less strict), the system can find available candidates. If no candidates are found at [*S O-1*] the reduction process can be repeated progressively until order 0 is reached. At this point any of the probable events on a given step with probability higher than zero can be played regardless of the past list of events. This technique solves the ZFS problem, given that, if the previous events are not important then the system can always find candidate events in a style and, once restored, catch up with the established order for the generation process. However, after a ZFS is reached and resolved by this methodology, the higher order can not be restored immediately but must be progressively increased.

In order to implement this variable-order methodology, the analysis must be performed for different orders (from order zero to the highest desired order). In this way, when the system is reproducing a sequence it can shift from one order to another if a ZFS is found at any time.

4.5 DrDrums

Throughout sections 4.3 to 4.4 a complete generative system based on analyzing drumming styles in symbolic format is presented in detail. Different steps of a generative system are covered. Extracting a style from a set of drum pattern examples and turning them into probability matrices, using the matrices to generate patterns, controlling the generation process by using style-based transformation techniques, and finally generating pattern variations. All these techniques, added to the monophonic transformation tool described in section 4.2, comprise a powerful toolkit which can be assembled into a complete generative music system.

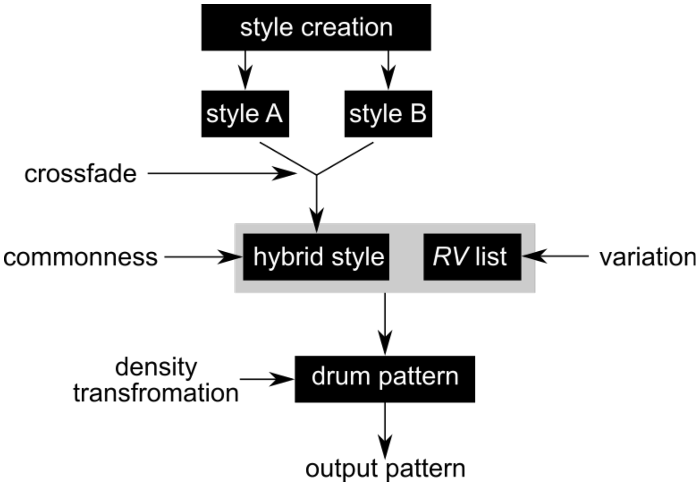

Figure 4.11. Block diagram of DrDrums based on the concepts and algorithms presented above.

In this section, the DrDrums application is presented as a tool which unifies the techniques and algorithms for style analysis, generation and transformation presented in this chapter[33]. DrDrums responds directly to the needs of EDM producers (as discussed in section 2.6) specially concerning dynamic stylistic generation and transformation of drum patterns. The inner workings of DrDrums are presented as a flow diagram in Figure 4.11. Style extraction allows to create a symbolic representation of a style based on processing a collection of patterns. The generated styles can be combined using a crossfade pattern, thus obtaining a new hybrid style. Such hybrid style can be processed using the commonness algorithms, and then using an RV a new drum pattern can be generated. Such RV list can be manipulated using the variation algorithms until, finally, the generated pattern can be post-processed using using the density transformation algorithm.

DrDrums is implemented in Pure Data (Puckette, 1996) as a fully functional app to work as a standalone tool with autonomous BPM and sound

[33]A video demo of DrDrums can be found here: http://bit.ly/2EuWJSB

reproduction engine; it can also be used as a DAW tool that can be synchronized with incoming MIDI clock and sending out MIDI note-on messages to be introduced in a DAW pipeline (Figure 4.12).

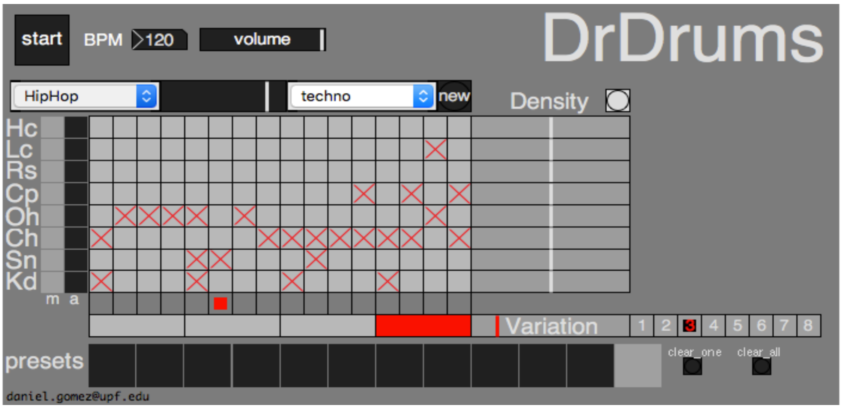

Figure 4.12. Screenshot of DrDrums interface.

4.5.1 DrDrums' new style creation

The first step to create a drumming style suiTable for DrDrums is to select the appropriate drum patterns in MIDI format. To test the capabilities of DrDrums, proper styles are created using the *RhythmToolBox, a set of* Python scripts developed during this thesis. The patterns collected for the styles are based on the exploration of EDM drumming styles throughout the literature presented in section 2.5.2. A total of 75 one bar patterns are collected, with a minimum resolution of 1/16th note. The authors have labeled these patterns in different styles, 70% of them belong to the House (28%), Breakbeat (26%), and Techno (16%) styles and the rest 30% belong to Garage, Drum n' Bass, Hip-Hop, Trance, Chillout, Dubstep, Jungle and Trip-Hop styles. By using the *MakeStyle* function of the rhythm toolbox, House, Breakbeat and Techno styles were generated using the following python script:

import rhythmtoolbox as rtb

```
house = rtb.midifolder2list('house')
rtb.makestyle(house,16,'styles','allhouse')
```

This script reads the folder /house to extract all MIDI files and then creates all the corresponding matrices of the style with the prefix '*allhouse*' and the suffix '*-O-S*' (order and step) following the methodology presented in section 4.3. These matrices are then inserted in the DrDrums/styles folder so they can be used in the generation.

4.5.2 Main controls

The main controls of DrDrums are those included in a typical grid-based drum machine, the Beats per Minute (BPM) controller to set the tempo, a volume control, a start/stop button, a grid to visualize the pattern generated and a playhead to visualize the current event being reproduced (see Figure 4.13).

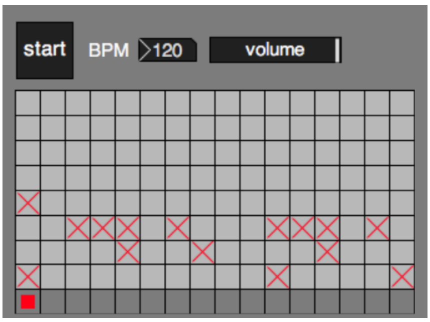

Figure 4.13. Main controls of the DrDrums application. Start/stop, BPM and volume controls (top), piano roll grid (mid), playhead (bottom).

4.5.3 Generation of patterns in style

Stylistic generation of drum patterns starts with the selection of an appropriate style and the generation of patterns. A drop-down menu presents different styles available which can be selected. Once a style is selected, a 'new' button is used to generate patterns in that style. This 'new' button can be selected on demand, obtaining a different and original pattern every time. The function of the 'new' button is to generate a new RV list so that under the same style a new pattern can emerge (see Figure 4.14).

Figure 4.14. Style selection menu in DrDrums.

4.5.4 Style combination

The style combination feature of DrDrums is presented as two drop down menus, which contain a list of the different styles available, connected by a continuous slider. Once both styles have been set, the combination of styles is operated via the slider, in order to achieve a proper balance of each style for the generation. As the slider is manipulated, new patterns emerge in real time in the pattern grid, as the style progressively mutates from one point to another. As mentioned above, as long as the 'new' button is not pressed, all the changes in the resulting pattern are reversible, allowing to fully recover a certain pattern found at a specific style combination (see Figure 4.15).

Figure 4.15. Style crossfade tool of DrDrums. Note the slider position quite close to techno and very far from HipHop

4.5.5 Inducing variation

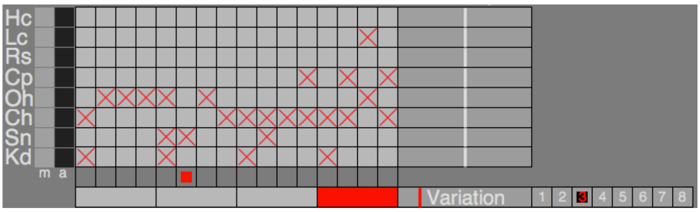

Figure 4.16. Variation control of DrDrums. Note the red slider at a position denoting a low amount of change.

All the parameters that control variation are located in the region below the playhead. There are four regions below the playhead which represent each of the four beats within one bar pattern. Each of these regions can be activated, indicating the beat which is going to be affected by the variation. The amount of variation is controlled by the 'Variation' slider indicating how drastic is the transformation that will occur. On the right side of the 'Variation' slider there is a time grid with a selector to control the number of bars that take place before each transformation (Figure 4.16).

4.5.6 Density Post Processing

The Agnostic Density Transformer (ADT), presented in section 4.2 as a post-processing tool for increasing or decreasing the number of onsets in a monophonic pattern, is used in DrDrums for each instrument. Using the ADT after all the stylistic generation process has taken place, adds the possibility to transform in real-time a new pattern, controlling the densities dynamically thus creating variations on the fly. A reset button that sets all densities to zero, to reset all densities to its original values, is located on top of all the sliders. The manipulation of these parameters adds a dynamic action to drum sequencing which enables musicians to convert a dynamic gesture to an immediate percussive result (Figure 4.17).

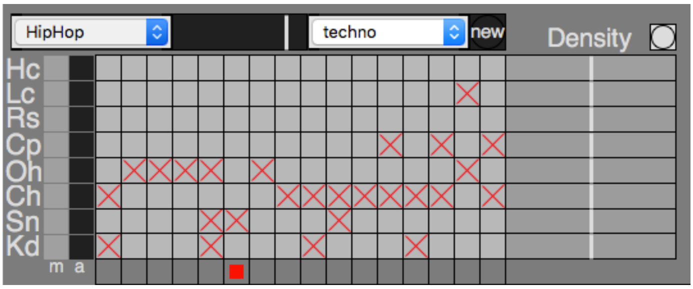

Figure 4.17. Density post processing sliders of DrDrums.

4.6 Quantitative evaluation of DrDrums

There is a tradition in the methodologies used for evaluating systems that have smart or intelligent features, all of them derived from the idea of the Turing test (Turing, 1950). The original conception of this test is to deduce if a machine can exhibit intelligence behavior which is indistinguishable from that of a human. Machines and their systems have diversified to solve a very wide array of specific problems, displaying intelligent behaviors in specialized and delimited areas. Therefore, tests for assessing intelligence in artificial systems have also specialized, focusing in specific areas rather than in a general idea of human intelligence.

Specifically, in the case of musical systems, there has been an emerging concern on how to evaluate intelligent musical systems (Ariza, 2009; Agres et al., 2016). The main ideas derived from the literature devoted to music intelligence evaluation are that evaluations must be performed at different levels of the system and that a certain subjectivity is hardly unavoidable in some of these levels of evaluation. In this section two different methodologies are presented intended to test the capability of *DrDrums* to pose as a human EDM drum composer.

Two procedures for evaluating the generative capabilities of *DrDrums* were carried out, both aimed at understanding the stylistic properties of the patterns generated and thus the effectiveness of the system in replicating a drumming style. Two EDM producers were selected given their international role in professional level music production and their special relation with percussion. The two musicians are Sebastian Hoyos, who uses the stage name Sano[34], and John-Erik Boska, who uses the stage name Boska[35]. Both of these producers are musically trained, Sano in Afro-Cuban piano and Boska in Latin and African percussion. Both producers were asked to create 10 drum patterns in their own style, as the ones they could use in their regular productions. They were asked to limit the MIDI velocity to a constant value and to use only the eight instruments available in DrDrums (kick, snare, closed hi-hat, open hi-hat, clap, rim shot, low conga and hi-conga). They were allowed to use their own kit of samples for reproducing each of the percussive instruments when creating their patterns.

