Percussive spaces

Definition and use of percussive spaces for analysis and interaction

Daniel Gómez Marín

MASTER THESIS UPF /2014

THESIS SUPERVISORS

Dr. Sergi Jordà, Dr. Perfecto Herrera

DEPARTMENT OF INFORMATION AND COMMUNICATION TECHNOLOGIES

Universitat Pompeu Fabra Barcelona

Abstract

This text presents the technical details of a system used to generate new rhythms based on similarity metrics to a reference pattern. Several important steps were done before the realization of such a system. First a thorough revision of the music cognition scientific knowledge that backs up the perception of rhythm was carried out. Then the algorithms from the most developed theories, namely the edit distance and the syncopation distance, were implemented, and compared and tested in a music genre classification task. The experience drawn from these activities is used as raw material to build up an interactive application for rhythm generation. The application produces rhythmic variations of a reference pattern using distance metrics such as the edit distance and the syncopation distance. Throughout this work it is realized that the apparently simple question of similarity between rhythms is still an open subject under current research. This fact makes of a cognition driven percussive system an interesting topic for research and development. Some contributions resulting from this project are the compilation of up to date scientific literature on the topic, the comparison of existing models of similarity, the development of new open source tools for rhythm analysis, the proof of concept of this tools in a real life analysis task and finally a working system for rhythm production and performance. Results show that the rhythmic metrics used are useful to generate new rhythmic patterns based on perceptual distances to a reference pattern.

Índex

1. INTRODUCTION

This work is inspired on the need to develop new music production tools based both on interpreting what a musician is playing and also on suggesting new interesting musical elements. Specifically, the topic of rhythm is addressed. The goal is to create an assistant capable of understanding an actual rhythmic context and proposing new rhythms, either for percussive or melodic instruments, in which the degree of transformation is controlled by a user. The assistant will receive digital musical information i.e. note messages, process it and prompt with new possible rhythmic structures to be used in the ongoing musical performance or production. As the idea is to create a musical tool, defining a small dimensional space suited for human retrieval and interaction is a major goal of this work. The starting point in the definition of this space is music cognition and the models proposed by scientists that study human rhythm perception.

Recent approaches to the comprehension of rhythm perception have been based on these two questions *how are two rhythmic patterns perceived as similar? What are the internal mechanisms of the brain when defining the closeness of two rhythms?* Although there is still no clear answer to these questions, there is some experimental research going on. Most of these researches propose new models that try to predict human judgment of similarity between rhythms. As of today there are mainly two broad hypotheses that explain rhythmic similarity. One is based on measuring the transformations of the elements of a pattern to become another one. The other one is based on understanding the relation of the onsets within a pattern with the beat it induces on a listener. Both approaches are still under development, and there is still no complete answer to the rhythmic similarity question. An answer to the rhythmic similarity questions would mean a step forward towards a definite model of percussive perception and to a deeper understanding of the way humans respond to periodic sound. The present work is an effort to materialize the current state of the art in rhythm cognition into an interactive musical tool for percussion.

The main objectives to be developed in this work are, first to define and validate a cognitive space for rhythmic patterns suitable for computer-aided exploration and navigation. The second objective will be to develop tools for real time interaction with drum sounds and rhythmic patterns based on the cognitive space.

The first part of this report presents a revision of publications around the topic of rhythmic similarity, music cognition experiments and engineering approaches to extract rhythmic properties from sound signals. Two main music cognition approaches to address the issue of rhythmic similarity providing experimental results are identified. Data mining and machine learning based approaches are also reviewed and commented. The cognitive approach is selected to be used throughout this work, specifically the theories based on two similarity metrics: the edit distance and the syncopation distance.

To carry out the activities of this work, the methodologies and algorithms described in the literature had to be implemented. Also two databases were created, one that included all possible rhythms created in a 16 step grid based exclusively on silences; and another one with songs in MIDI format to be classified according to their musical genre. The latter contains popular songs within different musical genres and the analysis is focused solely on the rhythmic structure of the bass track of each song.

Three activities are defined in order to test different aspects of the metrics and descriptors in which they are based. One is to compare the performance of the metrics when being used to measure the distance between a reference and large set of different rhythms. A second activity is to test if the rhythmic similarity metrics are useful in classifying songs by their genre. This is task is performed in order to test the rhythmic similarity descriptors from on the side of the performer and not the listener. As the descriptors come from experiments in human perception, meaning the exercise of listening; any proficiency in evaluating musical performance could be considered as a good sign for developing music creation tools.

Finally a system is implemented for the generation of rhythm. The rhythmic metrics are used as tools to generate new rhythms based on their similarity to a rhythmic pattern. One rhythm is used as input reference to the system and new rhythms are created controlling via sliders the similarity to the reference. This system could be used as an expansion to the common sequence-based rhythm machines. The implications of developing the system are described, explaining the dataflow and the details of the system's architecture. The results from the implementation stage show how metrics are useful to generate new patterns maintaining the distances set by the users. However, in some cases more than one pattern is retrieved instead of a group of patterns due to the resolution of the metrics versus the amount of possible patterns. This leads to the need to use new variables in the algorithm in order to get more precise and accurate results.

