Drum rhythm spaces: From polyphonic similarity to generative maps

Daniel Gómez-Marín , Sergi Jordà & Perfecto Herrera

Published online: 24 Aug 2020.

Submit your article to this journal ↗

Article views: 68

View related articles ↗

View Crossmark data ↗

Routledge
Taylor & Francis Group

Check for updates

Drum rhythm spaces: From polyphonic similarity to generative maps

Daniel Gómez-Marín [a], Sergi Jordà [b] and Perfecto Herrera [b]

[a]Faculty of Engineering, Universidad ICESI, Cali, Colombia; [b]Music Technology Group (MTG), Universitat Pompeu Fabra, Barcelona, Spain

ABSTRACT
This paper reports on the design and evaluation of drum rhythm spaces as interactive bi-dimensional maps used for the visualisation, retrieval and generation of drum patterns. We carry out two experiments exploring human processing of polyphonic drum patterns concluding with a list of descriptors that significantly influence similarity sensations. These features are used to build spaces based on drum pattern collections, where patterns are organised by similarity, modelled according to human perception. A drum-interpolation algorithm is introduced (and evaluated) to enhance rhythm space functionality by means of patterns that bound it, converting a discrete space to a continuous generative one.

1. Introduction

Rhythm, considered one of the main dimensions of music, appears when events are detected, compared and tracked for their repetition or variation across time. Studies on music cognition have shown that rhythm sensations are not just the result of processing objective information extracted from the physical data. It is in the combination of such data and processes of making sense of them (by means of entrainment, anticipation, previous knowledge, and hypotheses testing), that pulses, tempi and metric hierarchies 'emerge' (Clark, 2013; London, 2012; van der Weij et al., 2017). When human listeners and performers try to encode rhythmic information, pattern comparison becomes a powerful strategy to aid decisions on possible critical features, attention resources allocation, or preparation for the 'next' moment. Deciding on the likeness or the contrast between just processed patterns or between a recent one and some retrieved from memory is crucial. As we can judge how similar two scenes, melodies, or colours are, we can do the same for rhythm patterns. The mechanisms operating in rhythm similarity decisions have not been fully characterised yet, specially for the case of polyphonic rhythms, which involve many layers of sounds and events. That knowledge is fundamental for the sake of expanding our comprehension of rhythm cognition in humans, and it can also be profitable in music creation contexts by embedding rhythmic intelligence inside music composition tools. The research reported here follows this path

that, from music theory and music cognition, leads to music composition.

One of the techniques to study the way humans organise their memories and their pattern processing is by explicitly asking them about the degree of similarity between stimuli. Typically, a group of subjects assesses the similarity between each pair of elements within a collection of stimuli and, after analysing these results, new hypotheses or models (usually referred to as 'conceptual spaces', 'perceptual spaces' or 'mental maps') can be proposed to explain the mechanisms and relevant features behind the resulting similarity relations (Gärdenfors, 2000). Such models, although descriptive in nature, can also generate predictions about human similarity ratings for new, previously unseen, patterns.

This paper presents the research towards a model for understanding polyphonic rhythm similarity and its implementation in music composition tools.[1] Our research is based on experiments on similarity ratings between polyphonic drum patterns. Most of the patterns used here are characteristic of electronic dance music (EDM), a term referring to music created with electronic instruments and specifically designed for inciting people to dance. EDM is an umbrella term where different sub-genres avidly coexist, most of them differentiated by rhythm, timbre and the type of technology used to generate and process sounds (Collins et al., 2013; Reynolds, 2012).

The article is organised in the following sections: first we provide some definitions and research context on perceptual and conceptual spaces; then we focus on rhythm spaces and rhythm similarity. A discussion of potentially relevant features to compute polyphonic rhythmic similarity is followed by their validation using existing data and literature. We then address the characterisation of rhythm spaces for EDM sub-genres by means of an experiment with human listeners. Our results show the general capabilities of our proposed features to capture rhythm similarity relations. Once we have created reliable and robust rhythm spaces, we further stretch their capabilities by using them as a music generation tool. A final experiment validating the rhythm space generative algorithm is presented. Our discussions and observations provide the expected closure to the paper.

2. Perceptual spaces

Perceptual spaces have a scientific tradition for representing human knowledge. There are examples in many domains such as timbre (Grey, 1977), colour (Shepard, 1964), wine (Ballester et al., 2008), texture (Hollins et al., 2000), or even Dutch shirts! (Zwarts, 2015). Spaces derived from subject-based research reveal how the semantic memory for those domains could be organised. Such spaces are used to geometrically understand relationships between their elements. In these spaces, the concept of similarity becomes analogous to that of distance, where a small separation between elements suggests the perceptual closeness of such elements, and thus a high degree of similarity (Gärdenfors, 2000). In a perceptual space, each of its spanning dimensions represents a fundamental feature for the apprehension of the domain, a characteristic that can be measured and quantified in all the elements of the domain. The degree in which a characteristic is found determines its position in the given dimension. Thus, perceptual spaces can be composed of several dimensions, all of them revealing essential characteristics for the human cognition of a domain.

2.1. Spaces in sound and music

The use of conceptual spaces as a framework to model different music-related domains has led to the creation of new knowledge. Conceptual spaces have been used in musical domains such as timbre (Grey, 1977) and tonality (Krumhansl, 1979), where geometric models of how these domains could be organised in human minds are presented. For example, Grey (1977) found that human perception of timbre similarity could be characterised using three main dimensions: the spectral fluctuation, the

centroid of the spectrum and the attack time. This geometric model of timbre supports a framework that can be used to understand essential aspects of timbre cognition and it has been further explored by other researchers (Halpern et al., 2004; Hourdin et al., 1997; McAdams et al., 1995; Yee-King, 2011), but it also allows, by quantifying certain objective properties of those sounds, the prediction of how similar a group of sounds will be perceived by subjects. The structure of the domain-specific space (e.g. a timbre space) creates relations among elements that act as a human-optimized interface into the domain. That is, a timbre space can be used as an interface for controlling timbre. For example, Grey's timbre spaces were the foundation for a new research field on spatial timbre interaction, led by Wessel (1979) and further continued by other researchers (Graham et al., 2017; Turquois et al., 2016). Conceptual spaces proved to be robust enough to be used not only for organisation and visualisation, but also for retrieval of the musical elements (sounds in the case of timbre) (Einbond et al., 2009).

3. Rhythm spaces

The experiments of Gabrielsson (1973a, 1973b) are one of the first and most important precedents for establishing experimental procedures and providing results on polyphonic rhythm similarity. One of his main contributions is the development of conceptual rhythm spaces, where he locates polyphonic drum patterns in bi-dimensional and three-dimensional spaces as the result of subject-based similarity experiments.

3.1. Rhythm similarity

The first clues on how rhythmic similarity is processed can be found in the simpler case of monophonic rhythm. In this area there are two parallel approaches to derive similarity metrics, one is the use of rhythm-agnostic information-based metrics (Post & Toussaint, 2011; Toussaint, 2004), and the other involves using cognitive and perceptual knowledge of rhythm processing (Cao et al., 2014; Johnson-Laird, 1991).

The approach for measuring monophonic rhythm similarity using information-based metrics is based on simple algorithms. The edit distance, for example, has been reported to be correlated with monophonic similarity assessment (Guastavino et al., 2009; Post & Toussaint, 2011; Toussaint et al., 2011), and its algorithm is based on measuring the number of transformations needed for a string to become another one (by means of insertion, deletion and addition of characters). This metric, although straightforward to measure, ignores important
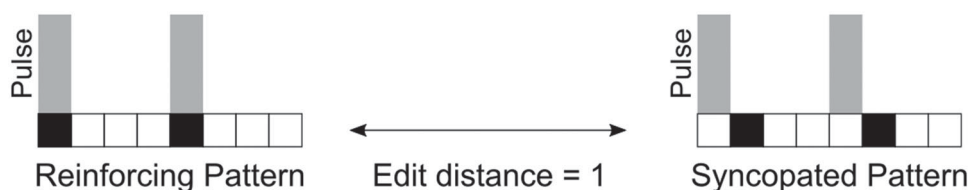
Figure 1. A pulse-reinforcing pattern and a highly syncopated pattern at small edit distance of 1. The pulse is represented by a grey bar and onsets by black squares.

cognitive aspects, allowing two patterns that can be perceived as very different rhythmically to have a very small edit distance value (Figure 1). The disentanglement of these rhythm-agnostic metrics from the basic concepts of rhythm cognition, namely pulse, meter (London, 2012), metric salience (Palmer & Krumhansl, 1990) and syncopation (Longuet-Higgins & Lee, 1984) is clearly noted by Paiement et al. (2007), suggesting they are not the ideal candidates to be expanded to measure similarity in polyphonic scenarios.

The other approach, cognition-based monophonic similarity metrics, is based upon current scientific knowledge on human rhythm processing. One of these metrics is that of *rhythm families* proposed by Cao et al. (2014), where the concepts of syncopation and *identical regions* are at play. Syncopation families arise as the result of classifying intra-pulse sub-patterns into three categories defined by their relation to the pulse (pulse reinforcement: R, syncopation: S or nothing: N). The concept of *identical regions* refers to sub-patterns which are common to two different monophonic patterns, but shifted in position (see Figure 2). Cao et al. report that both concepts (syncopation families and identical regions) influence similarity sensations in subjects. The existence of identical regions shifted in position, and the presence of same syncopation families in the same position, maximises the similarity sensations between two monophonic patterns.
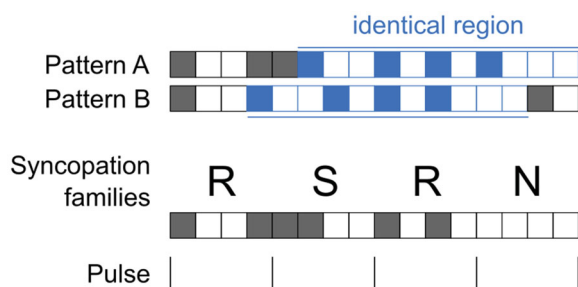
Figure 2. Identical regions and syncopation families as described by Cao et al. (2014) influence the similarity sensation when two monophonic patterns are compared. Syncopation families are analysed within pulses and can be classified as R: pulse reinforcement, S: syncopation, N: nothing.