The first experiment is a multi-subject activity, where the objective is to evaluate if listeners can establish a difference between the patterns created by the producers and the patterns generated by DrDrums in each producer's style. This experiment explores the generative properties of DrDrums from a listener's perspective, as the subjects learn the style of the producers by listening, and then compare new stimulus with the style.

[34] Sano's discography: http://futura-artists.com/sano/

[35] Boska's discography: https://www.beatport.com/artist/boska/139462

4.6.1 Method

4.6.1.1 Participants

The participants are 20 subjects, most of them Master students from the Music technology Group at the Pompeu Fabra University, with very diverse cultural and geographic backgrounds. All subjects report having at least 1 year of formal music training.

4.6.1.2 Material

Dr.Drums is trained with each producer's MIDI patterns, creating a drumming style model after each of them. This is done by using the rhythmtoolbox.py set of functions and by running the following python script:

```
import rhythmtoolbox as rtb
sano = rtb.midifolder2list('sano')
rtb.makestyle(sano,16,'styles','allsano')
```

This script uses the *midifolder2list* function to look for a local folder named */midi* and, inside, it will look for a subfolder named as indicated in the function, in this case 'sano'. Then, the *makestyle* function creates the style representation based on those MIDI files in the folder, which are 16 steps length, and then save them in a folder called 'styles' and naming all output files 'allsano'.

Once both styles are extracted, they are loaded into DrDrums, and then they are used to generate new patterns in each producer's style. With these new generated patterns the quantitative evaluations are carried out. This evaluation stage is based on the idea of style as a consequence of the existence of common, possibly preattentive, features in a group of musical pieces which can emerge in a listener's mind by exposition (Meyer, 1957).

4.6.1.3 Procedure

The procedure of the quantitative evaluation consists of selecting randomly five of the original patterns from each producer and reproducing them with a neutral drum kit (the original sounds of the Roland TR-808 drum machine), to create a context of the producer's style in the listener. Then, after the context is presented, subjects listen to 17 new evaluation patterns and determine how they fit the original drumming style. To select the evaluation patterns, a random set of *5* patterns originally created by the producer (not used in the context stage) are selected. Additionally, ten patterns generated using Dr.Drums within this style, and finally two patterns of the other producer's style are also included.

The task of the listener was to value the degree by which the evaluation patterns belonged to the same context that was presented in the beginning of the experiment. To do this, each pattern was judged using a continuous scale ranging from 0 to 5. Where 0 meant 'no relation with the context', and 5 'complete relation with the context'. The evaluation patterns, as well as the patterns in the context, were reproduced with the same neutral drum kit and at 120 BPM, so the timbre and the tempo were kept constant during the experiment.

4.6.2 Results

Every pattern in the original style, generated by DrDrums and belonging to a different style has a subjective rating of its relationship to the presented context. The median of each rating is extracted as the representative value of the relationship of a pattern with the training context.

Results show how the original patterns, and the ones generated by DrDrums have higher relation with the context than the patterns that be-

longed to the other producer (Figure 4.18). This strongly suggests the style was grasped by the listeners, as the original patterns, the ones created by the producer, are ranked as more related with the original ones than patterns by a different producer. This effect is evident for both producers.
For both producers the patterns created by DrDrums are also regarded as more related with the context than the patterns created by a different producer. This means that the patterns that DrDrums is capable of generating have also a distinctive quality from the patterns of the other producer. That is, the subjects can differentiate the patterns generated form DrDrums in a specific style as being different form the patterns of another style, with more than one point of difference.

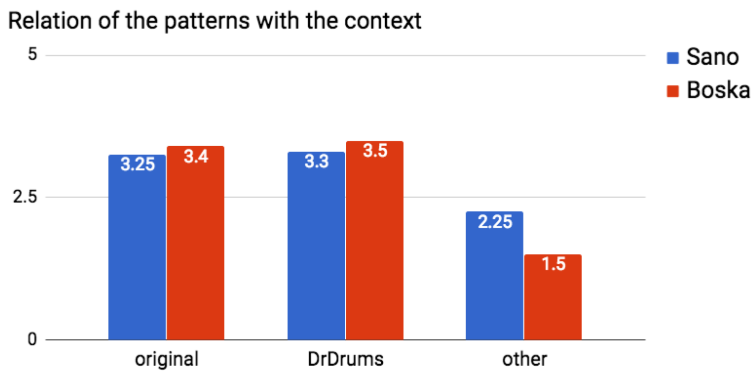

Figure 4.18. Subjective ratings of the relation of drum patterns with the context.

The magnitude in which the original patterns created by each producer, and the patterns generated by DrDrums differ from the patterns of another producer is very similar. This implies that for the subjects there is no perceivable difference between DrDrums' generated patterns and the original patterns created by the producer of the style. The patterns generated by DrDrums are no less stylish than the original patterns of the producer given the magnitude of the relation of the patterns with the context.

4.7 Qualitative Evaluation of DrDrums

The second method to evaluate DrDrums is a qualitative and subjective one, based on conversations with the two involved producers who listen to new patterns generated by DrDrums in their own style. This procedure seeks to evidence the stylistic generative properties of DrDrums reviewed from an author perspective. DrDrums is trained with an author's style and the generated patterns are judged by the authors who are (obviously) experts in their own styles. This is a different, perhaps more strict evaluation of the generative capabilities of the system, as the relation between an author's style and generated patterns goes deeper than acknowledging a relation between a pattern and a context: in this case a question of self replication and the quality of that self replication is at play.

4.7.1 Method

4.7.1.1 Participants

The participants are the same two producers who created the patterns in their own styles, Sano and Boska.

4.7.1.2 Material

The material for this quantitative procedure is the same material used on the qualitative experiment. That is, the ten original drum patterns created by each of the producers Sano and Boska.

4.7.1.3 Procedure

This evaluation is based on presenting the expert producers with their original drum patterns and then with ten patterns generated by DrDrums in their style. This evaluation is carried out several weeks after the ten original drum patterns were composed. After carefully listening to their ten patterns and the ten generated patterns, specific questions are asked, along with some contextual questions. These are the questions and the order in which they were asked:

- Do you acknowledge your ten original patterns as your own style?
- What do you think of the ten generated patterns?
- Do you feel your style is represented in those generated patterns?
- Can you envision a context in which you would use DrDrums?
- What do you think about the possibility of mixing styles?
- Do you see DrDrums useful for preserving your style?

The answers of each producer to these questions are transcribed in the following sections and the topics that emerge form the conversations are analyzed and commented in the discussion section.

4.7.2 Results

The following answers are edited transcriptions of the interviews carried out with Boska and Sano. Each of them was interviewed individually but for simplicity their answers to the same question are interleaved.

Daniel: Do you acknowledge your ten original patterns as your own style?
Sano: "Yes. But wow, they are very good (laughs) i could write a hit with each of them".
Boska: "Yes".

D: What do you think of the ten generated patterns in your style?
S: In general it replicates me. I can listen to myself there, in some intentions where I make the syncopations and the cadences that I use but, I can also recognize easily where it does not because I can feel where it is not working. For example patterns 3, 8 and 9, I would have not taken those decisions to make those rhythms. Two of them yes definitely I could have done them, patterns 0 and 4. And the rest, I could not have done them, there is a certain surprise factor that I like, and those I would have been interested in doing them. I like them a lot and would have liked to be able to come up with them.

B: Its amazing! Some of the patterns are very similar to the ones I made, 3 or 4 instruments are the same to my original patterns. But other patterns are not. I am really surprised by the sensitivity the drum machine is showing. In some patterns I wouldn't put so many notes because what I tried to do was to not put all instruments in all parts of the measures. I like to reserve some instruments at some positions of the measure. So then I notice that after some patterns the style gets very minimalistic and way more similar to the style I programmed and really, really sensitive. It would be so easy for a computer to program all of the 8 instruments in every beat of the bar. But the sensitivity here is amazing like its really musical sensitivity, because the rhythms are focused on one or two instruments and the rest of the instruments are only appearing in very specific spots. yeah I am really impressed by this.

D: would you use a system like DrDrums for replicating your style?
S: Many times the ideas that I have start with the rhythm. When I listen to something or try to replicate something that makes me curious. In that sense, to receive something automatic, I do not feel it inspiring. However, if i see it from the perspective of not being inspired and have to start creating music I see it as an interesting option but I would not be conformed with its output. I would always leave the door open to edit it and transform it afterwards. To use it as a seed in MIDI format. I miss the appropriation, as my way of working is always to disrupt, invert, transform and mute until I get this "aha!" feeling.

Potentially, someone who has to make many tracks would love to use this like "I have a certain amount of time to do a track and then I have to make another one." If you are against the clock this is a great tool. In that scenario it would be a great tool, a proper tool. Someone who loves automatic music creation and that his music has some of it, would make a lot of profit form this tool. For me it is difficult to assimilate to my music because of the way that I make my music. But there are some other expressions where DrDrums fits very well.

B: *Im thinking. Maybe in one scenario this might be useful for me. I work with Clavia on the Nord Drum[36], and making patterns that are actually good, like this, would be very useful for me for making demos of the sounds and the instruments. Because there is a difference when I make music for myself and when I make that stuff for production. And there is one specific challenge that occurs: when you make ten sound demos in two days, keeping the creativity up is really difficult, because what I always do there is that I always clear the pattern and start over to try to think fresh, to try to start over and make something new. If I had something like this I could randomize a beat and it would be a good starting point and I would be able to work from there. As what I do for Clavia is to generate sounds for presets. so my work is sound demo production and preset generation, so if I could take off the weight of creating patterns it would be great.*

For musicians who are not so adept with rhythm it is fantastic too. because for me it's very easy to come up with rhythms but the challenge with Clavia is to synthesize sound designs as well so it is an overload to my brain. to think of a musical idea and a complete sonic landscape. It would be good to have that kind of tool. For someone who is no so adept with rhythm it would be good to start with these rhythms.

D: What do you think about the idea of mixing styles. Lets say I can mix my style with someone's style or with the style of a great track you love. Would you be interested in such a tool?

S: When I make music my first intention is very clear but it is very difficult for me to go to a B part of the track: how to break the pattern and how to transform it and make it change? That is very difficult for me. This system would be useful for generating B parts of my patterns for a song without having to break my head around the variations. An automatic system that can suggest material to me without going away from the original idea A.

[36] Clavia is a Sweedish company that manufactures synthesizers and drum machines. More information: http://www.nordkeyboards.com/

Like a wizard, a one click tool that makes a variation. I see it analogous to the "swing" or "quantize" commands in a DAW, a very straight forward command, where I don't feel that the algorithm is doing anything creative for me.

However, it is a complex scenario. It would be great to test ideas. But, as I imagine it, it would not be a default tool to use in my production pipeline. It would be more as a curiosity. As I am currently trying to take my work away from the computer screen and if I see that the machine makes something that I can not make or that I can not think about… it would be frustrating in a way, as "How can the computer make this and I can not" or in the worst case scenario the new idea could take my concentration away. For another type of production scenario this could be a great starting point, for the genres that are more extreme (as breakcore or IDM) this tool could be perhaps more interesting.

B: Yes. Its super interesting. I think for example it changes the concept of collaboration in electronic music. And it brings a conceptuality, opens the door to think about the interaction between two musicians in a an almost conceptual way. Lets say I contribute my style to DrDrums and then other producer makes some music with it afterwards, music that I did not make but is made with traces of me. So, if he can make music with the patterns generated by the software, then I did not make it but I am still there. I love the intersection between the AI and the human. It's somewhat attractive. Its quite a very attractive prospect. Im not sure how many times I will be interested to do it, or if I would use it on a regular basis. It depends on the kind of work I have or how it turns out to be in the long run, but for sure form a purely conceptual standpoint is very interesting.

But I know how musicians would use it all the time. I have a friend who has problems with creating source material for her music. She is a good producer but she does not have many means for creating musical ideas so she mostly curates other musical elements that she samples or finds or records and this is very comfy for musicians. I mean this is why sampling *exists. If with this tool you can create patterns in MIDI and import them and edit them in Ableton Live and apply anything you want to them it, would be vey useful.*

D: If you were working on a record. But you wanted to evoke somehow what had happened rhythmically on a previous record. And you trained a system with that record to be recalled, do you think that this tool could be attractive for having some dynamic memory of what was done and to be useful as a new starting point? Would you like to preserve your style?