The overall outcome of this work shows that rhythmic similarity metrics proposed in the literature are not related in a simple way. Nevertheless, they are valid and useful in different areas such as genre classification scenarios and in the broad generation of new rhythms based on a reference. However, it is also shown that the actual implementation of these metrics is not completely developed for the task of defining one single rhythm based on perceptual directions. Based on these findings, further research objectives in this direction are suggested.

2. STATE OF THE ART

2.1 Introduction

The first part of this work consists of reviewing relevant literature of contemporary research on rhythm. Broadly, two main areas are identified. One is the extraction and analysis of rhythmic data from audio signals using MIR techniques. The second one is a symbolic analysis and transformation of rhythmic data (possibly extracted after using MIR techniques) based on music cognition theories. Extraction of rhythmic parameters from audio is still a difficult and unsolved task, therefore the rest of this work will be centered on pre existing symbolic representations or extracted from scores, MIDI files or sequencers.

2.2 Symbolic analysis of rhythmic patterns

Symbolic analysis of music and sound requires measuring the features of interest in order for them to be used in logical and mathematical operations. Once a feature is extracted from a sound, it is generally used to measure distances to compare, group or separate with other sounds. Particularly, rhythmic analysis of sound requires first the abstraction of its percussive elements into simple representations; second, to use those representations to create meaningful descriptors; and finally, to use appropriate metrics to measure distances between descriptors. These three aspects of percussive analysis (representations, descriptors and metrics) are still an open issue and will be described in this section.

a) Measuring distances among string representations

One useful way of representing musical information in a symbolic way is strings. Orpen and Huron (Orpen and Huron, 1992) converted music fragments to strings to analyze them with computer programs in search of patterns. Describing their analysis they theorize how exact replicas of a pattern are almost nonexistent in real life cases. So the notion of similarity, as an alternative to exactitude, is explored within musical concepts of melody and rhythm. The technique proposed to measure similarity of musical strings is based on the edit distance, particularly the Levenshtein version. A thorough exposition of the edit distance algorithm and its many versions is discussed on their text. Examples on how to analyze melodic and rhythmic similarity are given with Bach sonatas. However no correlation between their analysis and any music cognitional property is explored.

Following the use of strings in musical analysis, Toussaint focuses specifically in the analysis of rhythm, comparing different similarity measures among them (Toussaint, 2004). He compares Hamming, Euclidean interval vector, interval-ratio, swap and chronotonic (Gustafson, 1988) distances using African 4/4 clave patterns. A phylogenetic analysis of the set of patterns using the different distances is carried out arguing that, at a pure mathematical level, DNA and rhythmic patterns are coded as sequences of symbols so the use of genetic methodologies is valid. The results of the quantitative phylogenetic analysis are compared with the classifications of the same set of patterns done by classic qualitative musicology. The results suggest that the chronotonic distance (distance between chronotonic representations of two rhythms) is the best suited, among the ones used, to explain the musicological relations of parenthood among rhythms. On a follow up paper, Toussaint uses geometry to analyze strings of melodies, rhythms and chords. Although the text is not conclusive or experimental, he proposes clever tools to represent rhythms and explores different distance measures (Toussaint, 2010). According to the text, Hamming distance is attractive for pattern comparison because of its fast computation, but it is not appropriate for measuring rhythm dissimilarity because it does not measure how far the mismatch between two corresponding onsets occurs. The author suggests, from a computational point of view, the *fuzzy* hamming distance to measure the dissimilarity between rhythmic patterns, while preserving a low computational time. Swap distance (the pairwise disagreements between two ranking lists) is discarded as it is only useful with patterns with the same amount of onsets. However a "swap", the main operation of the swap distance, is interesting because is one of the three operations used to measure the edit distance.

The edit distance (the Levenshtein version of the algorithm) measured between two rhythmic patterns has been successfully correlated with human measures of rhythmic similarity (Post and Toussaint, 2011). This correlation is based on two experiments: One aimed at testing rhythms that are similar and how human perception of that similarity is correlated with the edit distance. The other experiment explores the use of the edit distance in predicting subjects' dissimilarity among rhythms. On both cases the edit distance was able to predict the perceived distance rated by the subjects, better than the swap distance and the measure given by the GTTM. The correlations found are of 0.64 for the first experiment and 0.82 for the second experiment. However, there were some interesting failures on the prediction: In the similarity experiment (experiment I) the edit distance failed to account for a 4/4 accentuation, because it is not related to metric hierarchies. So it remains an open issue how to account for this perceptive 4/4 accents. On experiment two, there were also some discrepancies between perception and the edit distance. This time the perceived distance ratings of two rhythms was opposed to the one measured by the edit distance. An open issue in this study is the use of more complex strings that could include accents.