In experiments previously reported (Gómez-Marín et al., 2015a, 2015b), syncopation families have proven useful for predicting subjective similarity sensations specifically in the case when the pulse is induced. That is, in a pulse-induced scenario (as most rhythmic music is experienced), measures based on syncopation are useful for predicting similarity sensations. Syncopation, and thus knowledge on rhythm processing, has proven fundamental for the prediction of monophonic similarity. We suggest that an extension of this approach using these rhythmic fundamentals can be useful as part of a model to measure polyphonic similarity.

Other advances have also been made in the direction of understanding how rhythm polyphony is processed by humans. For the specific case of polyphonic drum arrangements, it is proposed that the frequency range of the percussive instruments influence how the complete polyphonic stream is processed. Specifically, the patterns performed by instruments with a low-frequency range (i.e. kick drum) seem to have a higher importance in the induction or disruption of the pulse than instruments in a high-frequency range such as the hi-hats (Bouwer et al., 2014; Burger et al., 2017; Hove et al., 2014; Witek, Clarke, Kringelbach, et al., 2014). These observations have been used by researchers to propose a metric for measuring syncopation in polyphonic drum patterns. This polyphonic syncopation metric is used to find how a medium dose of syncopation maximises pleasure and desire to dance when listening to drum patterns (Witek, Clarke, Wallentin, et al., 2014).

As we already mentioned at the beginning of this section, another seminal reference in polyphonic similarity is the research carried out by Gabrielsson (1973a, 1973b) who performed several experiments in polyphonic similarity judgments.

These three sources, the polyphonic syncopation measure, the relevance of the frequency-range of percussive instruments (Bouwer et al., 2014; Burger et al., 2017; Hove et al., 2014; Witek, Clarke, Kringelbach, et al., 2014) plus Gabrielsson's experiments constitute the foundation of the novel method for measuring polyphonic similarity presented in this paper.

3.2. Spaces for visualising rhythm

As visualisation techniques were used to explore different relations between percussive rhythms, other rhythm research involving spaces has emerged. Desain and Honing have an extensive body of work on modelling human rhythm perception from a cognitive perspective. In several papers they use a three dimensional space for visualising rhythms. Each axis of the space represents one of the three inter-onset intervals (IOI) which exist between the four notes of their rhythms. In this informative space, a rhythmic structure is recognised by its position (Desain & Honing, 1999, 2003). Other authors have dealt with rhythm spaces and rhythm similarity in a polyphonic music audio (retrieval) context. Rhythm spaces are present in many MIR studies involving rhythm descriptors (Ellis & Arroyo, 2004; Makris et al., 2017; Paulus & Klapuri, 2002; Rocamora et al., 2014). Here, spaces are rarely depicted or used as such, probably because of their high dimensionality and because the aims lean more towards automatic music classification (Chen & Chen, 1998).

3.3. Rhythm in electronic dance music

As we mentioned above, the context of our research is electronic dance music (EDM). EDM evolves from African American dance music genres as disco and funk, as they were reinterpreted using a new generation of affordable synthesisers and drum machines released during the 1980s (Reynolds, 2012; Sicko, 2010). One of the main musical elements through which EDM rhythm sophistication is expressed is drumming, which constitutes the backbone of this type of music (Butler, 2006). Drum arrangements in EDM, also called *drum sequences* or *drum tracks*, are typically presented as the progressive concatenation of different polyphonic patterns of short duration (one or two bars), which are the focus of minute detail from music producers. In the context of EDM, knowledge of polyphonic drum similarity becomes of main relevance, and the rhythmic essence of this music becomes an excellent framework in which to do research on polyphonic rhythm cognition.

4. Features for polyphonic rhythm similarity computation

In order to understand the mechanisms underlying human processing of polyphonic rhythm, three different sources of knowledge are here revised and integrated into one main research methodology: the experiments of Gabrielsson (1973b), contemporary experiments on polyphonic processing of rhythms (Bouwer et al., 2014;

Burger et al., 2017; Hove et al., 2014; Witek, Clarke, Kringelbach, et al., 2014) and our previous research on syncopation and similarity in polyphonic scenarios. One of Gabrielsson's main contributions is presenting the results of his similarity experiments as rhythm spaces. Gabrielsson asked subjects to rate the similarity between different rhythms, obtained low dimensional spaces from those ratings. According to his research, rhythm similarity judgments are influenced by:

- 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.

The second source of knowledge that we consider is recent research on how humans process polyphonic rhythms Bouwer et al. (2014); Burger et al. (2017); Hove et al. (2014); Witek, Clarke, Kringelbach, et al. (2014). In these studies, the prominence of the main frequency of the different instruments of a polyphonic drum pattern is reported to influence listeners' rhythm processing. The predominant frequency of a drum sound can affect its power to disturb or to confirm 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 than high-frequency instruments (i.e. the hi-hats)in the establishment of a pulse or in its alteration (e.g. a syncopation). This view is experimentally backed by Witek, Clarke, Kringelbach, et al. (2014), who devise a new 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 then successfully used to study the impact of syncopation in the desire to dance. Finally, as a third knowledge source for guiding our research, the importance of syncopation is aligned with our previous results (Gómez-Marín et al., 2015a, 2015b), where it was reported how the syncopations present in a monophonic rhythm are a useful source of pattern differentiation, influencing subjects when assessing the similarity between two rhythmic patterns.

Some conclusions can be drawn from incorporating these sources of knowledge into a unified perspective on polyphonic rhythm similarity. There is a clear relevance of syncopation and meter in the processing of monophonic and polyphonic patterns that is transversal to the three different research approaches presented above. A differentiation between three frequency ranges for percussive sounds (low, mid and high) is also useful, 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 between sources while avoiding frequency overlapping. 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 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 to capture the different qualitative factors mentioned (frequency ranges, syncopation, density, number of instruments and the simplicity/complexity of the patterns), by means of simple and straightforward algorithms.

Our focus on symbolic patterns resides in the need to discard the effect of timbre in subjective similarity measurements, as other researchers (Gabrielsson included) remarked. The use of consistent timbres when rendering percussive patterns allows the subjects to just focus on the symbolic interpretation of rhythm when undertaking experiments. Thus, we have devised a new set of symbolic rhythm descriptors, that differ from others used in automatic rhythm classification research (for example Gouyon et al., 2004; Paulus & Klapuri, 2002) 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. We use the MIDI (Musical Instrument Digital Interface) protocol to represent drum patterns, coding instruments, notes, durations and dynamics in symbolic format.

4.1. Symbolic drum pattern descriptors

Herewe will adapt some ideas reviewed above, where the factors influencing polyphonic similarity sensations are related to metrical weight[2] 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 by mapping the General MIDI Level 1 Percussion Key Map1 (GMPKM)[3] to three instrument categories (low, mid and high),[4] based on the typical spectral centre of each sound (i.e. a low tom belongs to low frequency, a snare and a clap to the mid frequency instruments and all cymbals to

the high 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: low, mid and high. This procedure of using only three streams is an adaptation of the methodology used by Witek, Clarke, Kringelbach, et al. (2014) backed in experiments by Hove et al. (2014), Bouwer et al. (2014), Witek, Clarke, Kringelbach, et al. (2014) and Burger et al. (2017). It also resonates with the Auditory Scene Analysis theory (Bregman, 1994) in which the multiple and concurrent data generated during the parallel analysis processes in the auditory pathways and auditory cortex are simplified into a small number of auditory streams.

Once a symbolic drum pattern is converted into a combination of three band-wise patterns, these patterns are characterised according to different factors pointed out by Gabrielsson to influence similarity at a polyphonic level: syncopations, densities, number of instruments, meter, and the simplicity-complexity of the patterns. The crossover between the three frequency streams (low, mid and high) and the sources of information (presented in Table 1) encouraged us to define equations for these concepts, which can be extracted from symbolic drum patterns. Gabrielsson articulated his ideas just verbally, and we here attempt to turn some of them into MIDI-based computable features. 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 using the mapping presented in the complementary material. The descriptors are quantified as presented in the following subsections.

4.1.1. Number of instruments (noi)
This is the simplest metric to compute as it is just the amount of different instruments in the symbolic polyphonic pattern.

4.1.2. Hisync, midsync, and losync
Syncopations are quantified following Longuet-Higgins and Lee (1984), who propose a method based on a nested metric profile similar to the one presented by Lerdahl and Jackendoff (1985), where onsets coinciding with higher divisions of a pattern obtain higher metrical values. For each stream (low, mid and high), a metrical value is extracted when an onset is followed by a silence to compute a syncopation value for each stream. These values

---

[2] Lerdahl and Jackendoff (1985) propose a hierarchy of accents within a rhythmic pattern, where onsets coinciding with higher divisions of the pattern obtain higher metrical values.

[3] The MIDI association provides a list of percussion instruments and their suggested MIDI note values: https://www.midi.org/specifications-old/item/gm-level-1-sound-set.

[4] The complete mapping list is at https://github.com/danielgomezmarin/rhythmtoolbox/blob/master/MIDI-mapping.md.

Table 1. This table presents our descriptors resulting from combining the three different polyphonic research contexts: Gabrielsson's conclusions (the five columns), three different frequency layers (three rows: low, mid and high) and syncopation (second and third columns).