S: I would not use it. I can not project myself to such abstraction. I would rather use a loop or a track not the idea of a memory.

B: Gaining perspective of your ideas or your past or from the present. For me personally, no. I try to artistically focus not on the past or tradition or history or anything. So this is something I'd rather forget. I would not want to keep around ideas from my previous releases. I always try to come out with at least a new idea when I am making a song. So not for me, but I know for other artists this could be interesting. I have some friends who have drum programming as the main focus of what they do and who contribute to other production with rappers or other songs only with drums. Their full production was only drums because they are fantastic drum programers. So, one of this guys told me that when I make a new song I should make it in the file of the project of the song I produced before. So instead of starting from zero to try to make a song it would be good to start from where I left off the other idea and keep working from there. So I respect this perspective and I see that a lot of people have it, to start working where they left off. For someone like that, very analytical or design oriented, yes, it could be an interesting tool. But the tool should be capable of analyzing on its own the recordings. I do not know if an artist would convert all his patterns back to MIDI, it would have to be all automatic.

So i probably would not use it to analyze previous releases but maybe other people would. But for example the collaboration you mentioned before I find it very interesting, you already put my style and Sano's style together so now me and Sano can start having a musical exchange without having a musical exchange. The idea of making a collaboration with Derrick May without him knowing, thats something I would be interested in.

4.7.3 Summary and discussion

The main idea that comes up after both producers are exposed to patterns generated by DrDrums in their own styles, is the acknowledgement of the style imitation power of the system. Both music producers are positively surprised with the quality of the generated patterns, and they both argue that it replicates well their own styles. They also point out that some patterns are better than others, Sano claiming that 7 out of the 10 generated patterns are really good replicas of his style as well as Boska, who confirms the system has “musical sensitivity” although some of the patterns were too similar to the original patterns he created. In general the output is regarded as surprisingly good and accurate stylistically.
The idea of using the software on a daily basis in their artistic practice is somehow not appealing. For Sano, the idea of the machine improving over some work he can’t do, as making variations on the fly, is not completely attractive, although he acknowledges that it is one of his main difficulties for music production. For Boska, the machine can be a source of creativity but he, as a musician, feels that he must be in charge of the production process as he studied specially how to do so. Both of them offer possible scenarios where the use of DrDrums might have a proper impact:

- When a producer has to make a certain amount of compositions in a given amount of time. Boska explicitly says how, in his drum-sound design job, he would appreciate a tool that could at least solve one of the dimensions in which he has to be creative at, so he can focus on the other ones. He would use DrDrums for the patterns and focus in designing timbres.

- When you are a producer not very adept with rhythm or have a hard time being creative with drum sequencing and production; perhaps artists who use samples of other music as their composition technique. This scenario is particularly important, as these are the precise users the GiantSteps project was targeting. Artists that are still developing their skills but yet they have the desire to “sound like someone” who is inspiring, or perhaps as a referential artist in the EDM sub-genres they are interested in.

For Boska, more than Sano, the idea of mixing drumming styles is very appealing. The possibility to have virtual collaborations between them and other musicians, by loading the style of another drum programmer into DrDrums, triggers the curiosity and the imagination. However, the use of this system as some sort of abstract memory of a long gone past style, even if it is their own, does not appeal. They both agree that if they wanted to recall the style of a given period in their careers they would not record it in DrDrums but rather listen to the music they were doing at the time.

The results of this evaluation confirm the capabilities of DrDrums to replicate the style of a drum producer, validated by the responses of two EDM producers in which the system is trained on. The results of the listening tests carried out by both producers confirm that the methodology proposed in this thesis, for extracting knowledge from a group of symbolic drum patterns, is valid, useful, and can be profited in real life EDM production scenarios.

From a conceptual perspective, the success of the generative system implemented in DrDrums is the analysis of musical recurrences at specific moments in time. The results of this qualitative experiment confirm how the style, defined by the musical strategies and rules used by a drum pattern composer, must not necessarily be encoded explicitly in the set of rules to be used in a generative system. On the contrary, it is rather the effect of these rules and strategies, which are evidenced in the generated musical material, what is sufficient to encode a style into a replication machine as DrDrums. This confirms how the common recurrences found in a set of patterns form a mental representation of *style* in a listener's mind, in the form of expectations of certain events to occur at a given time, which are continuations of certain previous events (Meyer, 1967). In practical terms, it was not necessary an analysis, in explicit musical terms, of each producer's own material in order to recreate their styles; but only with the use of the manifestations of these rules in the form of sequences of drum events, the styles could be replicated. In this case, rules can be bypassed, as we are interested in the mental representation which is directly related with repetition and recurrence in drum sequences.

So far in sections 4.2 to 4.7 an algorithm-based system for generating drum patterns supported in the concept of style has been presented and evaluated. Every algorithm involved in this system was discussed independently, as well as its integration in a working prototype called DrDrums. This prototype was also evaluated in subject-based experiments where the stylistic properties are acknowledged, being regarded as capable to generate original patterns in the style of a human musician. The rest of the chapter is devoted to the automatic organization of drum pattern collections, based on the theoretical aspects discussed in Chapter 2 and further developed in Chapter 3.

4.8 Rhythm Spaces

For this thesis, EDM drum production has been studied in order to understand specific technological needs which help improve its current practice. One of such necessary improvements is to organize and explore a collection of drum patterns by their rhythm properties. Both activities expand the current state of browsing music files in a computer system, which is currently done in alphabetical order without taking into account rhythm or any other musical properties. Alphabetical browsing, although universally used, has proven to result in under-exploring collections of musical material (Turquois et al., 2016). Ideally, when organizing a collection of drum files by some of their meaningful properties, similar patterns should be close together so they can easily be browsed and retrieved. One structure that can deal with this type of arrangement is a low dimensional coordinate system, as it allows for the visualization of many elements located according to their values in the different dimensions. Such spaces favor an observer making sense of a collection, as she can both grasp the local relations of their elements as well as inspect the complete set. If, in addition, such rhythm space is made interactive, the observer can also point to a specific position and retrieve the chosen drum pattern directly to a drum machine, enabling its reproduction in real-time. In this sense the rhythm space becomes a percussive instrument which allows the sequencing of complete drum patterns by gesturing over an interface. The idea is then to create an interactive rhythm space that can be used for organizing, visualizing and retrieving drum pattern files.

Perceptual spaces of this sort, in which elements are arranged by similarity, have been commonly used in many different domains some of them being timbre (Gray, 1977) or color (Sheppard, 1962). These spaces propose frameworks for the organization of item collections based on perceptual properties, to make domains as colors or timbre browsable under specific interactive configurations. The property allowing elements in these perceptual spaces to occupy a proper location in relation to the other elements is perceptual similarity, as it becomes the attraction force to make some elements stick together and to stay away from others, while this organization makes sense to a human user. The backbone of a rhythm space for EDM drums is an adequate similarity distance, measurable from the patterns themselves and aligned with closeness sensations of human subjects. Given this perspective, it makes sense to use the results presented in Chapter 3, regarding polyphonic drum pattern similarity based on rhythmic descriptors, and use it as the bond supporting drum rhythm spaces.

In the process of studying rhythm spaces it was evidenced that using the appropriate metrics, any collection of drum patterns can properly be arranged into a low dimensional map where points represent patterns that can be retrieved. However, such space is discrete, limited to jumping from one pattern of the collection to another, restricting the possibility of a more continuous and nuanced exploration of rhythm. Expanding on this idea, and connecting with the previous experience of combining two different drumming styles (presented in section 4.4), new algorithms for drum pattern interpolation are developed (see Figure 4.19) as an expansion to a drum rhythm space. By implementing these algorithms, as a layer on top of the rhythm space, the space becomes continuous, where any blank point (a point where no pattern from the collection is located) can retrieve a new pattern not contained in the collection of the rhythm space, but created in real-time based on its neighbors. With these algorithms, a rhythm space is enhanced with generative capabilities, becoming a visualization tool that auto generates new elements on the fly, expanding its original components.

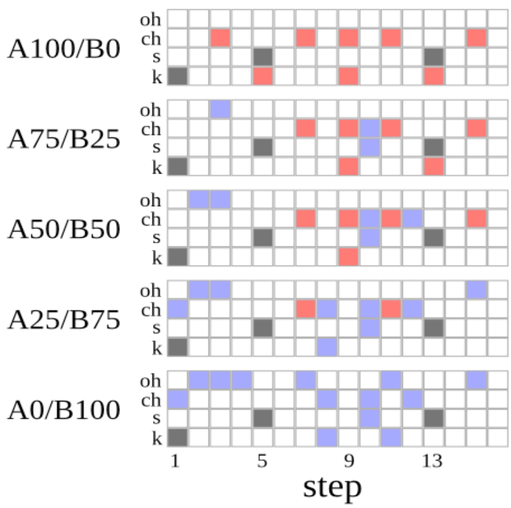

Figure 4.19. A progressive interpolation between two patterns based on one of our algorithms. Patterns A and B have three onsets in common: a kick on step 1 a snare on step 5 and 13. Snapshots of the resulting interpolated rhythm are presented at three equidistant values of the interpolation between A and B.

4.8.1 Requirements of a rhythm space

With a musical tool as the rhythm space, a musician can make sense of a collection of patterns, browsing through them in a perceptually meaningful way and generating new patterns in real time. The main features of this interactive rhythm space as it was described are:

- Low dimensionality for ease of navigation.
- Closeness of similar patterns and separation from different patterns.
- Spatial continuity.
- Real-time retrievability of patterns.
- Generativity (in the sense of producing new patterns, different from the ones in the collection, when navigating through an empty region)
- Editability (openness to the addition of new patterns, and to the removal of existing ones, so the pattern collection can be updated).

The exploration of polyphonic rhythm similarity measures presented in sections 3.5 and 3.6 advance some of the main features of the rhythm spaces presented here. By making human-based pairwise similarity comparisons between each drum pattern in a collection, a matrix of distances between the patterns can be established. Using the multidimensional scaling (MDS) technique such a dissimilarity matrix can be converted into a low dimensional space, a map, where the reported distances are respected so the drum patterns are located according to their (di)similarity. By taking this approach, an amorphous collection of patterns is converted to a coherent space where patterns that share common features, relevant to human perception, are close together. One of the contributions presented in Chapter 3 is an algorithm that can model the features which are relevant to human perception, so that distances among patterns, and thus rhythm spaces, can be predicted automatically. By using this method, subject-based similarity ratings are replaced by automated similarity computation, so rhythm spaces can be generated without the need of additional human similarity ratings. By using this system, symbolic drum patterns are input and the coordinates to locate the patterns in the space are output.

Once the exact coordinates of each pattern are established, an interactive system for navigating the space can be implemented. By comparing the coordinates of a pointer with the coordinates of each pattern, a proximity radius can be established so that when the pointer is very close to one pattern it is output by the system. Searches throughout the rhythm space can then be carried out by scrolling the pointer along the space, retrieving the drum patterns from the collection to a drum reproduction system as a drum machine or a sampler.

4.8.2 Rhythm Interpolation

As a means to add continuity to a rhythm space built up from a discrete collection of patterns, three different algorithms for drum interpolation are proposed. Based on a Delaunay triangulation (Lee and Schachter, 1980) of a 2D rhythm space, these algorithms weight the three surrounding patterns of any point in the rhythm space, in order to achieve smooth transitioning along the space. A transition within three different rhythms suggests a new hybrid pattern is created, with features that resemble the patterns surrounding. Our algorithms take care of smoothly introducing and removing onsets in the pattern, based on the interpolation values.

By using the triangulation, the position of a user-controlled pointer is always inscribed inside the area of a triangle, while it navigates through the rhythm space. Based on the drum patterns located on each vertex of the triangle and the distance from the pointer to each pattern, our algorithms generate a new pattern as an output.