b) Measuring distance using syncopation

A foundational study, cited in almost every further text on the topic, tries to explain how humans generate a metric sense out form a sequence of events (i.e. musical notes). The text suggests a correlation between a symbolic descriptor of a percussive pattern called Induction Strength and the ease with which subjects understand (or tap to) a rhythmic pattern (Povel and Essens, 1985). The authors propose a computer algorithm that tries to simulate the propensity of a pattern to induce an internal clock on a listener. It is assumed that if the clock is induced it will be used to specify the temporal structure of the pattern. The clock in the model has a hierarchical time that can be subdivided in portions (such as we do on everyday life with time of day, hour, minute, second). The unit of time is flexible and is adapting to the sequence that is being considered. Part of their theory is based on the fact that a single sequence of unaccented onsets induces a certain perceptual accentuation in some of the onsets according to their position within the pattern: (1) an isolated tone is perceptually accentuated, (2) the second onset of a cluster of two onsets is perceptually accentuated, (3) the initial and final onset of a cluster consisting of three or more onsets are also perceptually accentuated. So subjects perceive different rhythmic (non loudness) accentuations within a pattern generated by the positions of the onsets. Using this rhythmic accentuation, the authors generate rhythmic sequences and try to guess the “best” *internal clock* for understanding the pattern. The generated sequences are clustered according to the induction strength. They run 3 experiments to understand how humans code rhythmic patterns, specifically assessing the idea of the Internal Clock. They find that when a constant time cue is given, subjects are more accurate to reproduce an induced pattern. On every experiment they divide the patterns according to their induction strength. They prove that the lower the induction strength, the harder (takes more time) for subjects to reproduce it correctly.

Another well known source for symbolic analysis of music in general is the General Theory of Tonal Music (GTTM) by Lerdahl and Jackendoff. On the Metrical Structure section of the book the authors propose the use of weights or accents on certain onsets of a musical phrase depending on its metric (Lerdahl and Jackendoff, 1983). For example, on a 4/4 measure the beats that fall on the smaller even divisions have higher weights than the higher ones. See example on Figure 1.

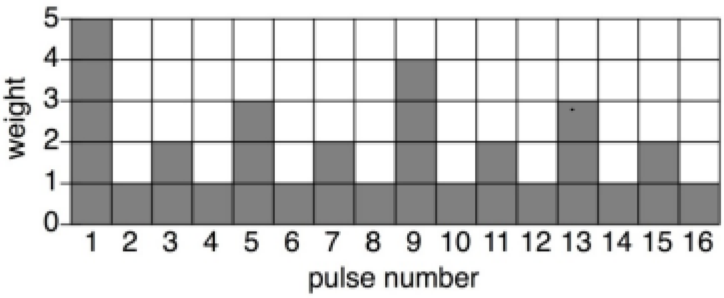

Figure 1: Weights for different pulses based on the GTTM

Although their rhythm approach is theoretical, it resembles the ideas of Povel and Essens. Also, as exposed by Post and Toussaint, one discontinuity present on their results of their second experiment (measuring dissimilarity) could have been avoided if GTTM weights were used instead of the edit distance.

Building on top of the meter induction theory and on the weightings of percussive events, *syncopation level* is proposed by Longuet-Higgins and Lee in 1984. This level is based on assigning weights to the notes on a musical phrase based on their rhythmic relations with their preceding notes. Fitch and Rosenfeld make a very good example of how syncopation is measured (2007).

Syncopation level is related conceptually with rhythmic complexity and then used for perceptive experiments by Fitch and Rosenfeld (Fitch and Rosenfeld, 2007) and by Ladinig (2009). Fitch and Rosenfeld use the syncopation metric to explore how humans memorize and reproduce rhythms with different syncopation levels. Their methodology consists of measuring micro timing accuracy of reproducing a rhythm by tapping. Rhythms were induced in subjects and reproduced immediately for one group, and with a 24 hour difference for another group. As a support for the usefulness of the syncopation level, they report a correlation between the syncopation level of a musical phrase and the difficulty for the subjects to reproduce a rhythm. Ladinig measures the syncopation level of patterns using different weight profiles.

Another relevant project that uses syncopation level for practical purposes was developed by Tutzer in 2011. Syncopation level is measured for each onset of a rhythmic sequence and then a histogram with the counts of each level is created. The author proposes measuring the distances between histograms as a measure of how different their syncope is. That measure is used as a rhythm similarity distance. His experiments show that his method, the metric weight histogram predicted with an accuracy of 76.6 % of human perceived similarity.

Recently, a variation of the syncopation level metric is used (Witek et al. 2014). In this work they use a polyphonic drum pattern composed of three instruments (bass drum, snare and high hat) and correlate their polyphonic syncopation with the 'desire to move' and the 'pleasure' experienced by subjects while listening to 4 repetitions of the same drum pattern. One of the important contributions of this project is the extension of syncopation level from the monophonic version on Longuet-Higgins and lee to a new polyphonic measure.

Thul and Toussaint present a report on the comparison of different complexity measures and their cognitive significance. (Thul and Toussaint, 2008).