Frequency	Gabrielsson's factors					
	Syncopation		Density		Instruments	Complexity
High	polysync	Hisync	hiD, hiness	stepD	noi	hisyness
Mid	midsync	midD, midness	midsyness			
Low	losync	lowD, lowness	losyness			
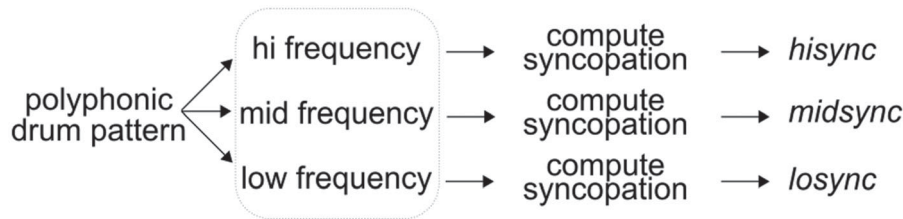
Figure 3. Block diagram description of *hisync*, *midsync* and *losync* descriptors computation.
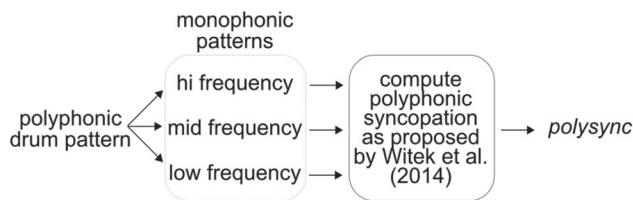
Figure 4. Block diagram description of *polysync* descriptor computation. Three different monophonic patterns (low, mid and high) are extracted form the polyphonic pattern. Polyphonic syncopation is then computed, based on these three monophonic patterns, as described by Witek, Clarke, Kringelbach, et al. (2014).

are reported as the *hisync*, *midsync* and *losync* respectively (see Figure 3).

4.1.3. Polysync

Polyphonic syncopation is computed following the method proposed by Witek, Clarke, Kringelbach, et al. (2014) to obtain a single value from a polyphonic drum pattern (see Figure 4). Their algorithm computes monophonic syncopation values for three instrumental frequency ranges (low, mid and high frequency), then assigns an inversely proportional weight to the instrumental frequency range (low-frequency range has higher weight than mid-frequency and high-frequency), and then these three weighted values are added to obtain a single value. The algorithm is fully documented in the Supporting Information section of their paper.[5]

4.1.4. HiD, midD, loD

Sum of onsets for each different instrument group, divided by the total number of steps in the pattern (see Figure 5).

4.1.5. Losyness, midsyness, hisyness

Quotient of the syncopation value and the sum of onsets for each instrument group (see Figure 6).
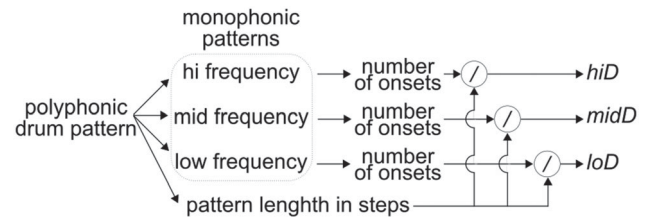
Figure 5. Block diagram description of *hiD*, *midD* and *loD*, descriptors computation. Three different monophonic patterns (low, mid and high) are extracted form the polyphonic pattern. The number of onsets of each monophonic pattern is divided by the pattern length in steps.
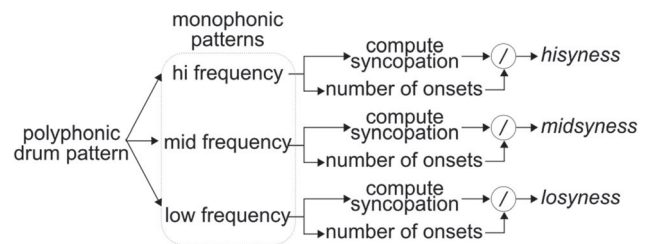
Figure 6. Block diagram description of *hisyness*, *midsyness* and *losyness*, descriptors computation. Three different monophonic patterns (low, mid and high) are extracted form the polyphonic pattern. The syncopation value of each monophonic pattern is divided by the number of onsets.

Figure 7. Block diagram description of *stepD*, descriptor computation. The number of steps that contain at least one onset is divided by the pattern length in steps.
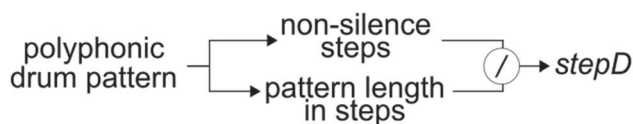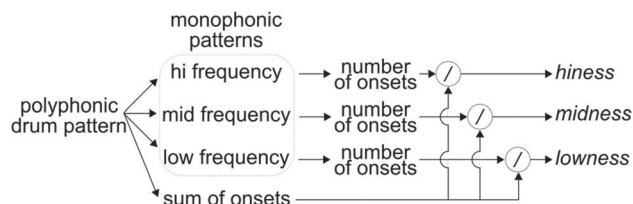

Figure 8. Block diagram description of *hiness*, *midness* and *lowness*, descriptors computation. Three different monophonic patterns (low, mid and high) are extracted form the polyphonic pattern. The number of onsets of each monophonic pattern is divided by the polyphonic pattern's sum of onsets.

4.1.6. StepD

Sum of the steps in the pattern which contain at least one onset, divided by the total amount of steps (see Figure 7).

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 (see Figure 8).

5. Experiment 1: Validation of rhythm features in Gabrielsson's space

Once a set of polyphonic drum descriptors has been defined, the next step is to test their performance in different real-life musical scenarios. As we mentioned above, Gabrielsson published a study on polyphonic drum rhythms' similarity (Gabrielsson, 1973b) which will be the starting point for our first experiment. Specially, the results of Gabrielsson's experiments 1 and 2 (GE1 and GE2) will be analysed, given the peculiarities of the rhythms selected for his experiment: all of them are reproduced with the same synthetic timbres of a drum machine, and with the same tempo (120 BPM). These factors reduce the intrusion of tempo and timbre in listener's similarity assessment. These experiments are also suitable 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) is kept constant (Collins et al., 2013).

The patterns used by Gabrielsson were the factory presets of the Ace Tone Rhythm Ace FR-3 drum machine which were recorded to magnetic tape. The patterns used in GE1 and GE2 were foxtrot, rockn'roll, rhumba, beguine, habanera and Waltz. These two experiments (GE1 and GE2) were designed to explore similarity between polyphonic drum rhythms and to create similarity matrices. For experiment 1 (GE1) 16 subjects (6 female and 10 male, with more than four years of amateur experience performing music) listened to triads of rhythms, and then selected which pair was the most similar. Summing across all subjects a similarity matrix was obtained. For GE2, 13 subjects (5 female and 8 male with same experience as subjects in GE1) listened to all possible pair combinations and rated 'how similar are the patterns among themselves', having 10 as perfect similarity. Sixteen subjects (6 women, 4 men), rated the stimuli.

The similarity matrices gathered by Gabrielsson from his two experiments (GE1 and GE2) were processed with a Multidimensional Scaling (MDS) algorithm (Kruskal, 1964). With this procedure he obtained two low-dimensional spaces where the drum patterns were located according to the distance reported by subjects. MDS was used to create a small dimensional representation of a dissimilarity matrix while minimising the distortion of the distances among instances of the matrix. The spaces obtained by Gabrielsson are a graphic representation of this dissimilarity matrix (Figure 9). While Gabrielsson's interpretations of the axes spanning the spaces are educated guesses without strong grounding on empirical data, our experiment will be devised to take advantage of the above-presented rhythm descriptors to interpret these axes.

In Gabrielsson's paper the patterns from the FR-3 drum machine are transcribed to symbolic musical notation. Our proposed polyphonic descriptors will be extracted from these transcribed patterns obtaining a descriptor vector for each pattern used in GE1 and GE2. These vectors will be used to approach the positions of each pattern in the rhythm space resulting from GE1 and GE2 (Figure 9). 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 both spaces. Lasso regression is a method typically used for variable selection as it helps to reduce the number of variables needed to get accurate predictions, hence enhancing both the interpretability and the accuracy of a model. Thus, the resulting set of descriptors should capture the essence of subjects' ratings revealed through the structure of the spaces.

Our 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 generalised to predict other spaces with patterns from different drumming styles.
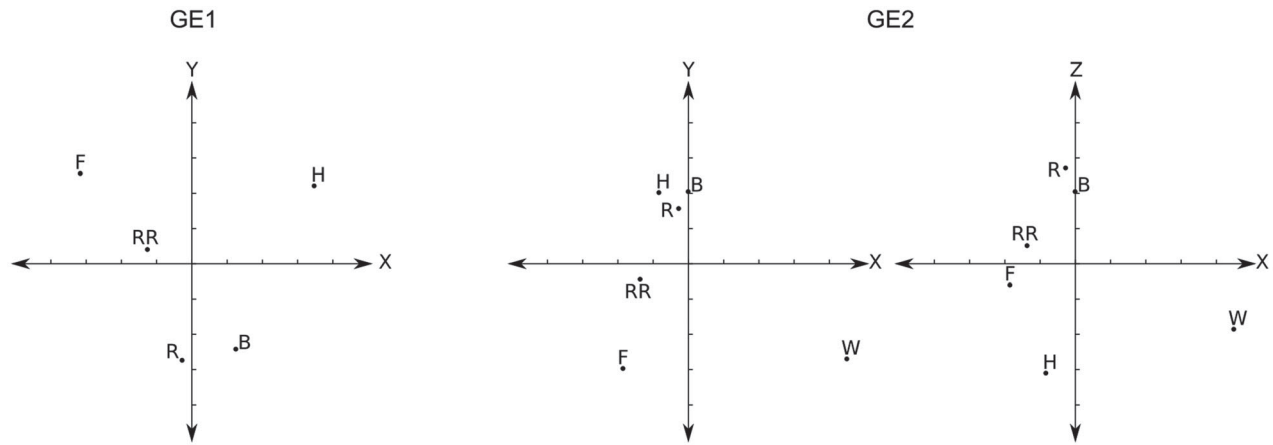
Figure 9. Gabrielsson's rhythm spaces from experiments 1 and 2. F: Foxtrot, RR: Rockn'roll, R: Rhumba, B: Beguine, H: Habanera, W: waltz (Gabrielsson, 1973b).