To compute the weights of each pattern at the vertex of the Delaunay triangle, the distance from each vertex (A,B,C) to the user pointer (P) is measured. Each distance ($\overline{AP}$, $\overline{BP}$, $\overline{CP}$) is normalized by dividing its magnitude by the distance of the line that starts in each vertex, ends in the opposite edge, and passes through P ($\overline{AD}$, $\overline{BE}$, $\overline{CF}$). The normalized value of each vertex to the pointer is subtracted from 1 in order to convert a normalized distance value into a closeness value (αA, αB, αC). The closer the pointer is to each vertex, the closer α to 1. This value is used as weight for the interpolation of each pattern associated with the vertex of a triangle in the rhythm space (see Figure 4.20). Having the three weights (αA, αB, αC) associated to each pattern in the triangle, the interpolation algorithms can be applied to generate a fourth resulting pattern located at point P.

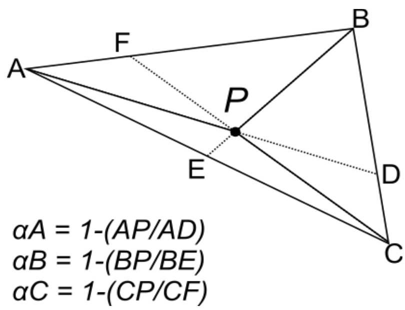

Figure 4.20. The weights for each pattern derived from the distance of point P to the three vertices of the triangle ABC.

The three different algorithms presented here, have the same basic approach which is simplifying pattern interpolation to a step-by-step interpolation. The problem is then reduced to process interpolation at the step level based on the three weights (αA, αB, αC), and then concatenate all output steps to reconstruct a complete interpolated pattern. Each algorithm is progressively more complex than the other regarding the representation of the steps input to the process. The first algorithm, Step by Step Naive Algorithm (STENA), uses only the event representation at each step. Events are used as the entity to be interpolated regardless of the instrumental onsets it represents (remember that each event represents combinations of different instruments at a single step). The other two algorithms, Step by Step Event to Onset (STEVO) and Step by Step Event to Onset and Density (STEVO-D) are based on the decomposition of each event in its onsets (see Figure 4.21). Additionally, the STEVO-D algorithm uses an additional *density* value to process the interpolation at each step.

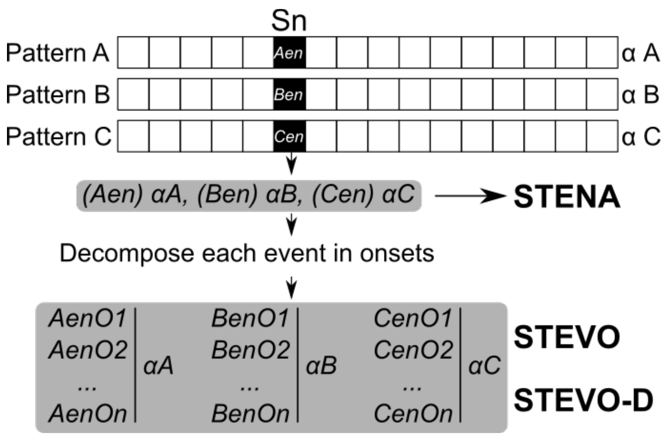

Figure 4.21. Scheme of the representations used by the different algorithms. Patterns are decomposed in events at each step *Aen, Ben, Cen* (top) and each pattern has a different weight *αA, αB, αC* (top right). The STENA algorithm is based on the event representation and the weights processed at each step. Events can be decomposed into its respective instrumental onsets (*Aen=AenO1+ AenO2+AenO3 ... AenOn*) and the STEVO and STEVO-D algorithms are based on that representation.

4.8.2.1 Step-by-step Naive Algorithm (STENA)

This algorithm processes each step independently, extracting a resulting interpolated event at each step. The final interpolated pattern is the concatenation of every interpolated event. To extract an event at step *Sn*, the STENA algorithm sums the weights of the common events found at each step (if any) and then selects the event at each step which has the highest weight. This operation is repeated for each of the steps in the pattern until a complete list of interpolated steps is output.

For example, if three given patterns A, B and C have respective interpolation weights of 0.4, 0.4 and 0.2, and if at a given step *Sn* the event for each respective pattern is 1, 3, 3, then the interpolation procedure is a factorization of the common events:

$$0.4\ (e1) + 0.4\ (e3) + 0.2\ (e3) = 0.4\ (e1) + 0.6\ (e3)$$

The algorithm selects the event at each step with the highest weight, in this case the event (3) which has a weight of 0.6.

This algorithm is easy to implement and works in the event representation without any further information of the onsets that compose the event. With this algorithm the event of the pattern that is closest to the pointer will always have a higher weight than the other two, and can only be overthrown in the factorization stage, if the other two have the same event with a weight sum higher than 0.5. This condition limits the area of effectivity of the algorithm to a triangular region inscribed within the interpolation triangle (see Figure 4.22). The triangle *A'B'C'* is defined by the midpoints of each edge of the triangle (*AC'=C'B, BA'=A'C, AB'=B'C*) and inside it the weights αA, αB and αC of each pattern are less than 0.5. This means that inside the triangle *A'B'C'* the STENA algorithm is effective.

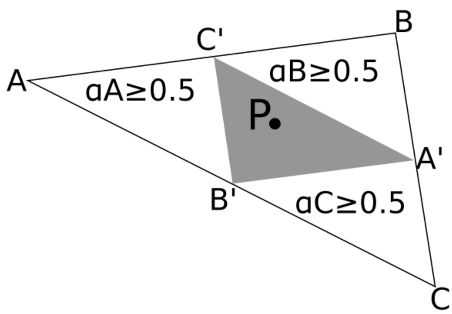

Figure 4.22. The ABC triangle is generated inside the Delaunay triangulation and the point P marks the interpolation point. The lines *B'C', A'C'* and *A'B'* delimit the areas where αA, αB and αC, respectively are greater than 0.5.

In the outer areas of the inscribed triangle, delimited by triangles $AC'B'$, $BA'C'$ and $CB'A'$, the weights αA, αB and αC respectively, are ≥ than 0.5 thus resulting in the inoperability of the STENA algorithm. These weight values cancel any possibility of the complementary patterns influencing the result of the interpolation, i.e. in region $AC'B'$ the only possible outcome of the STENA interpolation is pattern A. This geometric constraint limits the operation of the STENA algorithm within the boundaries of the $A'B'C'$ triangle.

4.8.2.2 Step-by-step Event to Onsets (STEVO) algorithm

As its name suggests, the Step by Step Event to Onsets (STEVO) algorithm is based on decomposing the events into onsets and using the weights (αA, αB and αC) derived from the position of point P and each vertex of the triangle ABC (see Figure 4.22). In every pattern at a given step n in the three different events namely Aen, Ben and Cen are decomposed in its onsets, i.e. Aen is converted into its onsets $AenO1$, $AenO2$, … $AenOn$. The resulting onsets and the weights of their corresponding patterns (αA, αB and αC) are grouped by instruments, and the weights of the common instruments are added. The result of the addition is a list of unique instrument onsets with their respective weights, from which a sublist is extracted and then the sublist converted into an event. This procedure is repeated for every step in the pattern until a new sequence of events conforming a pattern is obtained. In the STEVO algorithm the main decision is how to convert the complete list of onsets and weights into a sublist, in other words, how to filter the onsets that will make it to the output to the algorithm.

After processing the weights of each pattern and the events in a given step, the result is a collection of unique onsets and weights. The unique list of onsets ordered by weight can then be filtered by weight or by number of onsets. The STEVO algorithm is based on using the maximum value of the three weights (αA, αB and αC) as a filter to eliminate the onsets which have a lower weight value and thus do not make it to the output.

For example, lets say we have three patterns *A*, *B* and *C* and at a given step *Sn* they have the following decomposed events and weights:

Aen=[snare, closed hi-hat, high conga], αA=0.4
Ben=[kick, closed hi-hat], αB=0.3
Cen=[kick, high conga, clap, rimshot], αC=0.3

The resulting factorization of the onsets is:

snare (0.4)
closed hi-hat (0.4+0.3= 0.7)
high conga (0.4+0.3 = 0.7)
kick (0.3 + 0.3 = 0.6)
clap (0.3)
rimshot (0.3)

As the maximum weight value is 0.4 the onsets that are output at this step are those equal or above that value which are: snare, closed hi-hat, high conga and kick. This list of events is converted back to event representation. This procedure is repeated for every step in the pattern to achieve a complete list of events that finally conform an interpolated pattern.

4.8.2.3 STEVO-D with fixed density profile

The decisive factor of the STEVO and STEVO-D algorithms is the threshold used to control which onsets are output. STEVO-D proposes a threshold, related with the vertical density (the number of simultaneous instrument onsets at each step) of the interpolated patterns. Thus, an ideal density value to filter the onsets is obtained by interpolating the densities of the three events with the corresponding weight.

To extract the onsets at a given step *Sn*, the weighted density is computed by extracting the vertical density of each event, multiplying it by the respective weights and then summing all the components. To decide the output of step Sn, each event is decomposed to instrument onsets, each instrument onset is multiplied by the weight of the pattern it comes from and, finally, the common instrument onsets and their weights are factorized. The resulting list of instrument onsets and weights is filtered out by the interpolated density profile at step N. Only the instrument onsets which have a weight equal or above the weighted density are output. The output onsets are converted back to a single event.

This algorithm favors the mixture of elements at each step, allowing onsets that do not exist in the pattern with highest weight to make it to the output pattern.

4.8.3 Interpolation algorithm discussion

The three interpolation algorithms presented above use different techniques for selecting the onsets that will compose the resulting interpolated pattern. A comparative example of the way these algorithms work is presented in Figure 4.23 where three different events at a given step *Sn* with different weights αA, αB and αC generate distinct results.

The STEVO and STEVO-D algorithms offer an interpolation with more combination capabilities than STENA, as they decompose events into their most basic elements (i.e. drum onsets). With the example presented in Figure 4.23 the STENA algorithm outputs the same event as the highest-weight pattern, which is pattern *A*. As mentioned above, STENA allows a different output from the event with the highest weight only if the sum of the weights of the other two events is greater than the highest weight, and the two events are the same. On the other hand, the STEVO algorithms use information from a lower level, as events are decomposed into onsets, allowing lower-level mixtures of sounds, and thus better suited for the purpose of mixing.

The filters proposed for each of the STEVO algorithms differ conceptually and have different implications in the final result. The use of the maximum weight as the filter for deciding which onsets are output, as used in the STEVO algorithm, suggests the maximum value for every step in the interpolation process, but has no concern for the weights whatsoever. This density value will always be in tune with the pattern of the maximum density in the Delaunay triangle, regardless of its closeness to the interpolation point *P*. In the case where one of the interpolated patterns has a very high density compared to the other two patterns, it will force the output of the algorithm to include more onsets than the average. In general terms, this is an algorithm that leans towards the highest density at every step, inducing interpolated patterns to have more instruments. The STEVO-D algorithm, on the other hand, is more responsive to the interpolation values regarding the filtering of the number of onsets. Instead of having a constant value throughout the triangle (as the STEVO does), STEVO-D adjusts to the density of the highest-weight interpolated pattern.

One example of how the three different interpolation algorithms behave at the step level is presented in Figure 4.23, where, at a given step *n*, three events and their weights are used as input. Patterns *A*, *B*, and *C* have interpolation weights of 0.4, 0.3 and 0.3 respectively. The event *n* from pattern A, *Aen,* has onsets of instruments snare, closed hi-hat and hi-conga. *Ben* has onsets of instruments kick and closed hi-hat. *Cen* has onsets high conga, clap and rimshot.

The interpolation results for the step are presented in the lower portion of Figure 4.23. Where each algorithm outputs a different group of onsets for the same conditions. The STENA algorithm outputs the event with the highest weight, which in this case is event *Aen* composed of snare, closed hi-hat and high conga. The STEVO algorithm outputs the onsets with the highest weights after the factorization up to the highest density value (max density = 4). The output onsets by the STEVO algorithm are snare, closed hi-hat, high conga and kick drum. Mainly the same onsets as the STENA algorithm plus the kick drum. The STEVO-D algorithm limits the output of instrument onsets to the sum of the weighted density which in this case is 3. The onsets output by this algorithm are the closed hi-hat, the high conga and the kick drum.