On the most recent text (Cao, Lotstein and Johnson-Laird, 2014) propose the grouping of rhythms by families. Rhythmic families are those that share a same kind of syncope within a given time detail. The novelty is the way they measure syncope: every note or silence gets assigned a symbol (N, S or O) depending on their relation with the beat (N if the event reinforces the beat, S if it is syncope, O if it is neither). A rhythmic pattern can then be represented by a sequence of N, S or O. These sequences are the families. They theoretically explain how a rhythm with different IOI (inter onset interval) can belong to the same family. Their experiments are aimed to measure the perceived similarity between two rhythms varying the families or the IOI independently. Their results show that families can be more accurate than the edit distance because it improves certain aspects that the edit distance misses, specifically meaningful musical attributes such as syncopation. In general they show that the greater the distance between the families of two patterns, the greater the perceived distance between them. As a technique it is similar to the Syncopation Histogram mentioned above as both methods are based on measuring syncopation. The families maintain information of the order, but miss the strength of the syncopation. The syncopation histogram is more detailed between the types of syncopation but fails to record the order of the events.

As the reviewed literature suggests, the theories of rhythm cognition, their descriptors and metrics are still under development and results are promising but not definitive. The combined use of existing descriptors and metrics in order to improve the results is an idea that will be explored further in this text.

c) Other types of measures

When choosing the right distance to measure relationships among musical strings, different authors have explored the use of distances coming from areas unrelated to music. These explorations of different metrics include the Earth Mover's Distance (EMD) and the Proportional Transportation Distance (PTD) for melodic similarity

(Typke et al, 2003). They successfully use EMD and PTD within a dataset to find melodic occurrences outperforming previous approaches. Wiering, Typke and Veltkamp explore further both distances and formalize their application for querying large data sets of encoded melodies (Wiering, Typke and Veltkamp, 2004). In general, a good compendium on measuring distances using strings can be found in Cha and Srihari (2002) and Deza and Deza (2009). As an attempt to characterize measures and success for different distance measures, Table 1 presents a list of algorithms found throughout the literature review.

Table 1. List of the algorithms that have been used to measure rhythmic similarity among binary sequences found throughout literature review.

Name	Origin	Description	Authors	Correlation with perception
Rhythm family	music cognitions	Measure the syncope of pattern, assign characters to types of syncope, define families as strings of characters	Cao, Lotstein, Johnson-Laird	**possibly higher than edit distance**
Edit distance	strings	Amount of transformations applied to one pattern to become another	Levenshtein Damerau	**0.64 and 0.82**
Syncopation level	symbolic perceptual	Measure syncope of pattern. Measure frequencies of syncope.	Longuet-Higgins and Lee, Ladinig, Tutzer	**76.6%**
Euclidean	geometry		Euclid	-
Cosine	geometry		public domain	-
hamming	math	Minimum number of errors that could have transformed one string into the other.	public domain	-
and	math	Dot product	public domain	-
earth movers	math	Distance between probability distributions	public domain	-
swap	strings	Bocks to be swapped to achieve another block distribution (only works with patterns of same onsets)	Levenshtein	-
Chronotonic	Phonetics	Distance between sounds (Accounts for musicological relationships among patterns)	Kjell Gustafson on algorithm by Hofmann-Engl	-

2.3 Analysis of rhythmic patterns based on audio files

After reviewing different symbolic approaches to rhythm cognition, signal based approaches will be revised as an attempt to complete the picture of contemporary rhythmic analysis. The signal based research field is focused on extracting meaningful rhythmic data directly from audio signals and files. Most of this work is based on DSP and MIR techniques that draw useful percussive information from audio files and signals. The methods and metrics used are quite different from the symbolic ones and will be presented below.

In 2002 self similarity within audio files was used to define their BPM (Beat per minute) and to calculate rhythmic similarity (Foote, Cooper and Nam, 2002). The authors methodology is based on the *beat spectrum* (Foote & Uchihashi, 2001) a measure of acoustic self-similarity versus lag time, computed from a representation of spectrally similarity. Peaks in the beat spectrum correspond to major rhythmic components of the source audio. Although the spectral method proposed is useful as an abstract metric, the similarity proposed by the authors is focused on grouping songs together and not in reproducing any perceptual phenomena. This limits the scope of the method and the use for musical purposes. Even more, the metric used to determine the distance that defines the rhythmic similarity remains an open issue. In the authors words "Though there are many possible distance measures, it is not obvious that any will be at all correlated with perceptual differences. Thus it will be important to show that small 'distances' correspond to rhythmically similar music, and that larger distances are correlated with decreasing rhythmic similarity". In order to explore the effect of the distance metric, two experiments are carried out: the first one demonstrates the positive correlation of the beat spectrum with the BPM using the Euclidean distance. The second one tests three different distance metrics (Euclidean, cosine and Fourier beat spectral coefficients) to cluster 15 fragments of music extracted from 5 different files, based on the beat spectrum. The results for the second experiment show how more than 95% of the files were correctly clustered with both the cosine and the Fourier coefficients. These two experiments confirm that the beat spectrum can be useful in automated audio library browsing, but do not account for human perception. More so, it lacks any musical or symbolic explanation of what information within the files makes them similar. Never the less, the importance of research in metrics is stressed as the next step to follow.

A more recent study on the same area presents a more complex method for measuring rhythm similarity (Holzapfel and Stylianou, 2009). The authors use scale transforms to compare rhythmic features between audio files. Again, metrics are suggested but not supported under any specific perceptive argument; they rely on the positive clustering results obtained with different datasets. The highlights of their approach consist on the improvement of classification results over other approaches on same datasets. Also they claim that their method is very robust when comparing similar files in which tempo is very different.