5.1. Methods

5.1.1. Materials

The resulting low-dimensional spaces from Gabrielsson's first two experiments (GE1 and GE2) are used as a source from which the coordinates of each pattern are extracted. These coordinates are presented in Table 2.

5.1.2. Procedure

Each pattern used in GE1 and GE2 is transcribed to MIDI format and all our symbolic descriptors are extracted from them. Then a Multi Task Lasso regression is used (alpha 0.03), setting the coordinates 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 maximise correlation of a subset of descriptors with the positions of the patterns in each low-dimensional space resulting from Gabrielsson's experiments (GE1 and GE2).

5.2. Results

The output of the Lasso analysis shows that using this set of descriptors {*midD*, *hiD*, *hiness*, *lowsync*, *hisyness*} the results of both axes of GE1 are perfectly linearly correlated, yielding a Spearman correlation of 0.999 (*p*-value $<$ .005) for both GE1 X and Y axes. The weights of

Table 2. Patterns used in GE1 and GE2 with their coordinates.

	Coordinates	
Pattern	GE1	GE2
foxtrot	−0.64, 0.55	−0.4, −0.58, −0.1
rockn'roll	−0.27, 0.09	−0.3, −0.1, 0.1
rhumba	−0.06, −0.55	−0.09, 0.3, 0.55
beguine	0.24, −0.49	0.0, 0.4, 0.41
habanera	0.7, 0.45	−0.18, 0.41, −0.61
waltz	–	0.85, −0.51, −0.36

Table 3. Descriptors and their weights for perfect correlation with GE1, after Lasso analysis.

	Axis weight	
Descriptor	X	Y
midD	1.735	−0.66
hiD	−0.089	−0.86
hiness	−0.09	−0.068
lowsync	0.102	−0.266
hisyness	−0.357	0.118
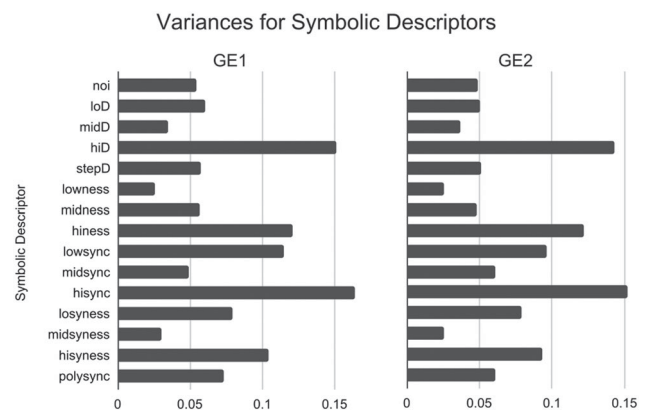
Figure 10. Variances of each symbolic descriptor after measuring the set of patterns used in Gabrielsson's experiments GE1(left) and GE2(right).

each descriptor are shown in Table 3 and the variances in Figure 10 (left).

For the space resulting from GE2, the Lasso analysis shows that the descriptor set {*midD*, *hiness*, *lowsync*, *midsync*, *hisync*, *losyness*, *hisyness*} yields perfect correlations with its three axes: Spearman correlation is 0.999 (*p*-value $<$ 0.005), 0.942 (*p*-value $<$ 0.005) and 0.999 (*p*-value $<$ 0.005) for axes X, Y, and Z, respectively. The weights of each descriptor are shown in Table 4 and the variances in Figure 10 (right).

Table 4. Descriptors and their weights for perfect correlation with GE2, after Lasso analysis.

Descriptor	Axis weight		
	X	Y	Z
midD	0.784	0.333	0.313
hiness	−0.072	−0.07	−0.032
lowsync	0.242	0.21	−1.12
midsync	0.574	−0.031	0.104
hisync	−0.708	1.005	0.81
losyness	0.0446	0.002	−0.052
hisyness	0.505	−1.411	−0.723

5.3. Leave-one-out descriptor analysis for GE1 and GE2

In order to scrutinise the set of descriptors obtained by fitting the patterns on GE1 and GE2, we performed a leave-one-out validation. One drum pattern was left out iteratively and a Multi Task Lasso regression was computed to obtain five different predictive models for both pattern collections used in GE1 and GE2. The resulting sets of descriptors from each predictive model was compiled in a histogram (Figure 11), as well as the predicted position of the complete set of patterns using each different model (Figure 12).

From the 15 descriptors available, 7 of them (47%) are never used as predictors for any model based on GE1 and 6 of them (40%) never used for any model based on GE2. Each resulting leave one out set is presented on Figure 11, ordered by the descriptors used in most models. The most common descriptors resulting from leave one out models of GE1 and GE2 and also belonging to the regressions with the complete set of patterns (GE1 and GE2) are {*lowsync, hisyness, hiness*} (Figure 11).

Comparing the precisions of the leave one out models (Figure 12) when predicting original positions of the patterns in GE1 and GE2, two trends can be observed: one where the left out pattern is the most distant to its original position (models 1, 3, 4 and 5 for GE1 and models 1, 5 and 6 for GE2), and another trend where some models are good predictors of the complete set (see flat response for models 2 in GE1 and 2, 3 and 4 for GE2 in Figure 12). Models in the first trend seem to have memorised the patterns of the training set and fail when predicting patterns not in the training set. Models in the second trend (good predictors) contain most of the descriptors that belong to the sets resulting from the regressions with the complete set of patterns. The set from model 2 and GE1 is {*lowsync, hiD, hisyness, midD, polysync*} and the set of the complete GE1 is {*lowsync, hiD, hisyness, midD, hiness*}. The same occurs in GE2, the set of the complete GE2 regression is {*midD, hiness, lowsync, midsync, hisync, losyness, hisyness*} and models 2, 3 and 4 are {*fhiness, lowsync, midsync, losyness, hisyness hiDg*}, {*fmidD, hiness, lowsync, midsync, hisync, hisynessg*}, {*hiness, lowsync, midsync, hisyness noig*}, respectively.

In general it can be observed that every good predictive model (either from the complete GE1 or GE2 sets of patterns or the good predictors of the leave one out models) contains the subset of patterns {*lowsync, hisynessg*}. Additionally, every good predictive model is either a subset or a very similar set to the one resulting from the regressions with the complete set of patterns (GE1 and GE2).

5.4. Discussion

Results show how syncopation is clearly a main factor for differentiating the patterns at the three different instrumental levels. Syncopation on the low-frequency
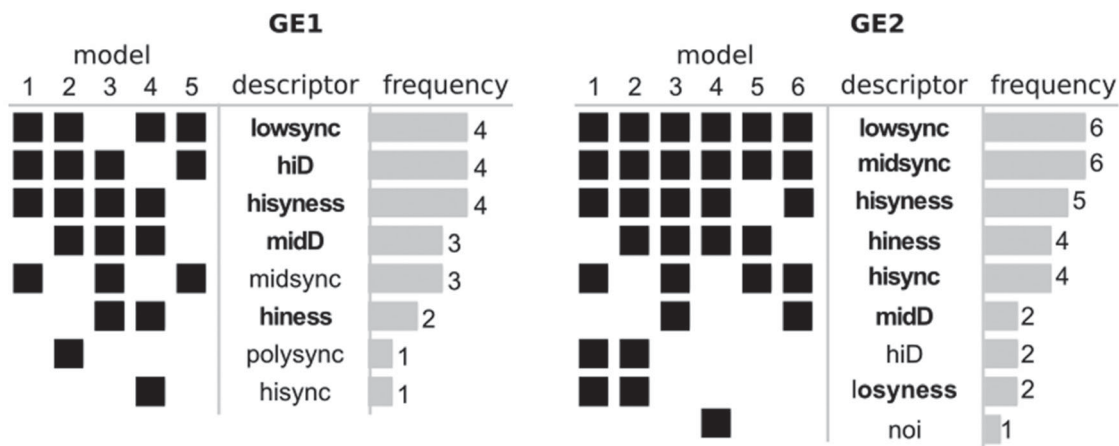
Figure 11. Models, descriptors and frequencies after performing the leave-one-out experiment for GE1 and GE2. Each model is a column where a square represents presence of a descriptor. A histogram summarises the presence of each descriptor. Descriptor names in b bold indicate the descriptor belongs to the set obtained by using the complete list of GE1 or GE2 patterns.
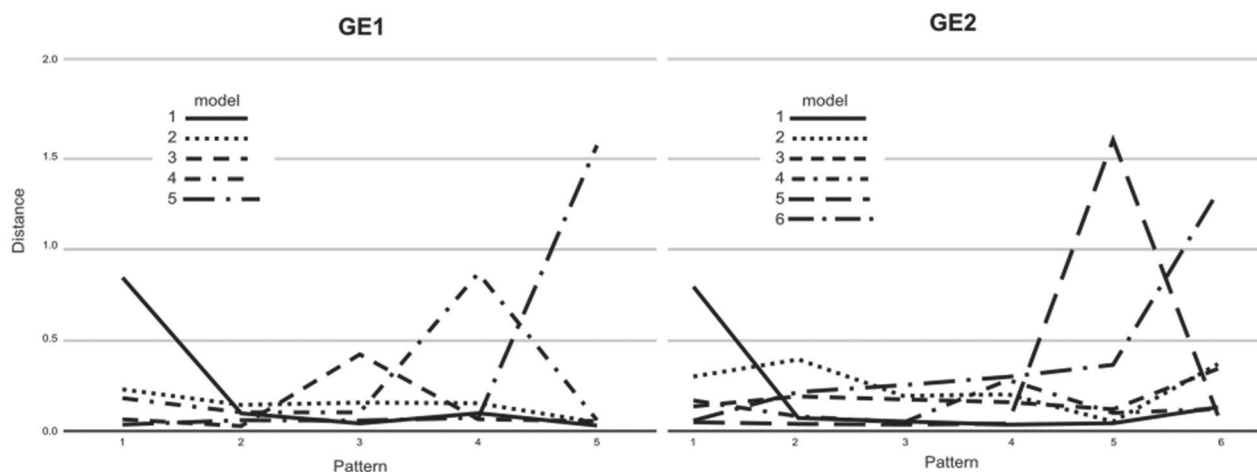
Figure 12. Distances from each original pattern to its predicted position by models created using the leave-one-out methodology for GE1 and GE2.

instrumental group {*lowsync*} is used to predict both axes of GE1 (Table 3). The quotient between the syncopation and the density is found relevant in the high and low frequency instrumental group {*losyness*, *hisyness*} (Tables 3 and 4). For both axes of GE1, the high density quotient is found relevant {*hiD*}. The density of the high and mid category of instruments is relevant, both axes coinciding in the importance of density in the mid instrumental level {*hiD*, *midD*}. The density percentage (instrumental density divided by total density) is found relevant for both axes in the high instrumental category {*hiness*}. Descriptors in the three instrumental categories are used to predict GE1 suggesting that they are useful for human similarity judgements of polyphonic rhythms. This validates the approach of mapping instruments to categories in the symbolic domain as discussed in section 4.