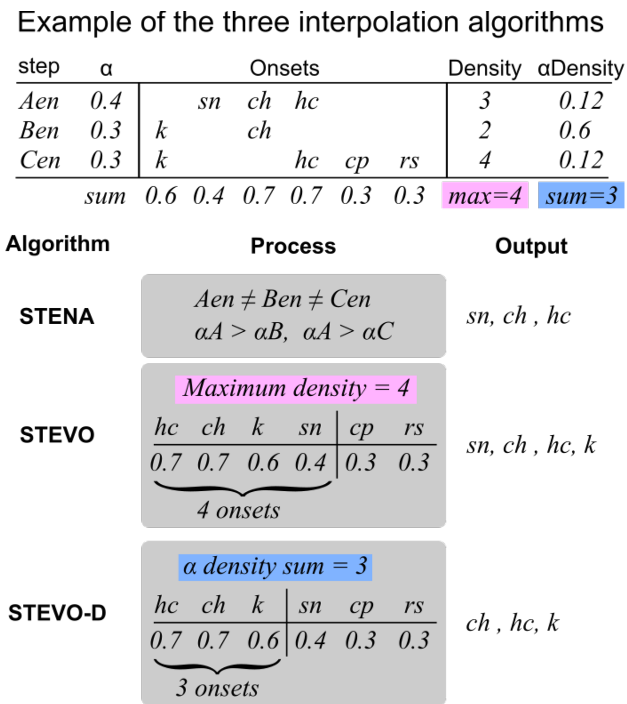

Figure 4.23. Example of the output of the three different interpolation algorithms. Three events *Aen*, *Ben* and *Cen* are presented on top, with respective weights 0.4, 0.3 and 0.3. The events have different instruments k: kick drum, sn: snare, ch: closed hi-hat, hc: high conga, cp: clap, rs: rimshot. The process of the different algorithms is presented below, each with a different output.

4.9 RhythmSpace Application

Up to this point all the constitutive elements for the automatic creation of a rhythm space have been presented. The descriptors to compute a distance metric and the appropriate dimensional reduction technique (presented in Chapter 3), the method to define triangles for the patterns in the space generated using a Delaunay triangulation, and different interpolation algorithms (Sections 4.7 and 4.8). With these elements an application is created in PureData which starts from a collection of patterns and results in an interactive rhythm space (see Figure 4.24).

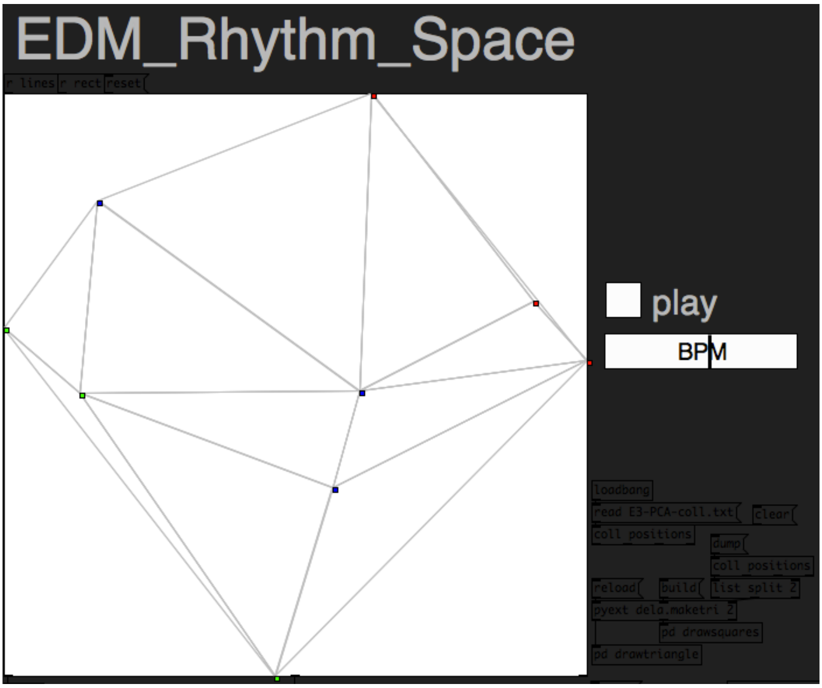

Figure 4.24. Screenshot of the EDM rhythm space.[37]

The pipeline (Figure 4.25) for creating a rhythm space starts with the selection of a folder which contains MIDI drum patterns. These patterns are all analyzed, extracting the descriptors found in Chapter 3 to be the best predictors of polyphonic similarity. Following the methodology from chapter 3, the Euclidean distance is measured between the five-dimensional vectors extracted from each pattern in the collection. With this informa-

[37] Video available: http://bit.ly/2EuWJSB

tion, a diagonal dissimilarity matrix is created where distances between each pair are reported. This matrix is finally processed by a multi dimensional scaling algorithm (MDS) which is set to retrieve a bi-dimensional solution for the dissimilarity matrix. In this final stage, the two coordinates to locate the different patterns in a space are obtained, representing a similarity-based organization. The coordinates are then input to a Delaunay triangulation algorithm which defines which sets of three patterns conform the triangles found in the bi-dimensional space.

With this architecture setup, the space is ready for interaction. A pointer input by a musician indicating a region in the rhythm space, represented by a bi-dimensional coordinate, is used to find a bounding triangle from the ones in the Delaunay list. Having the bounding triangle, the three patterns for the interpolation and the coordinate of the pointer are fed to one of the interpolation algorithms to obtain a resulting new pattern.

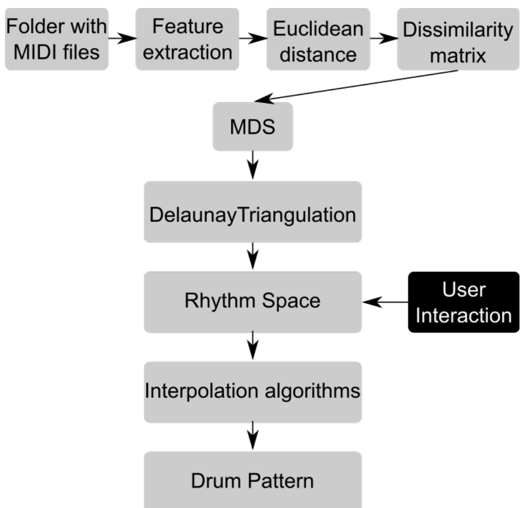

Figure 4.25. Block diagram of the process of converting a collection of patterns into a rhythm space.

The implementation of the RhythmSpace application is done in the Pure-Data programming language, using the py object to recall different python classes which are executed inside the PureData environment. The Python classes used in this application are all part of the RhythmToolbox collection as well as the RhythmSpace application which are available as open source download[38].

4.10 Conclusions

In this Chapter we have introduced several systems, aimed at processing, organizing or generating symbolic drum patterns. The first system is the agnostic density transformer (ADT) aimed at adding or reducing the onsets of a monophonic drum pattern, exploiting current knowledge of human rhythm cognition. The second system is *DrDrums*, a drum sequencer with style knowledge. Each constituent algorithm of DrDrums, related to style extraction or processing, was also presented in detail, along with its functional motivation. The last system, *RhythmSpace*, a tool for organizing, visualizing retrieving and interpolating drum patterns is introduced. RhythmSpace is derived from the original research results presented in sections 3.4 to 3.6.

The conceptual assumptions related to note salience and beat salience which inspired the Agnostic density Transformer (ADT) are currently being tested in real life products. Although the ADT has not yet been tested in experiments involving human subjects, its simplicity and straightforward application and use make it a practical tool for transforming the monophonic components of a polyphonic drum pattern. The implementation of the ADT in a real life drum machine product serves the complete purpose of the Giant Steps project, as it was designed to develop and transfer knowledge between academic and private partners.

The results of the two experiments carried out to evaluate the stylistic properties of DrDrums, suggest the system can successfully imitate drumming styles. The qualitative evaluation offers a first-person view on the

[38] Online repository: https://github.com/danielgomezmarin/rhythmspace-demo

generative system, as the reviewers are the same authors of the patterns of the style in which DrDrums is trained. According to the comments of the reviewers, more than a half of the patterns generated in their styles have distinctive features which they relate with their own way of composing drum sequences. The quantitative experiment offers a third person view on the system, as the style is learned by the subjects when they begin the experiment and then used to compare and assess new patterns. This experiment reveals how, for the listeners, there can be no difference between the patterns generated by DrDrums and the patterns generated by the subjects, in terms of their relation to the style. This behavior is observed in both styles used as source for the stimuli, evidencing a constant appraisal of the generated patterns although the style is changed. This experiment also reveals how the patterns that belong to a different style from the one used to train the subjects, are judged as less related than the ones generated by DrDrums. This last finding shows how subjects can discriminate patterns from the training style (both original and generated) apart from patterns generated in a different style, thus validating the judgment of the subjects involved in the experiment.

As mentioned by Boska, one of the producers involved in the qualitative evaluation of DrDrums, generative systems as the one developed in this thesis can offer new perspectives in musical collaboration. The act of two musicians collaborating in a musical piece, each of them contributing creatively, can reach new meanings via style-based tools as DrDrums. With the current capabilities, the style of a musician can be extracted from a collection of his patterns and be available for interaction by means of the style crossfade capabilities. This procedure is indeed disruptive for the creative process, as one musician, or all of them, can be absent from the production process itself while still contributing with their style to the musical piece being created.

The RhythmSpace application offers a solution for the arrangement of drum patterns based on similarity. The system is based on knowledge about polyphonic similarity, derived from experimental results presented on Chapter 3. This application serves the purpose of automatically organizing a drum pattern collection by creating a bi-dimensional map where patterns are located by similarity. This rhythm space solves the problem of browsing through collections of symbolic drum patterns using non-musical dimensions, as the typical alphabetical order used in computer browsers. The functionality of these RhythmSpaces involves the visualization of every pattern in the collection and the capability of browsing the patterns by signaling a point in the space, making the system retrieve the sequence of the pattern so it can be reproduced. To expand this browsing functionality, three interpolation algorithms are created which allow the retrieval of patterns from an empty section of the space by creating, in real-time, a new pattern based on the surrounding patterns from the collection. With these interpolation algorithms, the discrete space initially defined only by the patterns inside the collection, becomes continuous, offering a user uninterrupted browsing and the retrieval of additional patterns that do not exist (yet) in the current collection.

Dynamic gesture sequencing of drums, one of the milestones for drum production proposed in this thesis, is achieved both by the use of the ADT and the rhythm interpolation algorithms. Thanks to DrDrums and RhythmSpaace's implementation in the production pipeline, sequencing becomes a dynamic act again, enabling the process of drum creation to become fluid and produced by means of continuous human gestures, as it had been since the early stages of mankind.

5 CONCLUSIONS AND FUTURE WORK

Throughout this thesis, specific problems found in drum composition and manipulation of electronic dance music (EDM) have been discussed and approached scientifically. In the process of researching these problems, some basic questions regarding human assessment of rhythm similarity in music creation scenarios have been advanced. These advances were processed and used as a foundation to create novel tools which help EDM producers undertake their work. Particularly, the problems approached are:

- The complexity of composing drum patterns in specific EDM styles without explicit musical definitions of what an EDM style is.
- The transformation and variation of EDM drum patterns in real-time, during the process of creating *dance tracks*.
- The organization and efficient search of symbolic EDM drum pattern libraries.

These problems led to carry out new experiments aimed at understanding human processing of drum patterns in terms of similarity sensations, and also inspired the creation of novel techniques for visualizing, generating and transforming EDM drum patterns. The outcome of this work is materialized in a series of algorithms that are articulated in two software applications, *DrDrums* and *RhythmSpace*. The agnostic density transformer (ADT) and *RhythmSpace* contribute with new gestural and continuous manipulation of drum pattern sequencing, which is a transversal topic explored in this thesis.

The original experiments presented here reveal how the similarity perception between two different monophonic patterns is affected by inducing or not a pulse sensation during the exposition of the acoustic stimuli. When the pulse is absent, a *syntactic* mechanism operates in the listener's mind evaluating similarity based on comparing the notes-and-silences pattern of onsets among both patterns being judged. On the contrary, when the pulse is induced before and during the exposition of the patterns, a *semantic* mechanism guides the similarity comparison, based on higher level rhythm concepts as pulse meter and syncopation. These experimental results also illustrate how, in pulse-induced scenarios, metrics based on syncopation are profitable for predicting the similarity between two monophonic patterns.