A different view on MIR analysis of sound files for rhythmic similarity is taken by Paulus and Klapuri (Paulus & Klapuri, 2002). Unlike the previous authors, they generate their dissimilar percussive data set in a very clever way: subjects with different musical skills are asked to perform the same rhythm with different percussive sets and their performances are recorded. This is done for three different rhythms with three different sound sets. Their experiments are aimed at using their system to classify correctly the different recordings of each rhythm. The authors claim that their system was able to give high similarity ratings to the recordings of the same rhythms even though they were performed by different subjects and with different percussive sets. These results are achieved using as a metric the normalized spectral centroid of the samples, which is a product of the spectral centroid and loudness. The creation of the dataset for the experiment is interesting because rhythmic similarity is implicit in the reproduction of the rhythms by the subjects. There is a bias in the fact that subjects may recognize the rhythm but may not perform it properly. So his evaluation, although a valid one, is made against the similarity of interpretation and not of solely of perception. Therefore, no correlates to perceptual features are claimed.

A more complete work on the area of MIR for rhythm similarity is done by Smith in 2011. As Tutzer (discussed in the previous section), he defines his metrics of similarity based on rhythm complexity (syncopation levels) as proposed by Ladinig. Additionally he measures meter, grouping and tempo as features for the comparison. The distance measures used are Euclidean and Cosine. His MIR method to derive the measures is based on detecting the tempo based on energy peaks and computing energy in two non overlapping spectral bands (60-100Hz and 3500-4000Hz). To evaluate his method he uses a ballroom dance dataset with a support vector machine, building a machine learning algorithm with a fragment of the data set and then using it against the rest. His results are 67% of accuracy against a database of 458 rhythms. Defining the general approaches of the MIR community towards rhythm and onset detection he points out that an untested solution could be based on the statistical reformulation of the Longuet-Higgins and Lee model.

2.4 Conclusions

Both major areas MIR analysis and symbolic analysis are complementary. The first, seeks to extract the symbolic information correctly from signals, and the latter to meaningfully decide what to do with that symbolic information. The ideal scenario would be to go directly from a sound wave to a rhythmic symbolic interpretation of it, but as it is reviewed above this is far from completion. Extracting a rhythmic representation from an audio signal still leads to errors and gaps from a musicological point of view.

As presented on Table 1, the *edit distance*, the *syncopation distance* and the *rhythmic similarity distance* are the principal elements for evolved rhythmic similarity theories. All theories advance in parallel but there is no definite way of selecting one over other. Apparently all of them are partially correct although they both seem to address very different perceptual mechanisms. Considering a combination of them to improve similarity predictions, or even the comparison between the results generated by them are important issues to address in the near future.

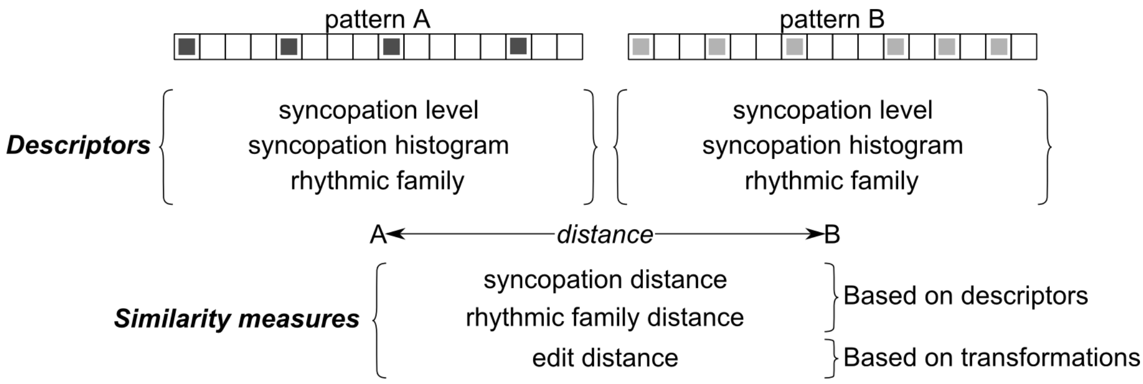

Figure 2: Scheme of symbolic rhythmic information based on music cognition.

As an additional source to give a structure to the findings, a summary of descriptors and distance measures derived from the symbolic approach to rhythm is presented on Figure 2. There are some descriptors that can be extracted directly from a rhythmic pattern: the syncopation level, the syncopation histogram and the rhythmic family. Distance among two different patterns is based on similarity measures, some based on the descriptors and one based on transformations needed for one pattern to become the other. All three descriptors are based on the *syncopation level* thus the syncopation distance and the rhythmic family distance are too. All similarity measures but the edit distance, come from the same root. This structure and these metrics are going to be used throughout this work.

Even though it is not clearly mentioned by any of the authors that work with binary strings of onsets vs. silence, the symbolic notation of ones and zeroes is extremely limited compared to a human musical performance. A distinction of at least three dynamic levels to compose the strings (silence, medium accented onset and hard onset) could be explored in order to account for more real life scenarios. A good panorama for this approach is that metrics and methods for comparing strings are robust and are used to work with higher complexities.