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

Pattern density, another of the factors described by Gabrielsson to influence similarity, is definitive for predicting the positions of the patterns in both experiments. Density, in the form of {*midD*, *highD*, *hiness*}, is a relevant factor for perceiving similarities. It is important to note that, contrastingly, the densities of low frequencies {*lowD*, *lowness*} are not present in the sets of relevant descriptors.

Although the number of different instruments in a sequence {*noi*} is another factor proposed by Gabrielsson to influence similarity and it is one of the descriptors computed, it had no relevance for the prediction of space GE1 or GE2, according to our analysis.

Given the results of the leave one out validation, we presume that the sets obtained by the Lasso analysis with the complete set of patterns for both GE1 and GE2 might not result as an overfitting artefact but instead, are the best possible sets. Therefore, in the next section we will explore how good predictors these sets are for a different collection of drum patterns.

6. Experiment 2: Generalisation of our feature set to EDM spaces

In the previous experiment, small sets of descriptors were found to quantitatively describe and predict Gabrielsson's GE1 and GE2 (Gabrielsson, 1973b) results (similarity ratings and the perceptual space computed from them). The question here is how general these features can be. Are these descriptors fitted to the particularities of the rhythms used by Gabrielsson, or would they work when other, quite different, rhythm patterns are rated? Computationally, this would mean that if we have a new set of drum patterns and their location in a human-based rhythm space, by extracting only the descriptors defined on the previous experiment (Section 5), and by using a dimensional reduction technique (as Gabrielsson did), the rhythm space could be configured with some accuracy. Two very well known dimensional reduction algorithms will be 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, used by Gabrielsson, is based on the dissimilarity matrix of the patterns, which is computed as the Euclidean distance between the descriptor vectors of each pattern.

For this experiment, and following Gabrielsson's methodology for experiment 2 (GE2) (i.e. rating the similarity between two patterns), a new EDM rhythm space is created based on subject ratings: selecting a collection of EDM patterns and presenting pairwise combinations to subjects who report their similarity, and then using either PCA or MDS to create a new rhythm space. An EDM drum rhythm collection is compiled specifically for this experiment.

6.1. Methods

6.1.1. Participants

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

6.1.2. Materials

In order to get a subject-based rhythm space, a set of rhythm patterns was needed, so we turned to the EDM production literature (Adamo, 2010; Brown & Griese, 2000; Emmerson, 2013; Hewitt, 2009; Snoman, 2012) and collected drum patterns explicitly associated to a certain EDM style. All patterns were 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 remaining 30% belonged to Garage, Drum n' Bass, Hip-Hop, Trance, Chillout, Dubstep, Jungle and Trip-Hop.

With the whole collection we created a preliminary rhythm space in order to sample the space and getting the stimuli for the experiment. First we extracted the complete list of symbolic descriptors (Section 4.1) and then using PCA we visualised them 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. The idea was to select the nine most different patterns in the space, according to the complete list of symbolic descriptors. This procedure was carried out for simplicity, as we could profit from 2D euclidean geometry to easily segment the space and proceed to pick one central and eight peripheral patterns. In this way, the list of 75 patterns was reduced to 9 patterns (see Table 5), each one intended to be representative enough of the variability of the included categories.[6] In order to be played

Table 5. Patterns selected from the subregions of the preliminary space. Left, centre, right and top, centre bottom represent the subdivisions of the space as explained in the text.

	Left	Centre	Right
Top	techno grinding analogue	techno industrial	techno hardcore
Centre	deep house	dirty house	deep tech house
Bottom	break synthetic subs	funk break	break funky drummer

in the rating experiment, the 9 patterns (from Techno, House and Breakbeat styles selected for the experiment) were rendered to audio at a constant tempo of 120 beats per minute combining single shot samples from the Roland 707, 808 and 909 drum machines. Although Breakbeat patterns use mostly sample-based drums, the drum machine sounds used in this experiment are extensively used throughout EDM styles. All selected patterns use instruments from the following set: Low Conga, Bass Drum, Side Stick, Maracas, Hand Clap, Snare, Closed Hi-Hat, Low Tom and Open Hi-Hat.

6.1.3. Procedure

A computer program in Pure Data was prepared to carry out the experiment. Before the subjects started the experiment, several patterns were presented for making them familiar with the timbre range of the percussive sounds used. In addition, examples of 'identical', 'potentially similar' and 'completely different' pairs of patterns were provided as reference to the range of variability to be found. These pairs were carefully selected from the 75 pattern list, as this collection contains very contrasting and rhythmically dissimilar items and also patterns that hold some rhythmic similarity. For the most dissimilar example, we selected a pattern that suggested very different rhythmic sensation in terms of number of instruments, syncopations and densities. For the similar example, we made sure both patterns shared some onsets at the same position preceding a silence, had a similar density as well as a similar number of instruments. To evaluate all combinations between the 9 patterns, subjects rated 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 control pairs were presented twice for determining the consistency of each subject's ratings, so, in total, subjects rated 40 pairs of patterns. Pairs were presented in a random order, preventing the same pattern to appear in consecutive pairs. Before a pair was reproduced, the order of its two patterns was also randomised so the same pair was presented in the two possible arrangements (i.e. a-b or b-a) to different subjects. Subjects listened to the same pair as many times as they needed, and the similarity value

---

[6] This strategy is used given the huge amount of time needed to compare 75 patterns pairwise in a human-based experiment, as subjects would have to rate 2775 different pairs of patterns.

was reported using 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). This procedure follows the one used by Gabrielsson in GE2 explained in Section 5. When the subjects completed the experiment, they answered 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, and the average number of times they listened to the pairs before answering. Finally, the possibility to leave a comment on the experiment was provided.

6.2. Results

In order to clarify the analysis, the 10 point scale was mapped to a 5 point scale where each value of the new scale groups two values of the original one (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 they 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 was 1.8 units (over 5) which is a 36% of

maximum variation. In order to approximate our analysis to that of Gabrielsson, we created a subgroup of subjects compliant with the musical background reported in his experiments, which consisted of '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. The inter quartile range mean for the rating is 1.81 units for the General group and 1.48 units for the Musicians group, suggesting more agreement in the latter than in the former group.

The observed mean for each assessed pair presents 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 'Deep Tech House' have 4 changes between groups. The difference between the spaces generated by the Musicians group and the general group were not big to keep their spaces separated so, from now on, we operate with all results. Then, we use the median rating for each pair to create the dissimilarity matrix.

An MDS is applied to the obtained dissimilarity matrix generating a bi-dimensional space (Figure 13, left). We can observe there 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
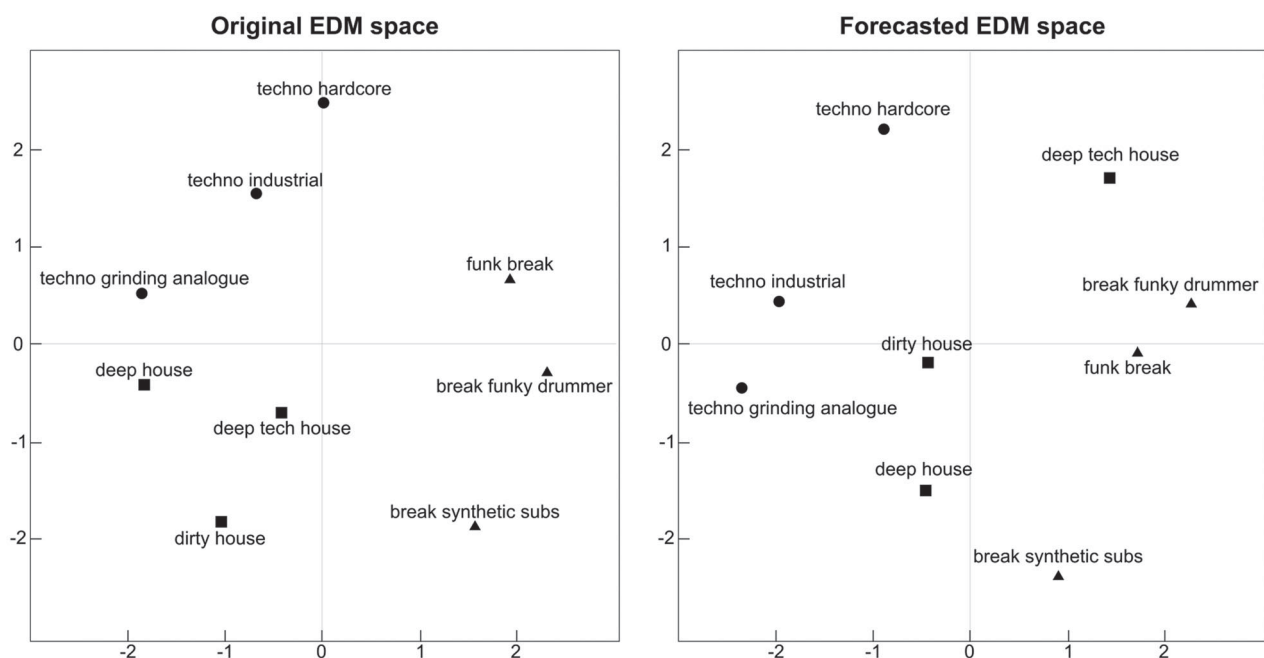
Figure 13. Original and forecasted EDM spaces. Left: Bi-dimensional space obtained by using MDS on the dissimilarity matrix of subject ratings. Right: Bi-dimensional space obtained by our descriptor-based model.