The experiments on human processing of polyphonic drum similarity led to the design of new symbolic rhythm descriptors, and to find how a few set of those descriptors (*lowsync, midD, hiD, hiness, hisyness*) is capable of predicting human polyphonic similarity sensations. These useful descriptors are based on syncopation, pattern density and rhythm complexity measured over low, mid and high frequency ranges. Using this set of descriptors and multidimensional scaling (MDS) a collection of symbolic drum patterns can be analyzed and processed to obtain a low dimensional map where all patterns are organized by similarity. These maps, called *rhythm spaces*, were created and evaluated in two different subject-based experiments. From these experiments it can be concluded that rhythm spaces created with this methodology align with subject-based rhythm spaces. Specifically, the arrangement of two subject-based spaces, one using EDM drum patterns and another using global dance rhythms, were successfully predicted using the methodology proposed. These results validate the whole approach taken to create rhythm spaces, comprising selected symbolic descriptors and a given multidimensional scaling technique.

Considering the accuracy for predicting human-based rhythm spaces, this method for analyzing and organizing rhythm patterns was converted into a novel tool for the visualization, retrieval and generation of drum patterns. An application called *RhythmSpace* uses the same descriptors and MDS technique discussed above, to convert a collection of symbolic drum patterns into a bi-dimensional space which organizes patterns automatically, given their rhythmic properties. The rhythm spaces created depict the collection of patterns analyzed as points in a space, located according to their similarity, with patterns perceived as being alike close together, and separated from different ones. This map-like structure is exploited in RhythmSpace to visualize a complete collection of patterns and explore it, retrieving patterns by pointing at them. This whole procedure targets the actual need for browsers specialized in music content, as drum patterns, allowing for complete collections to be visualized and explored, expanding the actual (limited) possibilities of music content browsing. The process of designing the *RhythmSpace* application was naturally complemented with the development of different algorithms for pattern-interpolation. These algorithms broaden the capabilities of a rhythm space as they allow to explore regions where no patterns are located, and to generate new patterns based on the surrounding ones. This feature turns a discrete rhythm space, capable of retrieving only the patterns in the collection, into a continuous space that generates new patterns, beyond the contents of the collection as it is browsed. RhythmSpace becomes then a tool that goes beyond organizing and visualizing a drum pattern collection, making possible the generation of new patterns in positions where the collection falls short of items.

The development of tools addressing stylistic EDM drum creation and variation can be successfully tackled, as we hope to have illustrated, by devising style-informed generative methods and variation processes. These tools extract drumming styles from examples and are then able to combine and transform style information in real-time in order to obtain new drum patterns. This style-based process offers a real-time, flexible and informed approach to drum generation, expanding the current compositional possibilities and shortening the gap between a simple rhythmic idea and the construction of a complete *drum track*.

The stylistic properties of the generative system devised were evaluated in two different subject-based experiments. The results of both experiments evidence how the system is capable of generating drum patterns, in a specific style, which can can be perceived as "integrated in" or "homogeneous with" those patterns in which the style is based. This quality is even appreciated when subjects, specialized in drum pattern composition, listen to patterns cerated in their own style, and confirm how the system is capable of “copying” it while being original. It can be concluded from these results that the system has capabilities for extracting a style from a collection of patterns and to use the the detected stylistic features to generate new patterns resembling those in the style.

5.1 Original Contributions

During this thesis, different original contributions were accomplished. A summary of the main contributions is presented below.

5.1.1 Algorithms

Several algorithms were designed and implemented in the course of this thesis:

- Symbolic rhythm descriptors analyzing 15 different features of polyphonic drum patterns.
- Agnostic density transformer (ADT) (section 4.2).
- Style extractor (section 4.3).
- Pattern generator in style (section 4.4.1).
- Style crossfader (section 4.4.2.1).

- Style commonness (section 4.4.2.2).
- Pattern variation algorithm (section 4.4.2.3).
- Pattern crossfader STENA (section 4.8.2.1).
- Pattern crossfader STEVO (section 4.8.2.2).
- Pattern crossfader STEVO-D (section 4.8.2.3).
- Rhythm space organizer (section 4.9).

5.1.2 Corpora

- Alf Gabrielsson's Experiment 1. 5 Ace Tone FR-3 drum patterns in symbolic format and their 2D coordinates after MDS from subject ratings (section 3.5.1.1).
- Alf Gabrielsson's Experiment 2. 6 Ace Tone FR-3 drum patterns and their 3D coordinates after MDS (section 3.5.1.1).
- EDM patterns. 75 techno, house and breakbeat patterns in symbolic format (section 3.6.1.2).
- EDM experiment. 9 EDM Patterns with their 2D coordinates from subject ratings after MDS. 36 different subject similarity ratings (section 3.6.2).
- Monophonic rhythm similarity experiment. 36 patterns in symbolic format plus 21 subjective distance ratings when the beat is induced and when it is not (section 3.2)
- DrDrums experimental data. 10 original drum patterns from EDM producers Boska and Sano. 10 generated patterns by DrDrums in Boska and Sano style. 20 subject ratings of how much original and generated patterns belong to each producer's style (section 4.6).

5.1.3 Applications

- *DrDrums* (section 4.5).
- *RhythmSpace* (section 4.8).

5.1.4 Experimental methods

- Monophonic rhythm experiment stage 1 (beat not induced) (section 3.2).
- Monophonic rhythm experiment stage 2 (beat induced) (section 3.2).
- Alf Gabrielsson's symbolic descriptors experiment (section 3.5).
- EDM polyphonic similarity experiment (section 3.6).
- Dr Drums quantitative experiment (section 4.6).
- Dr Drums qualitative experiment (section 4.7).

5.1.5 Software Libraries

The *RhythmToolbox* library was created in Python to process rhythmic patterns in symbolic format.

5.2 Further Research

Further research is needed in different complementary aspects of the ones presented in this thesis. From the rhythm experiments presented in Chapter 3, there are two principal results which can be followed further in order to gain deeper comprehension of rhythm processing. The first line of research, susceptible of being expanded, is that of *awareness* or the distinctive importance of the intra-pulse patterns when evaluating their similarity. This idea is explored in section 3.3 where the importance of the first intra-pulse patterns is observed to be of a higher magnitude than that of the rest of the intra-pulse patterns when predicting subject-rated similarity. This effect might reflect the well documented (in short-term memory research) “primacy” effect where the first items are best remembered and thus they could be primarily used in the similarity comparison. Despite this plausible explanation, this effect can be further researched in order to establish its role in human similarity sensations and thus help improve current algorithms used for similarity prediction.

The second line of future could be to expand the current research on polyphonic drum similarity to new human-based rhythm spaces. The small imprecisions found between the forecasted rhythm spaces and those derived from human ratings suggest the current method can be improved, specially if new similarity relations between the patterns in the stimuli collection are considered. The two human-based rhythm spaces used during this research are considered to be spanning over a different similarity range (see section 3.6.3), Gabrielsson's patterns are a very diverse collection, while the patterns used in our EDM space are less diverse as they all belong to the same musical genre. Further research could be addressed towards exploring the predictability of rhythm spaces of even smaller diversity, such as the drum patterns involved in an EDM song. Such results, combined with the ones presented in this thesis, could help increase the use-range and robustness of algorithms used to predict polyphonic rhythm similarity.

There is also additional research that can be undertaken to gain comprehension of the generative system presented in Chapter 4. Although the generative properties in a single style have been confirmed experimentally, the style crossfading capabilities are still susceptible of being improved. Questions regarding the effectiveness of a crossfade between two styles at different proportions can be approached experimentally in order to understand the aesthetic implications of using a system such as Dr-Drums. Also, evaluating stylistic generation based on different types of collections, as complete EDM songs or even albums can be researched to gain additional insight on the capabilities and limits of a system as Dr-Drums. Finally, the practical limitations of the system implemented in this thesis, as the absence of velocity or dynamics, and the reduced set of percussive instruments used, can be narrowed and explored much further.

5.3 Closing Remarks

Despite the tremendous amount of effort needed to complete this thesis, there is a general sensation of achievement and completion. This is due to the logical and progressive process of researching a topic, finding its main issues, designing and carrying out experiments, interpreting the results and, based on those interpretations, building useful systems that finally solve these issues. The topic of EDM has proven to be a fertile ground for research. The facts that EDM has a continuously-growing amount of global adepts, and that it is still evolving aesthetically and technologically, manifesting its own ways and idiosyncrasies, surely make this thesis one of many more studies to come focused in electronic dance music. As it was revealed during this research, the topic of rhythm is a very strong node, around which many different scientific and artistic disciplines articulate. I, optimistically, hope that the work presented throughout this thesis can serve as basis for further research in EDM and a small step towards the comprehension of musical rhythm.

APPENDIX A: LIST OF PUBLICATIONS

These are the papers published during the PhD thesis process.

Gómez-Marín, D., Jordà, S., Herrera, P. (2017) Drum rhythm spaces: from global models to style-specifc maps. 13th International Symposium on Computer Music Multidisciplinary Research (CMMR) Porto 25-28th September 2017.

Gómez-Marín, D., Jordà, S., Herrera, P. (2016) *Strictly Rhythm: Exploring the Effects of Identical Regions and Meter Induction in Rhythmic Similarity Perception*. In Music, Mind, and Embodiment.

Gómez-Marín, D., Herrera, P., Jordà, S. (2016) *Drumming with style: From user needs to a working prototype*. International Conference on New Interfaces for Musical Expression (NIME) Griffith, 11-15th July 2016

Gómez-Marín, D., Jordà, S., Herrera, P. (2016) *Rhythm Spaces*. 4th International Workshop on Musical Metacreation (MUME 2016) June 27th, 2016

Gómez-Marín, D., Jordà, S., Herrera, P. (2015) *PAD and SAD: Two Awareness-weighted Rhythmic Similarity Distances*. 16th International Society for Music Information Retrieval Conference. Málaga 26th-30th October 2015.

Gómez-Marín, D., Jordà, S., Herrera, P. (2015) *Evaluating rhythm similarity distances: The effect of inducing the beat.* Rhythm production and perception workshop. Amsterdam 6-8 July 2015

Gómez-Marín, D., Jordà, S., Herrera, P. (2015) *Strictly Rhythm: Exploring the effects of identical regions and meter induction in rhythmic similarity perception*. 11th International Symposium on Computer Music Multidisciplinary Research (CMMR) Music, Mind, and Embodiment. Plymouth, 16-19 June 2015.

APPENDIX B: GLOSSARY

There are specific terms adopted for this thesis that refer to particular elements of symbolic percussive rhythms (see Figure B.1) and mathematical functions. These terms are defined below in order to improve the reading experience.

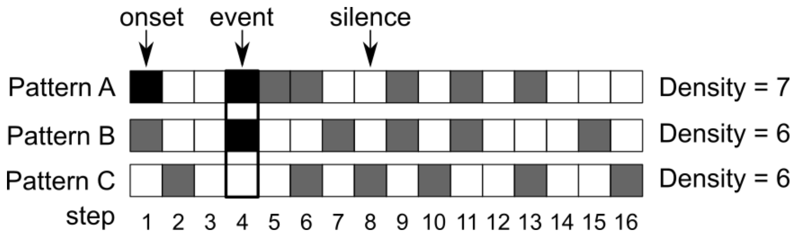

Figure B.1. Terms adopted to describe drum patterns. A polyphonic pattern of 16 steps is composed of three monophonic patterns A, B and C. Each of them having densities 7, 6 and 6 respectively.

Step: A discrete time position in a sequence.

Onset: The single hit of a drum instrument at a given step.

Silence: The absence of an onset at a given step.

Monophonic Pattern: A sequence of onsets and silences of a specific length in steps.

Polyphonic pattern: A stacked arrangement of monophonic patterns of equal length in which the first step is aligned for all the patterns.

Event: A vertical combination of onsets at a given step in a polyphonic pattern.

Density: The sum of onsets in a monophonic pattern.