3. METHOD
3.1 Introduction

As stated on the previous section, the approach selected for the development of the percussive musical application is the music cognition approach. The basic structure and elements needed for such an approach, based on the literature revised, is presented on Figure 2 and Table 1. These cognitive models are being used with two data sets. One is the group of all possible rhythmic patterns consisting of onsets and silences that have duration of 16 steps. The other is a group of songs in MIDI format each of them belonging to a certain musical genre. Both data sets are constructed during the course of this work.

Three activities are developed that use both the algorithms and the data sets First, the similarity measures resulting from the previous chapter are used to compute the distance from a reference pattern to all possible rhythmic patterns with 16 steps. Second, the descriptors and metrics are used as the only method in a genre classification task. Third, the cognitive model is used to extract from the data set of all possible 16 step patterns, patterns that are at certain distances from one reference rhythm created by a user. As Post and Toussaint (2011) state, at the heart of a symbolic percussive application there is always a metric used to analyze the closeness rhythmic patterns. These activities will be addressed in detail in the following chapters.

Results generated by these activities are measured using specific techniques suited for the type of data and the type of observation desired. For the first activity the focus will be in determining the amount of clusters generated by each distance metric given the different reference conditions. For the second activity, the measurements will be focused on the amount of songs and the different genres correctly clustered by the algorithms. For the final activity the measurements will be focused on the performance of the algorithm, specifically its proficiency in delivering a single resulting pattern from a single search operation.

It is important to state that the algorithms of the percussive descriptors and the distance metrics they have variations and that imply decisions to be made. These decisions and their implications will be explained in the following sections.

3.2 Implementing the descriptors and distances

The implementation of algorithms and software in this work was done using the Pure Data extended platform. As the major objective of this project s to create an interactive application, the Pure Data Extended platform is chosen for its fast prototyping and real time interaction features. The use of the Extended version with external libraries is preferred over the 'vanilla' version because of the additional computation tools for dealing with arrays and other data types such as lists.

a) Descriptors

The syncopation distance is based on the syncopation level proposed by Longuet-Higgins and Lee (1984). They propose a weight profile for the different onsets of a rhythmic pattern located at different subdivisions of the beat. Ladinig (2009) proposes a variation, using different profiles given the background of the evaluated subjects. The syncopation level is the weight difference between a note and the silence after it. According to the L weight profile (the originally proposed by Longuet-Higgins and Lee) there can only be eight different syncopation levels (See figure 3). Notes followed by silences on positions having negative values are not considered syncopations (i.e. note previous to a silence in uneven positions). Notes followed by silences on positions having positive values are considered syncopations (i.e. notes previous to a silence in even positions). The syncopation histogram is an eight sized vector with the sum of every value of syncopation available. The types of possible syncopations are in a range from -4 to 4 excluding the 0 value and are shown on top Figure 3.

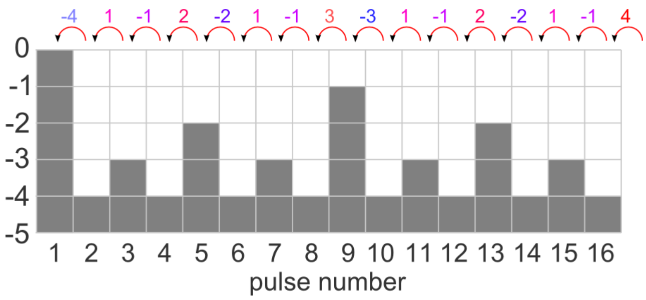

Figure 3. Syncopation L weight profile (grey) and syncope value for different note positions followed by silences (top in color). Negative values are for note positions followed by silences that are not syncopated, positive values for syncopated positions of notes followed by silences.

As proposed by Cao et al. (2014) another alternative arrangement based on the syncopation level is the syncopation family. It consists of a list of characters, each character representing if on a section of a pattern a positive, negative or no value of syncopation was found. In the original implementation by the authors, the resulting list is based on letters, S for positive syncopation, N for negative and O for none. In the implementation for this work numbers are used instead of letters for the further convenience of computing the distances using numerical methods. In this implementation, the rhythmic pattern is divided in sections of equal length; and the level of syncopation of each section is reported as: 1 if the syncope value is positive, -1 if the syncope value is negative and 0 if neither of them.

As an example, the syncopation level, the syncopation histogram and a syncopation family of the son pattern 1 0 0 1 0 0 1 0 0 0 1 0 1 0 0 0 are computed. The syncopation level, using on the L weight profile is -4 0 0 2 -1 0 0 0 -1 0 -2 0 0 0. Based on this, the syncopation histogram is the count of the different types of syncopation which can be -4, -3, -2, -1, 1, 2, 3 and 4 respectively. The count of different syncopations in the son pattern is: one syncopation of type -4, one syncopation of type -2, two syncopations of type -1, and one syncopation of type 2. There are no syncopations of type -3, 1, 3 or 4. This leads to a syncopation histogram of 1 0 1 2 0 1 0 0. The syncopation family is the place (not the count) where syncopations are found. In the case of the son it is -1 1 -1 -1 0 -1 -1 0 using a resolution of 1/8 of note. The first two steps have a negative syncopation (-4 + 0), the next two have a positive value (0 + 2), then negative (-1 + 0), then zero (0 + 0), then negative (-1 + 0), then negative (-2 + 0) and then zero (0 + 0).