Table 6. Spearman Rank correlations between each EDM axis and the prediction using the descriptor sets from GE1 rendered using PCA and MDS. The E1 set MDS (in bold) presents the highest correlations and lower *p*-value.

| Method | Correlation values | |
	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

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, 2006), 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 the same style end up located in specific and independent regions.

Finally, from the patterns in the EDM space, we extract the two descriptor sets found in the previous experiment (exploring GE1 and GE2 with our descriptors). The descriptor set for GE1 is {*midD*, *hiD*, *hiness*, *lowsync*, *hisyness*} and for GE2 is {*midD*, *hiness*, *lowsync*, *midsync*, *hisync*, *losyness*, *hisyness*}. Using these descriptor values we compute PCA and MDS to evaluate whether the locations of the EDM patterns can be predicted with any of the sets. Table 6 presents the correlations between the positions of each pattern in the predicted space and in the resulting EDM space.

6.3. Relevant set of descriptors

The descriptors derived from the Multitask Lasso analysis yielding 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 normalised 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.

Using the set of descriptors resulting from our first experiment (E1), for analysing the patterns of the EDM experiment and then applying MDS to those results (E1 set MDS), we observe Spearman correlations with values 0.67 (*p*-value < 0.05) and 0.78 (*p*-value < 0.05) for the *x* and *y* axis, respectively (see Table 6 and Figure 11, right). The other two combinations of sets and dimensional reduction techniques that are almost significantly correlated with the EDM space are E1 set PCA and E2 set MDS, but none of them yield statistical significant correlations for any of the axes. As a conclusive statement, we have shown that using the E1 set, and then MDS, captures the distance sensations reported by the subjects both in Gabrielsson's experiment 1 and also in our EDM experiment.

6.4. 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 the construction of very general rhythm spaces as reported (without such descriptors-based analysis) in early literature. We have thus seen that these descriptors, based on main concepts of rhythm cognition, allow us to construct stylistic meaningful spaces constrained to EDM. Consequently, we could use these descriptors in systems which present, visualise and manipulate pattern collections. This way, users could have 2D representations which, being close to their mental representations, could be exploited for meaningful 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 that arises from our second experiment. This can be taken as a demonstration of how musical concepts such as House, Techno and Breakbeat emerge based on listening to a handful of instances, each occupying a specific region of a conceptual space. That said, 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 on how patterns are both perceived and their similarity judged.

Drum patterns used in Gabrielsson's experiment 1 come from different dance music cultures, namely Western, Afro Cuban and South American. 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 experiment.

A further clarifying experiment, beyond the scope of this article, would consist of adding some EDM patterns to those used in Gabrielsson's, in order to see how far EDM ones are from the rest, and how close between them they appear in such a big picture.

To conclude, our results show that using a reduced set of descriptors, namely {*midD*, *hiD*, *hiness*, *lowsync*, *hisyness*}, computing Euclidean distance between the vectors describing each rhythm pattern, and then using MDS, a perceptual rhythm space composed of EDM patterns is reproduced with significant correlation values with listeners' ratings. Although this is a significant advancement towards a system capable of automatic drum pattern organisation, further experiments must be carried out to evaluate its robustness when applied to larger datasets.

7. A rhythm space for rhythm generation

In the project that framed the research reported here,[7] EDM drum production has been studied in order to understand specific technological needs which help improve its current practice. One such necessary improvement is to organise and explore a collection of drum patterns (which usually contains hundreds or thousands of 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 organising a collection of drum files by some of their meaningful properties, similar patterns should be close together so they can be browsed and retrieved easily. One structure that can deal with this type of arrangement is a low dimensional coordinate system, as it allows for the visualisation of many elements located according to their values in the different dimensions. Such spaces favour that observers can make sense of a collection, as they can both grasp the local relations of their elements as well as inspect the complete set. Additionally, if such a rhythm space is made interactive, the user can also point to a specific position and enable 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 application that can be used for organising, visualising and retrieving drum pattern files.

The backbone of a rhythm space for EDM drums is an adequate similarity distance, measurable from the patterns themselves and aligned with the closeness sensations that human subjects may report. Given this perspective, it makes sense to use the results presented in Section 6 as the foundation supporting drum rhythm spaces.

In the process of studying rhythm spaces it was suggested that, using the appropriate metrics, any collection of drum patterns can be properly arranged into a low dimensional map where points represent patterns that can be retrieved. However, such a 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, new algorithms for drum pattern interpolation can be developed as an expansion to an interactive drum rhythm space. By implementing these algorithms as a layer on top of the rhythm space, the space becomes continuous. Any blank point (a point where no pattern from the collection is located) can retrieve a new pattern created in real-time based on its neighbours. With these algorithms, a rhythm space is enhanced with generative capabilities, becoming a visualisation tool that auto-generates new elements on the fly, expanding its original contents.

7.1. Rhythm space interpolation

As a means to add continuity to a rhythm space built up from a discrete collection of patterns, a drum interpolation algorithm is proposed. Based on a Delaunay triangulation (Lee & Schachter, 1980) of a 2D rhythm space, this algorithm weights 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, with features that resemble the surrounding patterns. Our algorithm takes care of smoothly introducing and removing onsets in the pattern, based on the interpolation values.

The main elements of our drum pattern interpolation algorithm (DPIA) are the onsets, defined by an instrument and the step (rows and columns respectively in Figure 15). The first phase for interpolating is computing, for each pattern, the step density (the sum of onsets at each step) and the pattern weight ($p\omega$) which is proportional to the proximity of the interpolation pointer (p) to each pattern being interpolated. Each distance (ap, bp, cp) is normalised by dividing its magnitude by the distance of the line that starts in each vertex, ends in the opposite edge, and passes through p (ad, be, cf) (see Figure 14). In parallel, by making a stepwise weighted sum, the step
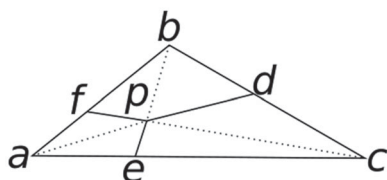
Figure 14. Pattern weights ($ap\omega$, $bp\omega$, $cp\omega$) are derived from the distance of point P to the three vertices of the triangle ABC.

densities of each pattern are combined (the sum of the different onsets) to produce a step density vector (SDV) having a specific density value per step.

Onsets in each pattern have three weights assigned: syncopation value ($s\omega$), frequency class ($f\omega$) and density ratio ($d\omega$). Highest weights are assigned respectively to onsets with higher syncopation (derived from the metrical weights) (Longuet-Higgins & Lee, 1984), lower frequency instruments and most common onsets within a pattern (See Figure 15).

An onset pattern weight ($Op\omega$) is obtained by dividing the pattern's weight ($p\omega$) by the number of onsets at every step (e.g. if onsets are snare and kick and the pattern weight is $p\omega = 0.5$, then Op for the kick and the snare is 0.25). Using the four onset weights ($Os\omega$, $Of\omega$, $Od\omega$, $Op\omega$), each onset of every interpolated pattern gets a final weight ($O\omega$) obtained by multiplying the four weights: $O\omega = Os\omega \times Of\omega \times Od\omega \times Op\omega$. Finally at every step, the weights from common instruments from all interpolated patterns are added (i.e. snare $0.2 +$ snare $0.1 =$ snare 0.3). A final set of unique onsets is obtained and sorted in descending order according to their final weights ($O\omega$). At every step, the value of the step density
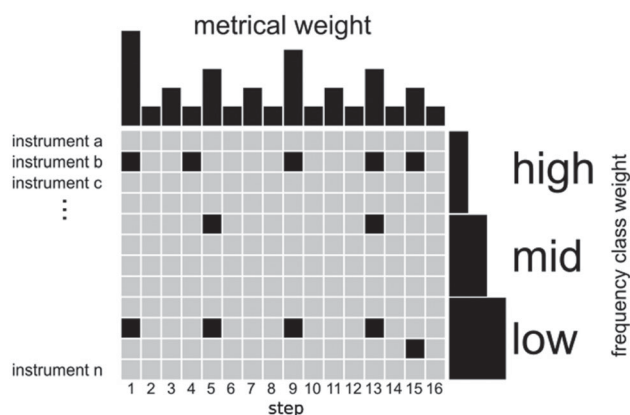
Figure 15. Piano roll representation of a drum pattern with time in steps on the *x* axis and instruments on the *y* axis. Frequency class weights ($f\omega$) are assigned to each onset given the frequency class of the instrument. Syncopation weights ($s\omega$) are assigned given the syncopation of each onset derived from the metrical weight.

vector (SDV) is used as a filter to control the number of output onsets. For example, if the SDV is three, the three onsets with the highest weights are output. This procedure is carried out at every pattern step (i.e. 16th note slot) until a new interpolated pattern is generated.

7.2. Experiment 3: Interpolation experiment

In order to explore how the DPIA works, we designed an experiment to evaluate how subjects perceive the interpolated pattern in relation to the original ones. We want to explore if an interpolated pattern conveys rhythmic information traceable to the parent patterns and proportional to their weights. For simplicity, our experiment is based on two parent patterns (*A* and *B*) which form a resulting interpolated pattern (*x*). In an ideal interpolation scenario, the sensation of listening to *x* resulting from combining *A* and *B* in equal proportions will resemble both *A* and *B*, but it will not be completely identified with either of them. As *x* is a synthesis of *A* and *B*, a transition of *x* followed by *A* or by *B* will present smaller structural differences than that of a transition from *A* to *B*. Our hypothesis is that if subjects listen to *A*, *B* and *x* focusing only on rhythm, the contrast between *AB* pairs would be rated as more contrasting than the contrast between *Ax* pairs or than the contrast between *Bx* pairs.