Discrete Probability Distribution Function (DPDF): Is a function used to select a value i from a probability distribution, using a random number as input. The DPDF is based on a cumulative distribution function, used to find the upper and lower boundaries of each element in a probability distribution.

$$i = f(i, r) = \sum_i P(x_i) < r \leq \sum_{i+1} P(x_{i+1})$$

Where r is a random number and i is the index of the element in the probability distribution that is larger than the value of the cumulative distribution $\sum_i P(x_i)$ and smaller than the value of the cumulative distribution $\sum_{i+1} P(x_{i+1})$.

As an example, given the probability distribution $pd = (x_1=0.1, x_2=0.3, x_3=0.4, x_4=0.2)$, and a random value $r = 0.6$. The *ith* value that satisfies $\sum_i P(x_i) < 0.6 \leq \sum_{i+1} P(x_{i+1})$ is $i=2$, as $\sum_2 P(x_2) = 0.1+0.3 = 0.4$ and $\sum_3 P(x_3)= 0.1+0.3+0.4 = 0.8$ and $0.4 < 0.6 \leq 0.8.$

BIBLIOGRAPHY

Adamo, Mike. *The breakbeat bible*. Hudson Limited, 2010.

Adkins, A. W. (1962). Heidegger and language. *Philosophy*, *37*(141), 229-237.

Agres, K., Forth, J., & Wiggins, G. A. (2016). Evaluation of musical creativity and musical metacreation systems. *Computers in Entertainment (CIE)*, *14*(3), 3.

Ames, C. (1989). The Markov process as a compositional model: A survey and tutorial. *Leonardo*, 175-187.

Andersen, K., & Knees, P. (2016, May). The dial: Exploring computational strangeness. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 1352-1358). ACM.

Anderson, C., & Eigenfeldt, A. (2011) "A New Analytical Method For The Musical Study of Electronica." Proceedings of the Electroacoustic Music Studies Conference, Sforzando! New York.

Anderson, C., Eigenfeldt, A., & Pasquier, P. (2013). The generative electronic dance music algorithmic system (GEDMAS). In *Proceedings of the Artificial Intelligence and Interactive Digital Entertainment (AIIDE'13) Conference*.

Arce-Lopera C., Salamanca J., Gomez D. (2017) An Elastic Interface for Artistic Composition and Performance. In: Rebelo F., Soares M. (eds) Advances in Ergonomics in Design. AHFE 2017. Advances in Intelligent Systems and Computing, vol 588. Springer, Cham.

Argamon, S., Burns, K., & Dubnov, S. (Eds.). (2010). *The structure of style: algorithmic approaches to understanding manner and meaning*. Springer Science & Business Media.

Ariza, C. (2009). The interrogator as critic: The turing test and the evaluation of generative music systems. *Computer Music Journal*, *33*(2), 48-70.

Aucouturier, J. J., & Pachet, F. (2002). Music similarity measures: What's the use?. In ISMIR (pp. 13-17).

Aucouturier, J. J., & Pachet, F. (2005). Ringomatic: A Real-Time Interactive Drummer Using Constraint-Satisfaction and Drum Sound Descriptors. In *ISMIR* (pp. 412-419).

Baker, A. (Producer) & Dunn, A. (Director). (2015). *808* [motion picture]. United states: You Know Films.

Bardet, René-Pierre (1987) 260 Drum Machine Patterns. Hal Leonard Corporation, Milwaukee,Wi, USA

Bernardes, G., Guedes, C., & Pennycook, B. (2010). Style emulation of drum patterns by means of evolutionary methods and statistical analysis. In *Proceedings of the Sound and Music Conference* (pp. 1-4).

Bidder, S. (2001). Pump up the volume: a history of house [Television series]. London: BBC, Channel 4.

Blánquez, J., & Morera, O. (2002). Loops: Una historia de la música electrónica. Reservoir Books.

Boden, M. A. (1998). Creativity and artificial intelligence. *Artificial Intelligence*, *103*(1-2), 347-356.

Bolton, T. L. (1894). Rhythm. *The american journal of psychology*, *6*(2), 145-238.

Bouwer, F. L., Van Zuijen, T. L., & Honing, H. (2014). Beat processing is pre-attentive for metrically simple rhythms with clear accents: an ERP study. *PLoS One*, *9*(5), e97467.

Brewster, B., & Broughton, F. (2017). *Last night a DJ saved my life*. Le Castor Astral éditeur.

Brooks, F. P., Hopkins, A. L., Neumann, P. G., & Wright, W. V. (1957). An experiment in musical composition. *IRE Transactions on Electronic Computers*, (3), 175-182.

Brown, R. J., & Griese, M. (2000). *Electronica dance music programming secrets*. Prentice Hall.

Burton, A. R. (1998). *A hybrid neuro-genetic pattern evolution system applied to musical composition* (Doctoral dissertation, University of Surrey).

Butler, M. J. (2006). *Unlocking the groove: Rhythm, meter, and musical design in electronic dance music*. Indiana University Press.

Burger, B., London, J., Thompson, M. R., & Toiviainen, P. (2017). Synchronization to metrical levels in music depends on low-frequency spectral components and tempo. *Psychological Research*, 1-17.

Calopa, P. (2016) "Drums and Bass Interlocking." Master thesis, Universitat Pompeu Fabra, Barcelona.

Cao, E., Lotstein, M., & Johnson-Laird, P. N. (2014). Similarity and families of musical rhythms. *Music Perception: An Interdisciplinary Journal*, *31*(5), 444-469.

Chen, J. C., & Chen, A. L. (1998). Query by rhythm: An approach for song retrieval in music databases. In *Research Issues In Data Engineering, 1998.'Continuous-Media Databases and Applications'. Proceedings., Eighth International Workshop on* (pp. 139-146). IEEE.

Collins, Nick. (2001). Algorithmic Composition Methods for Breakbeat Science. Proceedings of Music Without Walls, De Montfort University, June 21-23, 2001.

Collins, Nick. (2002) The BBCut Library. Proceedings of the International Computer Music Conference, Gothenburg, Sweden.

Collins, N. (2008a). The analysis of generative music programs. *Organised Sound*, *13*(3), 237-248.

Collins, N. (2008b). Infno: Generating Synth Pop and Electronic Dance Music on demand. In *ICMC*.

Collins, N., Schedel, M., & Wilson, S. (2013). Electronic Music. Cambridge Introductions to Music (Cambridge, United Kingdom: Cambridge University Press, Cambridge, United Kingdom).

Conklin, D. (2003, April). Music generation from statistical models. In *Proceedings of the AISB 2003 Symposium on Artificial Intelligence and Creativity in the Arts and Sciences* (pp. 30-35). London: AISB Society.

Conklin, D., & Witten, I. H. (1995). Multiple viewpoint systems for music prediction. *Journal of New Music Research*, *24*(1), 51-73.

Cont, A., Dubnov, S., & Assayag, G. (2006). Anticipatory model of musical style imitation using collaborative and competitive reinforcement learning. In *Workshop on Anticipatory Behavior in Adaptive Learning Systems* (pp. 285-306). Springer, Berlin, Heidelberg.

Cook, P., Rouse, A., Wilson, M., & Reichmuth, C. (2013). A California sea lion (Zalophus californianus) can keep the beat: motor entrainment to rhythmic auditory stimuli in a non vocal mimic. *Journal of Comparative Psychology*, *127*(4), 412.

Cooper, G., & Meyer, L. B. (1963). *The rhythmic structure of music*. University of Chicago Press.

Cope, D. (2004). *Virtual music: computer synthesis of musical style*. MIT press.

Deliège, I. (2001). Similarity perception↔ categorization↔ cue abstraction.

Desain, P., & Honing, H. (1999). Computational models of beat induction: The rule-based approach. *Journal of new music research*, *28*(1), 29-42.

Desain, P., & Honing, H. (2003). The formation of rhythmic categories and metric priming. *Perception*, *32*(3), 341-365.

DeSantis, D. (2015). *Making Music: 74 Creative Strategies for Electronic Music Producers*. Ableton AG.

Dictionary, O. E. (2004). Oxford English dictionary online. *Mount Royal College Lib., Calgary*, *14*.

Dictionary, C. (2015). Cambridge dictionaries online.

Drake, C., Jones, M. R., & Baruch, C. (2000). The development of rhythmic attending in auditory sequences: attunement, referent period, focal attending. *Cognition*, *77*(3), 251-288.

Eck, D. (2001). A network of relaxation oscillators that finds downbeats in rhythms. In *International Conference on Artificial Neural Networks* (pp. 1239-1247). Springer, Berlin, Heidelberg.

Eigenfeldt, A., & Pasquier, P. (2013). Evolving structures for electronic dance music. In *Proceedings of the 15th annual conference on Genetic and evolutionary computation* (pp. 319-326). ACM.

Emmerson, S. (2013). *Living electronic music*. Ashgate Publishing, Ltd..

Ellis, D., & Arroyo, J. (2004, October). Eigenrhythms: Drum pattern basis sets for classification and generation. In *ISMIR* (Vol. 2004, p. 5th).

Eshun, K. (1998). *More brilliant than the sun: Adventures in sonic fiction*. Interlink Publishing Group Incorporated.

Faraldo, A., Gómez, E., Jordà, S., Herrera, P., (2016) Key estimation in Electronic Dance Music. Proceedings in ECIR 2016.

Fischer, C., Luaute, J., and Morlet, D. (2010). Event-related potentials (MMN and novelty P3) in permanent vegetative or minimally conscious states. Clin. Neurophysiol. 121, 1032–1042.

Fitch, W. T., & Rosenfeld, A. J. (2007). Perception and production of syncopated rhythms. *Music Perception: An Interdisciplinary Journal*, *25*(1), 43-58.

Fitch, W. (2013). Rhythmic cognition in humans and animals: distinguishing meter and pulse perception. *Frontiers in systems neuroscience*, *7*, 68.

Fujioka, T., Trainor, L. J., Large, E. W., & Ross, B. (2012). Internalized timing of isochronous sounds is represented in neuromagnetic beta oscillations. *Journal of Neuroscience*, *32*(5), 1791-1802.

Gabrielsson, A. (1973a). Similarity ratings and dimension analyses of auditory rhythm patterns. 1. *Scandinavian Journal of Psychology*, *14*(1), 138-160.

Gabrielsson, A. (1973b). Similarity ratings and dimension analyses of auditory rhythm patterns. II. *Scandinavian Journal of Psychology*, *14*(1), 161-176.

Gardiner, C. (2009). *Stochastic methods* (Vol. 4). Berlin: springer.

Gärdenfors, P. (2004). *Conceptual spaces: The geometry of thought*. MIT press.

Gasser, M., Eck, D., & Port, R. (1999). Meter as mechanism: A neural network model that learns metrical patterns. *Connection Science*, *11*(2), 187-216.

Goodwin, A. (2004). Rationalization and democratization in the new technologies of popular music. *Popular Music. Critical Concepts in Media and Cultural Studies*, *2*, 147-169.

Gouyon, F., Dixon, S., Pampalk, E., & Widmer, G. (2004, June). Evaluating rhythmic descriptors for musical genre classification. In *Proceedings of the AES 25th International Conference* (pp. 196-204).

Grahn, J.A., Brett, M., 2007. Rhythm and beat perception in motor areas of the brain. Journal of Cognitive Neuroscience. 19 (5), 893–906.

Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *the Journal of the Acoustical Society of America*, *61*(5), 1270-1277.

Hampton, J. A. (2001). The role of similarity in natural categorization. In U. Hahn & M. Ramscar (Eds.), *Similarity and categorization* (pp. 13-28).

Hedges, S. A. (1978). Dice music in the eighteenth century. *Music & Letters*, *59*(2), 180-187.

Hein, Ethan (2013). Designing the Drum Loop: A constructivist iOS rhythm tutorial system for beginners. Master Thesis, NYU

Hennig, H. (2014). Synchronization in human musical rhythms and mutually interacting complex systems. *Proceedings of the National Academy of Sciences*, *111*(36), 12974-12979.

Hewitt, M. (2009). *Composition for computer musicians*. Nelson Education.

Holden, J. (2015) James Holden: On Human Timing, 2015.