To compute the syncopation level of a rhythmic pattern, a Pure Data abstraction is programmed. That abstraction uses the L weight profile.

Based on the syncopation level abstraction three descriptors are programmed:

- The *syncopation level sum*: The sum of all the syncopations present on a pattern.
- The *syncopation histogram*: A histogram counting the amount of each syncopation level found in a rhythmic pattern.
- The *syncopation family*: A sequence of numbers representing the places on a pattern where positive, negative or non syncopation is encountered
- The *density*: The sum of all the onsets present on a pattern.

b) Distances

The syncopation distance used in this work is defined as the distance between two syncopation histograms, based on the work performed by Tutzer (2011). As exposed on the previous chapter, this distance can be measured by different means such as cosine, Hamming, Euclidean or other distance measure. The first section of the next chapter is dedicated to understanding the impact of the different metrics when computing this distance. All the objects used to compute the distances were also programmed during the development of this work.

Another type of rhythmic similarity metric, but not based on syncopation it is the edit distance (Post and Toussaint, 2011). The edit distance is a measure of how many transformations should be done in one string to become the other. The possible transformations are deletion, addition and substitution. As an algorithm, it is based on iterating through a matrix that has rows and columns the same size as the size of each string. Iterations compare between one element of the string and the others. The result is the last element of the matrix which has accumulated the total amount of transformations. An object for Pure Data was programmed in C and then compiled to improve computation (https://github.com/outer-space-sounds/edit-distance). In this work, distances between rhythmic patterns are computed either using Euclidean, Hamming or cosine distance between syncopation histograms or family vectors, or are computed using the edit distance.

3.3 Creation of the data sets

a) The all pattern universe

The set of all possible parameters consisting of any number of onsets within a 16 step grid is created using a position shifting algorithm. This algorithm displaces the rightmost onset on the pattern one step on every iteration. When the rightmost onset reaches the last available step of the grid on the right, the onset on the left (if available) shifts one position and iteration starts again. Implementing and putting to work this algorithm generates a total of 65.334 patterns. They are presented on Table 2 separated by the amount of onsets present on the pattern.

Table 2. Number of onsets and amount of possible combinations on a 16 step grid.

Onsets	Possible combinations
1	16
2	120
3	560
4	1820
5	4368
6	8008
7	11440
8	12870
9	11440
10	8008
11	4368
12	1820
13	560
14	120
15	16
Total	65534

b) The MIDI song data set

To create the MIDI song data base, a group of songs were downloaded from the Internet via free sharing MIDI file sites. All the songs belonging to the Latin genre were downloaded from http://lacubanaza.tripod.com/ and the rest of the songs were found in http://www.cool-midi.com.

Table 3. The twenty four songs used.

Genre	Artist	Song
Latin	cheo feliciano	el ratón
	orquesta matamoros	son de la loma
	tradicional	Guajira
	oscar de león	Llorarás
	ruben blades	pedro navaja
Hip-hop	2pac	Changes
	50 cent	candy shop
	50 cent	in da club
	50 cent	Pimp
	snoop dog	drop it like is hot
Pop	madonna	like a virgin
	madonna	into the groove
	madonna	Celebration
	justin timberlake	rock your body
	justin timberlake	Señorita
	michael jackson	beat it
	michael jackson	billy jean
Rock	acdc	highway to hell
	guns n roses	welcome to the jungle
	nirvana	smells like teen spirit
	radiohead	karma police
	rollingstones	anybody seen my baby

It is important to say that the quality of the files was assessed just by listening to the midi file. They were not thoroughly compared with the original scores. So errors in this experiment may have been induced due to the divergence of the MIDI files from the recorder originals.

Every MIDI file was edited to make sure the bass track was track 1 and beginning of the file was on the first measure. A list of the used songs and the genres can be found on Table 3. The genres used were Latin, Hip-hop, Pop and Rock. The selection of the songs in each category is based on a subjective classification.

3.4 Measurements

a) Assessing distance metrics

This activity is designed to explore the different types of organizations provided by the different distance metrics when clustering all the possible 16 step patterns. The outcome of this activity is a count of how many different clusters is provided by each metric in order to classify a very large data set (such as the 65.534 possible patterns). It is of interest for the interactive application that the number of clusters is big and that patterns are more differentiated among them by a certain metric. This would help individualize patterns and extract the smaller amount when recalled using this parameter.

b) Genre classification exercise

In order to complete the classification task of the MIDI files the rhythmic information of the bass track in every MIDI file has to be measured. The bass track is split in groups of 16 semiquavers and note onset information is used to generate simple patterns representations of ones and zeroes. The syncopation level, the syncopation histogram, and the density (as explained in section 3.2 a) are measured in every pattern of 16 semiquavers of the bass.