7.2.1. Participants

Twenty-one subjects participated in our experiment. They were mostly university students from Asia, Europe and Latin America, all of them with at least two years of music lessons and an average age of 4.8 years listening to electronic dance music.

7.2.2. Materials

The same nine patterns used for the EDM experiment (E2, Section 6.1.2) were used in this interpolation experiment. From each of the 36 different combinations derived from these 9 patterns, a new interpolated pattern at a 50% distance from each was generated. We will refer to the original patterns as A and B and to the interpolated pattern as *x*. Patterns were reproduced at a constant 120 BPM tempo using the same sound samples extracted from Roland TR 707, 808, and 909 drum machines used for the EDM experiment.

7.2.3. Procedure

Subjects were asked to interact with a computer program implemented in Pure Data where they could read the instructions of the task. Three buttons were presented, each one associated with a Likert scale with a range from 1 to 5. Instructions invited them to 'Click on each of the

buttons to listen to a pair of drum patterns. After listening carefully, rate how abrupt the change is between the two patterns within the couple; in other words, how contrasting you find them'. Each button played back a couple of patterns (two bars of one pattern and two bars of another one) with the same tempo and without any silence between them. Couples formed by combining patterns A, B and *x* were randomly distributed within the three buttons (*AB*, *AX* and *BX*). The order in which patterns were played back was also randomised (i.e. *AB* or *BA*). This procedure was repeated with 36 different *AB* couples, resulting in an experiment duration of approximately 20 minutes.

7.3. Results

In 71.6% of the rounds, *AB* pairs were rated as being more or equally contrasting than *Ax* and Bx. In 56.7% of the rounds *AB* pairs were rated as being more contrasting than pairs *Ax* and *Bx*. Ratings of the stimuli were consistent as 94% of *AB* pairs were considered by more than 50% of the subjects to be equally or more contrasting than *Ax* and *Bx*, and 58% of *AB* pairs were considered by 75% of the subjects to be equally or more contrasting than *Ax* and *Bx*. However, two stimuli pairs (5%) were specifically problematic as only 5% of the subjects rated

them as being equally or more contrasting than *Ax* and *Bx*. The AB pairs with maximum and minimum contrasts had average scores of 3.2 and 1.36 respectively. A one way analysis of variance (ANOVA) presents strong evidence that the expected values in the two different experimental conditions (interpolated: *Ax*, *Bx* and non-interpolated: *AB*) differ (p-value < 0.05, F-value $= 226.719$, degrees of freedom $= 1$). Post-hoc analyses show significant independence among the three experimental conditions (p-value < 0.05). The means for the interpolated *Ax*, *Bx* and the non-interpolated *AB* are 1.82 and 2.63 respectively.

7.4. Discussion

As reported in 71.6% of the listening rounds, new interpolated *x* patterns are less or equally contrasting perceptually when transitioning to *A* and *B*, than the perceptual contrast for transitioning from *A* to *B*. This suggests that our algorithm is proficient in grasping rhythmic aspects of two patterns (*A* and *B*) and then blending those aspects to create a new pattern. The proposed DPIA's method for assigning weights to patterns' onsets seems to be an appropriate process for extracting important cues from two patterns which are to be interpolated.
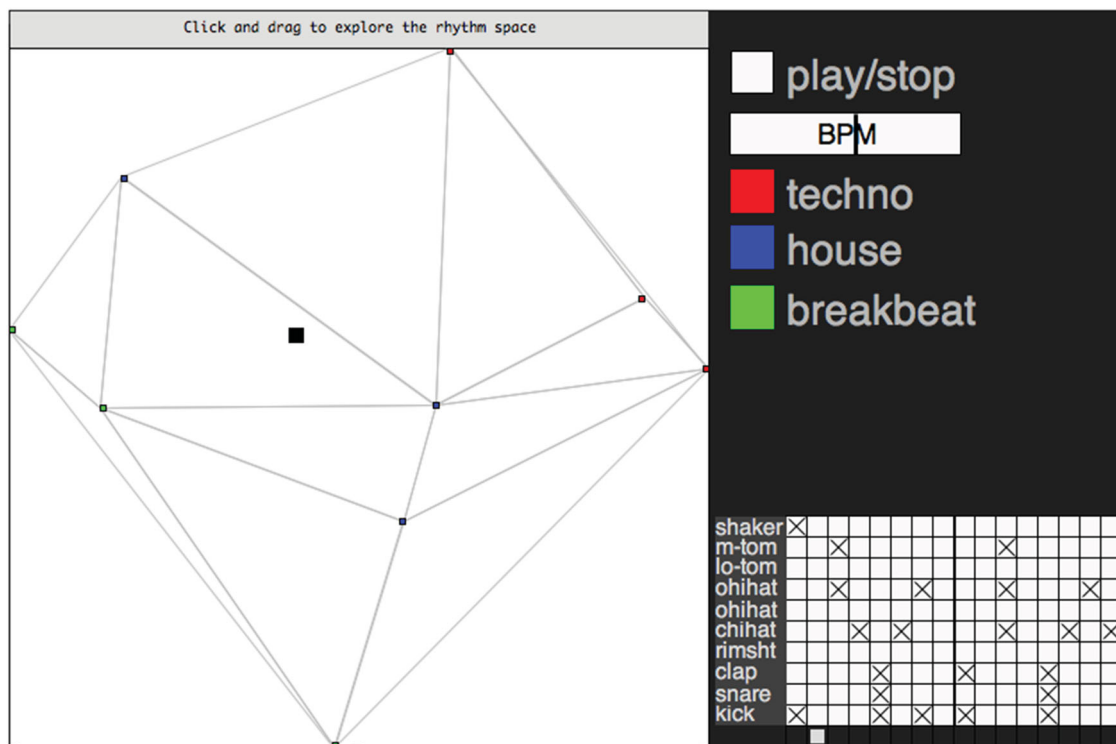
Figure 16. Screenshot of the rhythm space user interface. The white region on the left holds the nine patterns used in our EDM experiment, represented by the small dots that conform the vertices of each triangle. The black square indicates the user's browsing position. The grid in the lower right region presents the output drum pattern.

We found that the 14.9% difference between rating *AB* equally or more contrasting than *ABx* (71.6%), and rating strictly *AB* as more contrasting than *ABx* (56.7%) is due to some sort of perceived asymmetry in the interpolated pattern. That is, the generated *x* pattern was perceived as not in the middle between *A* and *B*, but leaning towards one of them, causing the reported contrast to be unbalanced. This can suggest that if *x* is not perceived as being in the middle but closer to *B* then the contrast between *Ax* and *AB* could be reported as the same.

Exploring the cases where the reported sensation of contrast does not comply with our hypothesis (*AB* was reported as less contrasting than *Ax* and *Bx*) it was found that it happened mainly with Breakbeat patterns. All three Breakbeat patterns used in the experiment are composed only of kick, snare and hihat, while all Techno and House patterns contain more than three instruments. The difficulty to interpolate pairs that include Breakbeat patterns can be produced by the DPIA dividing the 50% interpolation weight between few types of sounds, especially when, in Breakbeat patterns, there is at most a two-onset coincidence at one step. This makes every Breakbeat pattern onset to be highly weighted, especially the kick drum, as it is multiplied by a higher frequency class weight (f). On the contrary, House and Techno patterns typically contain coincidences of 3 and even 4 instruments on the same step. This weakens the final weight of all instruments, even the low frequency ones as the kick. A further enhancement of the algorithm can be to introduce the number of instruments as an additional source of discrimination.

Although the experiment was carried out by interpolating between two patterns, the interpolation process (as described in section 7.2) is suited to interpolate between two or more patterns. For the rhythm spaces (described in section 7.1), in which three patterns are interpolated (via Delaunay triangles), the presented interpolation algorithm is suited for continuous bi dimensional explorations of drum pattern collections. An example of a working prototype of a rhythm space user interface is presented in Figure 16.

8. Conclusions

The original research presented in this paper explores the domain of electronic dance music drum patterns from a cognitive perspective. A method for analysing collections of symbolic drum patterns and organising them according to a rhythm cognition model is presented, and a musical application for processing drum rhythms that is developed and described from our reported findings.

One contribution of this work is the articulation of different notions of polyphonic rhythm processing into a set of MIDI-based computable features. Specially Grabrielsson's notions on the human drivers of polyphonic rhythm processing, which he described verbally but did not formalise in any mathematical model, have been turned into computable features.

The experiments on human processing of polyphonic drum similarity led to designing new symbolic rhythm descriptors, and to finding how a sub-set of these 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 analysed and processed to obtain a low dimensional map where all patterns are organised 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 a collection of multicultural 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 the multidimensional scaling technique.

Considering the accuracy for predicting human-based rhythm spaces, our method for analysing and organising rhythm patterns was converted into a novel method for the visualisation, retrieval and generation of drum patterns. A novel software application uses the same descriptors and MDS technique discussed above, to convert a collection of symbolic drum patterns into a bi-dimensional space which organises patterns automatically, given their rhythmic properties. The rhythm spaces created this way depict the analysed collection of patterns as points in a space (which are located according to their similarity). Hence, patterns perceived as being alike appear close together, and separated from dissimilar ones. This map-like structure is used to visualise a complete collection of patterns and explore it, retrieving patterns by pointing at them. This whole procedure targets the actual need for browsers specialised in music content (as it is the case for drum patterns), allowing for complete collections to be visualised and explored, and expanding the actual (limited) possibilities of music content browsing. The process of designing this application was naturally complemented with the development of an algorithm for pattern-interpolation. This algorithm broadens the capabilities of a rhythm space as it allows the exploration of regions where no patterns are located, and

the generation of 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, as it is browsed, generates new patterns beyond the contents of the collection. The novel application presented here becomes then a tool that goes beyond organising and visualising a drum pattern collection as a rhythm space, making possible the generation of new patterns in places and regions where the collection falls short of items. In conclusion, by means of our research rhythm spaces were turned from a navigational and representational metaphor into a generative one.