Hollins, M., Bensmaïa, S., Karlof, K., & Young, F. (2000). Individual differences in perceptual space for tactile textures: Evidence from multidimensional scaling. *Perception & Psychophysics*, *62*(8), 1534-1544.

Honing, H. (2002). Structure and interpretation of rhythm and timing. *Tijdschrift voor Muziektheorie*, *7*(3), 227-232.

Honing, H. (2012). Without it no music: beat induction as a fundamental musical trait. *Annals of the New York Academy of Sciences*, *1252*(1), 85-91.

Hove, M. J., Marie, C., Bruce, I. C., & Trainor, L. J. (2014). Superior time perception for lower musical pitch explains why bass-ranged instruments lay down musical rhythms. *Proceedings of the National Academy of Sciences*, *111*(28), 10383-10388.

Huang, A., & Wu, R. (2016). Deep learning for music. *arXiv preprint arXiv:1606.04930*.

Jaques, N., Gu, S., Turner, R. E., & Eck, D .(2016) Generating Music by Fine-Tuning Recurrent Neural Networks with Reinforcement Learning. Deep Reinforcement Learning Workshop, NIPS

Johanson, B., & Poli, R. (1998). *GP-music: An interactive genetic programming system for music generation with automated fitness raters* (pp. 181-186). University of Birmingham, Cognitive Science Research Centre.

Johnson-Laird, P. N. (1991). Rhythm and meter: A theory at the computational level. *Psychomusicology: A Journal of Research in Music Cognition*, *10*(2), 88.

Jones, M. R. (1976). Time, our lost dimension: toward a new theory of perception, attention, and memory. *Psychological review*, *83*(5), 323.

Jones, M. R. (1987). Dynamic pattern structure in music: Recent theory and research. *Perception & psychophysics*, *41*(6), 621-634.

Jones, M. R. (1990). Learning and the development of expectancies: An interactionist approach. *Psychomusicology: A Journal of Research in Music Cognition*, *9*(2), 193.

Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological review*, *96*(3), 459.

Keane, M. T., & Smyth, B. (2001). Dynamic similarity: A processing perspective on similarity.

Keim, D. A. (2002). Information visualization and visual data mining. *IEEE transactions on Visualization and Computer Graphics*, *8*(1), 1-8.

Kilchenmann, L., & Senn, O. (2015). Microtiming in swing and funk affects the body movement behavior of music expert listeners. *Frontiers* *in psychology*, *6*, 1232.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., & Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology*, *14*(4), R36.

Krumhansl, C. L. (2001). *Cognitive foundations of musical pitch*. Oxford University Press.

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, *29*(1), 1-27.

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological review*, *106*(1), 119.

Lee, D. T., & Schachter, B. J. (1980). Two algorithms for constructing a Delaunay triangulation. *International Journal of Computer & Information Sciences*, *9*(3), 219-242.

Lehmann, A., Arias, D. J., & Schönwiesner, M. (2016). Tracing the neural basis of auditory entrainment. *Neuroscience*, *337*, 306-314.

Lerdahl, F., & Jackendoff, R. (1985) *A generative theory of tonal music*. MIT press.

Logan, B., & Salomon, A. (2001). A Music Similarity Function Based on Signal Analysis. In ICME (pp. 22-25).

London, J. (2012). *Hearing in time: Psychological aspects of musical meter*. Oxford University Press.

Longuet-Higgins, H. C. (86). Lee, CS (1982). *The perception of musical rhythms*. *Percep-tion*, *11*, 115-128.

Longuet-Higgins, H. C., & Lee, C. S. (1984). The rhythmic interpretation of monophonic music. *Music Perception: An Interdisciplinary Journal*, *1*(4), 424-441.

Maurer, D., & Mondloch, C. J. (2004). Neonatal synaesthesia: A reevaluation. In L. Robertson (Ed.), Synaesthesia: Perspectives from cognitive neuroscience (pp. 193-213). New York: Oxford University Press.

Meumann E. (1894) Untersuchungen zur Psychologie und Asthetik des Rhythmus. Philosophische Studien 10:249-323,393-431

Merchant, H., Grahn, J., Trainor, L., Rohrmeier, M., & Fitch, W. T. (2015). Finding the beat: a neural perspective across humans and non-human primates. *Phil. Trans. R. Soc. B*, *370*(1664), 20140093.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *Journal of experimental psychology*, *90*(2), 227.

Meyer, L. B. (1956) *Emotion and meaning in music*. University of chicago Press.

Meyer, L. B. (2010). *Music, the arts, and ideas: Patterns and predictions in twentieth-century culture*. University of Chicago Press.

Mills, S. & Menagh, M. (1990, June). Rock stars score the future, OMNI 42 – 47, 80-82.

Moore, A. F. (2001). Categorical conventions in music discourse: Style and genre. *Music & letters*, *82*(3), 432-442.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clinical neurophysiology*, *118*(12), 2544-2590.

Nierhaus, G. (2009). *Algorithmic composition: paradigms of automated music generation*. Springer Science & Business Media.

Nozaradan, S., Peretz, I., Missal, M., & Mouraux, A. (2011). Tagging the neuronal entrainment to beat and meter. *Journal of Neuroscience*, *31*(28), 10234-10240.

Nozaradan, S., Peretz, I., & Mouraux, A. (2012). Selective neuronal entrainment to the beat and meter embedded in a musical rhythm. *Journal of Neuroscience*, *32*(49), 17572-17581.

Pachet, F. (2002). Playing with virtual musicians: The continuator in practice. *IEEE MultiMedia*, *9*(3), 77-82.

Pachet, F. (2003). The continuator: Musical interaction with style. *Journal of New Music Research*, *32*(3), 333-341.

Pachet, F. (2006). 19 Enhancing individual creativity with interactive musical reflexive systems. *Musical creativity*, 359.

Pachet, F., & Roy, P. (2011). Markov constraints: steerable generation of Markov sequences. *Constraints*, *16*(2), 148-172.

Paiement, J. F., Grandvalet, Y., Bengio, S., & Eck, D. (2007). A generative model for rhythms. In *NIPS Workshop on Brain, Music and Cognition* (No. LIDIAP-CONF-2007-035).

Palmer, C., & Krumhansl, C. L. (1990). Mental representations for musical meter. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(4), 728.

Pampalk, E., Flexer, A., & Widmer, G. (2005). Improvements of Audio-Based Music Similarity and Genre Classificaton. In ISMIR (Vol. 5, pp. 634-637).

Parncutt, R. (1994). A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Perception: An Interdisciplinary Journal*, *11*(4), 409-464.

Patel, A. D., & Daniele, J. R. (2003). An empirical comparison of rhythm in language and music. *Cognition*, *87*(1), B35-B45.

Patel A. D., Iversen J. R., Bregman M. R., Schulz I. (2009). Experimental evidence for synchronization to a musical beat in a nonhuman animal. Curr. Biol. 19, 827–830 10.1016/j.cub.2009.03.038

Paulus, J., & Klapuri, A. (2002). Measuring the similarity of Rhythmic Patterns. In *ISMIR*.

Pearce, M., & Wiggins, G. (2004). Improved methods for statistical modelling of monophonic music. *Journal of New Music Research*, *33*(4), 367-385.

Pearce, M., Conklin, D., & Wiggins, G. (2004). Methods for combining statistical models of music. In *International Symposium on Computer Music Modeling and Retrieval* (pp. 295-312). Springer, Berlin, Heidelberg.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical neurophysiology*, *118*(10), 2128-2148.

Post, O., & Toussaint, G. (2011). The edit distance as a measure of perceived rhythmic similarity.

Pouthas, V. (1995) The development of the perception of time and temporal regulation of action in infants and children. In I. Deliège and J. A. Sloboda (Eds) Musical Beginnings: The Origins and Development of Musical Competence. New York: Oxford University Press, pp. 115–41.

Puckette, M. (1996). Pure Data: another integrated computer music environment. *Proceedings of the second intercollege computer music concerts*, 37-41.

Ramscar, M., & Hahn, U. (Eds.). (2001). *Similarity and categorization*. Oxford University Press.

Reynolds, S. (2013). *Energy flash: A journey through rave music and dance culture*. Faber & Faber.

Rocamora, M., Jure, L., & Biscainho, L. W. (2014). Tools for detection and classification of piano drum patterns from candombe recordings. In *Proc. of the 9th Conference on Interdisciplinary Musicology (CIM14)* (pp. 382-387).

Roldán, A. (1930) Rítmica No. 5: for Percussion Instruments. New York City: Southern Music.

Schuback, M. S. C. (2003). The Poetics of Language: Readings of Heidegger's On the Way to Language. In *Metaphysics, Facticity, Interpretation* (pp. 195-215). Springer, Dordrecht.

Shannon, C. E. (2001). A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, *5*(1), 3-55.

Shepard, R. N. (1962). The analysis of proximities: multidimensional scaling with an unknown distance function. I. *Psychometrika*, *27*(2), 125-140.

Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of mathematical psychology*, *1*(1), 54-87.

Snoman, R. (2012). *The dance music manual: tools, toys and techniques*. CRC Press.

Spectrasonics. Stylus RMX, 2005.

Steedman, M. J. A generative grammar for jazz chord sequences. Music Perception, pages 52–77, 1984.

Technemedia. Different Drummer, 2015.

Temperley, D. (2007). *Music and probability*. Mit Press.

Théberge, P. (1997). *Any sound you can imagine: Making music/consuming technology*. Wesleyan University Press.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267-288.

Tidemann, A., Öztürk, P., & Demiris, Y. (2009). A groovy virtual drumming agent. In *International Workshop on Intelligent Virtual Agents* (pp. 104-117). Springer, Berlin, Heidelberg.

Toussaint, G. T. (2005). The Euclidean algorithm generates traditional musical rhythms. In *Proceedings of BRIDGES: Mathematical Connections in Art, Music and Science* (pp. 47-56).

Toussaint, G. T. (2004). A Comparison of Rhythmic Similarity Measures. In *ISMIR*.

Toussaint, G. T., Campbell, M., & Brown, N. (2011). Computational models of symbolic rhythm similarity: Correlation with human judgments. *Analytical Approaches to World Music*, *1*(2), 380-430.

Toussaint, G. T., & Oh, S. M. (2016). Measuring musical rhythm similarity: Edit distance versus minimum-weight many-to-many matchings. In *Proceedings on the International Conference on Artificial Intelligence (ICAI)* (p. 186). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp).

Tulving, E., & Craik, F. I. (Eds.). (2000). *The Oxford handbook of memory*. Oxford: Oxford University Press.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind 59*(236), 433.

Turquois, C., Hermant, M., Gómez-Marín, D., & Jordà, S. (2016, March). Exploring the Benefits of 2D Visualizations for Drum Samples Retrieval. In *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval* (pp. 329-332). ACM.

Tversky, A. (1977). Features of similarity. *Psychological review*, *84*(4), 327.

Urtubia, H. (2015) Robotic Drums.

Varèse, E. (1930) Ionisation : for Percussion Ensemble of 13 Players. Milano: Ricordi.

Verguts, T., Ameel, E., & Storms, G. (2004). Measures of similarity in models of categorization. *Memory & Cognition*, *32*(3), 379-389.

Vogel, C. (2015) Rhythmic Computation Lab.

Weaver, W. (1953). Recent contributions to the mathematical theory of communication. *ETC: a review of general semantics*, 261-281.

Witek, M. A., Clarke, E. F., Wallentin, M., Kringelbach, M. L., & Vuust, P. (2014a). Syncopation, body-movement and pleasure in groove music. *PloS one*, *9*(4), e94446.

Witek, M. A., Clarke, E. F., Kringelbach, M. L., & Vuust, P. (2014b). Effects of polyphonic context, instrumentation, and metrical location on syncopation in music. *Music Perception: An Interdisciplinary Journal*, *32*(2), 201-217.

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and intelligent laboratory systems*, *2*(1-3), 37-52.

Wooller, R., & Brown, A. R. (2011). Note sequence morphing algorithms for performance of electronic dance music. *Digital Creativity*, *22*(1), 13-25.

Zicarelli., D. M and Jam Factory. Computer Music journal, dec(1):13 – 29, 1987.