Acknowledgements

Funding

ORCID

Daniel Gómez-Marín http://orcid.org/0000-0001-7694-3919
Sergi Jordà http://orcid.org/0000-0003-2964-4013
Perfecto Herrera http://orcid.org/0000-0003-2799-7675

References

Adamo, M. (2010). *The breakbeat Bible*. Hudson.

Ballester, J., Patris, B., Symoneaux, R., & Valentin, D. (2008). Conceptual vs. perceptual wine spaces: Does expertise matter? *Food Quality and Preference*, *19*(3), 267–276. https://doi.org/10.1016/j.foodqual.2007.08.001

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. https://doi.org/10.1371/journal.pone.0097467

Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. MIT Press.

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

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*, *82*, 1–17. https://doi.org/10.1007/s00426-017-0894-2.

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

Cao, E., Lotstein, M., & Johnson-Laird, P. N. (2014). Similarity and families of musical rhythms. *Music Perception*, *31*(5), 444–469. https://doi.org/10.1525/mp.2014.31.5.444

Chen, J. C. C., & Chen, A. L. P. (1998). Query by rhythm: An approach for song retrieval in music databases. In Avi Silberschatz, Aidong Zhang, & Sharad Mehrotra (Eds.), *Proceedings eighth international workshop on research issues in data engineering, 1998. Continuous-media databases and applications* (pp. 139–146). IEEE Computer Society.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Collins, N., Schedel, M., & Wilson, S. (2013). *Electronic music. Cambridge introductions to music*. Cambridge University Press. https://www.cambridge.org/core/books/electronic-music/4FD04D2A538AB21D7504B9CEA054DB4F.

Desain, P., & Honing, H. (1999). Computational models of beat induction: The rule-based approach. *Journal of New Music Research*, *28*(1), 29–42. https://doi.org/10.1076/jnmr.28.1.29.3123

Desain, P., & Honing, H. (2003). The formation of rhythmic categories and metric priming. *Perception*, *32*(3), 341–365. https://doi.org/10.1068/p3370

Einbond, A., Schwarz, D., & Bresson, J. (2009). Corpus-based transcription as an approach to the compositional control of timbre. In *International computer music conference proceedings* (Vol. 2009, pp. 223–226). International Computer Music Association.

Ellis, D. P. W., & Arroyo, J. (2004). Eigenrhythms: Drum pattern basis sets for classification and generation. In Claudia Lomelí Buyoli & Ramón Loureiro (Eds.), *Ismir 2004: 5th international conference on music information retrieval* (pp. 101–106). Audiovisual Institute, Pompeu Fabra University.

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

Gabrielsson, A. (1973a). Similarity ratings and dimension analyses of auditory rhythm patterns: I. *Scandinavian Journal of Psychology*, *14*(1), 138–160. https://doi.org/10.1111/j.1467-9450.1973.tb00105.x

Gabrielsson, A. (1973b). Similarity ratings and dimension analyses of auditory rhythm patterns: II. *Scandinavian Journal of Psychology*, *14*(3), 161–176. https://doi.org/10.1111/j.1467-9450.1973.tb00106.x

Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought* (1st ed.). MIT Press.

Gómez-Marín, D., Jordà, S., & Herrera, P. (2015a). *PAD and SAD: Two awareness-weighted rhythmic similarity distances*. 16th international society for music information retrieval conference ismir. http://mtg.upf.edu/system/files/publications/pad%20and%20sad_0.pdf.

Gómez-Marín, D., Jordà, S., & Herrera, P. (2015b). *Strictly rhythm: Exploring the effects of identical regions and meter induction in rhythmic similarity perception*. 11th international symposium on computer music multidisciplinary research (cmmr), Plymouth.

Gouyon, F., Dixon, S., Pampalk, E., & Widmer, G. (2004). Evaluating rhythmic descriptors for musical genre classification. In Daniel R. von Recklinghausen (Ed.), *Aes 25th international conference* (pp. 1–9). Audio Engineering Society.

Graham, R., Manzione, C., Bridges, B., & Brent, W. (2017). *Exploring pitch and timbre through 3D spaces: Embodied models in virtual reality as a basis for performance systems design*. New Interfaces for Musical Expression Proceedings.

Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *The Journal of the Acoustical Society of America*, *61*(5), 1270–1277. https://doi.org/10.1121/1.381428

Guastavino, C., Gómez, F., Toussaint, G., Marandola, F., & Gómez, E. (2009). Measuring similarity between Flamenco rhythmic patterns. *Journal of New Music Research*, *38*(2), 129–138. https://doi.org/10.1080/09298210903229968.

Halpern, A. R., Zatorre, R. J., Bouffard, M., & Johnson, J. A. (2004). Behavioral and neural correlates of perceived and imagined musical timbre. *Neuropsychologia*, *42*(9), 1281–1292. https://doi.org/10.1016/j.neuropsychologia.2003.12.017

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

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. https://doi.org/10.3758/BF03212154

Hourdin, C., Charbonneau, G., & Moussa, T. (1997). A multidimensional scaling analysis of musical instruments' time-varying spectra. *Computer Music Journal*, *21*(2), 40–55. https://doi.org/10.2307/3681107

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. https://doi.org/10.1073/pnas.1402039111

Johnson-Laird, P. N. (1991). Rhythm and meter: A theory at the computational level. *Psychomusicology: A Journal of Research in Music Cognition*, *10*(2), 88–106. https://doi.org/10.1037/h0094139

Krumhansl, C. L. (1979). The psychological representation of musical pitch in a tonal context. *Cognitive Psychology*, *11*(3), 346–374. https://doi.org/10.1016/0010-0285(79)90016-1

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, *29*(1), 1–27. https://doi.org/10.1007/BF02289565

Lee, D.-T., & Schachter, B. J. (1980). Two algorithms for constructing a Delaunay triangulation. *International Journal of Computer & Information Sciences*, *9*(3), 219–242. https://doi.org/10.1007/BF00977785

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

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

Longuet-Higgins, H. C., & Lee, C. S. (1984). The rhythmic interpretation of monophonic music. *Music Perception*, *1*(4), 424–441. https://doi.org/10.2307/40285271

Makris, D., Kaliakatsos-Papakostas, M., Karydis, I., & Kermanidis, K. L. (2017). *Combining LSTM and feed forward neural networks for conditional rhythm composition*. International conference on engineering applications of neural networks.

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research*, *58*(3), 177–192. https://doi.org/10.1007/BF00419633

Paiement, J.-F., Grandvalet, Y., Bengio, S., & Eck, D. (2007). *A generative model for rhythms*. Nips workshop on brain, music and cognition.

Palmer, C., & Krumhansl, C. L. (1990). Mental representations for musical meter. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(4), 728–741. https://doi.org/10.1037/0096-1523.16.4.728

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

Post, O., & Toussaint, G. (2011). The edit distance as a measure of perceived rhythmic similarity. *Empirical Musicology Review*, *6*(3), 164–179. https://doi.org/10.18061/1811/52811

Reynolds, S. (2012). *Energy flash: A journey through rave music and dance culture*. Soft Skull Press.

Rocamora, M., Jure, L., & Biscainho, L. W. P. (2014). Tools for detection and classification of Piano drum patterns from Candombe recordings. In Timour Klouche & Eduardo R. Miranda (Eds.), *Proc. of the 9th conference on interdisciplinary musicology (cim14)* (pp. 382–387). CIM.

Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, *1*(1), 54–87. https://doi.org/10.1016/0022-2496(64)90017-3

Sicko, D. (2010). *Techno rebels: The renegades of electronic funk*. Wayne State University Press.

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

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Toussaint, G. T. (2004). A comparison of rhythmic similarity measures. In Claudia Lomelí Buyoli & Ramón Loureiro (Eds.), *Proceedings of the 7th international society for music information retrieval conference (ismir 2004) ismir* (pp. 242–245). Audiovisual Institute, Pompeu Fabra University.

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). https://doi.org/10.4407/jims.2012.12.002.

Turquois, C., Hermant, M., Gómez-Marín, D., & Jordà, S. (2016). Exploring the benefits of 2D visualizations for drum samples retrieval. In ACM (Ed.), *Proceedings of the 2016 acm on conference on human information interaction and retrieval* (pp. 329–332). Association for Computing Machinery.

van der Weij, B., Pearce, M. T., & Honing, H. (2017). A probabilistic model of meter perception: Simulating enculturation. *Frontiers in Psychology*, *8*, 824. https://doi.org/10.3389/fpsyg.2017.00824

Wessel, D. L. (1979). Timbre space as a musical control structure. *Computer Music Journal*, 45–52. https://doi.org/10.2307/3680283

Witek, M. A. G., Clarke, E. F., Kringelbach, M. L., & Vuust, P. (2014). Effects of polyphonic context, Instrumentation, and metrical location on syncopation in music. *Music Perception*, *32*(2), 201–217. https://doi.org/10.1525/mp.2014.32.2.201

Witek, M. A. G., Clarke, E. F., Wallentin, M., Kringelbach, M. L., & Vuust, P. (2014). Syncopation, body-movement and pleasure in groove music. *PLoS One*, *9*(4). https://doi.org/10.1371/journal.pone.0094446.

Yee-King, M. J. (2011). *Automatic sound synthesizer programming: Techniques and applications* [Unpublished doctoral dissertation]. University of Sussex.

Zwarts, J. (2015). Conceptual spaces, features, and word meanings: The case of Dutch shirts. In Peter Gärdenfors & Frank Zenker (Eds.), *Applications of conceptual spaces* (pp. 57–78). Springer.