As the rhythmic distances are measured between two patterns, using the methodologies discussed above, one reference has to be selected. The *son* pattern (1 0 0 1 0 0 1 0 0 0 1 0 1 0 0 0) is chosen as a reference and edit distance and the syncopation histogram distance are measured to it. This pattern is used as reference following its use the literature, specially the texts by Toussaint where he recurrently uses the *son* pattern as reference or example in different research experiments. Having selected the son as reference, the edit distance and the syncopation distance (Euclidean) are extracted for every bass pattern of each song. Additional statistical values are computed for each track: mean, standard deviation, maximum, minimum, flatness, dispersion of the two distances and the density.

In order to provide a better comprehension of the data, a distribution analysis is performed for the distances so we know how many different categories (differentiated rhythmic patterns) exist in a song and how dense (how many times each pattern is repeated). In other words, the frequency of each category present on the dimension is analyzed to create a histogram. On these histograms, its *flatness* (computed as the division of the geometric mean by the arithmetic mean) has been measured. Also the amount of different categories present in each song is measured for each histogram, this measure is called *dispersion*. So if a song has 3 different bass rhythmic patterns along the song we are able to tell how many times each one of them was played.

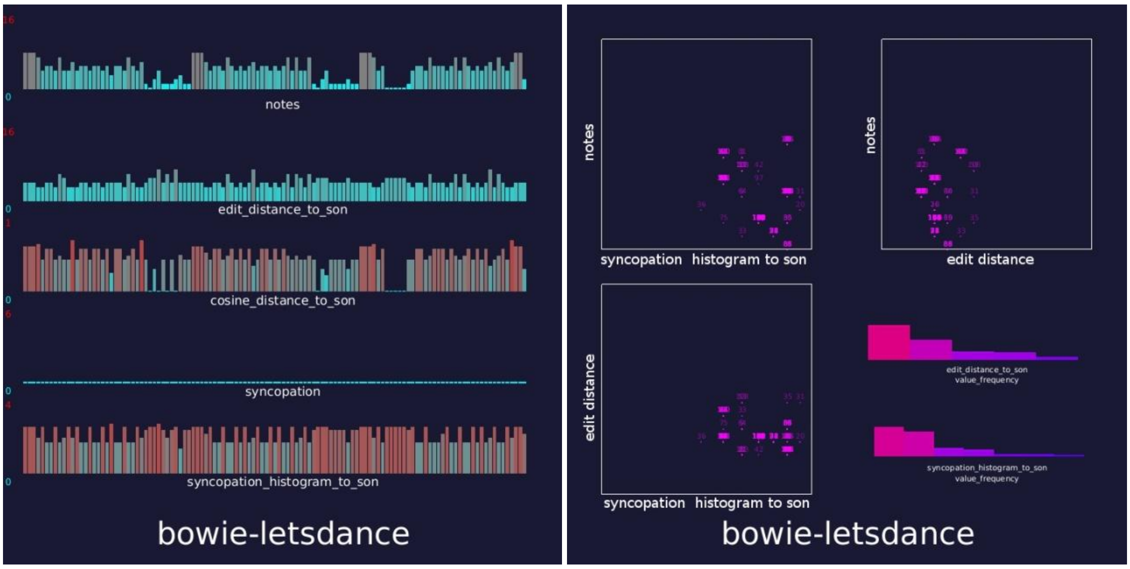

Figure 4. Data generated for the song "Let's dance" from David Bowie. The data on the left is the notes (Density), edit distance to son, cosine distance to son, syncopation and syncopation histogram to son measured for every pattern of the song. Time goes from left to right. On the right there are 2D plots of the same data plus the class analysis of the edit distance to son and the syncopation histogram to son. They are visualizations of how many different classes are on each set and how the data is distributed among them.

Each song generates a data summary of 18 values like this: Density [mean, sdev max, min, flatness, dispersion], Edit distance to the son [mean sdev max min flatness, dispersion], Syncopation distance [mean sdev max min flatness, dispersion]. Each analysis generated three specific two dimensional plots of the three measures (Figure 4 right): syncopation histogram to son vs. notes, edit distance to son vs. notes, edit distance to son vs. syncopation histogram to son.

4. DEFINING A GENERAL SPACE FOR RHYTHMIC ANALYSIS: EXPLORING ALL POSSIBLE PATTERNS

4.1 Comparing distances

Some experimental results presented on the state of the art explore the perceptual phenomena of rhythmic similarity by selecting few and specific rhythms to be used in experiments. Correlation with human ratings is then deduced statistically and this is how current metrics appear.

Although correlation is not on doubt, when this metrics are proposed for analysis or generation of rhythm they show several limitations. For example, the maximum possible outcome value of comparing two rhythmic patterns of 16 steps, via the edit distance, is between 0 and 16 minus the total amount of onsets. The possible amount patterns that have i.e. 8 onsets is more than 12.000. This means that if we organize all patterns of 8 onsets based on their edit distance to a given pattern, we would have at most 8 levels to grade their distance. This grading depth would be significantly smaller than the population of possible rhythms and insufficient to differentiate among many patterns. Many different patterns would fall in the same edit distance category. Perhaps a more distinctive metric could assign a different distance to each and every pattern. Table 2 presents the number of possible different rhythmic patterns to be created on a 16 step grid for each number of onsets.

As it will be seen below, the syncopation distance is also limited. The computation can be done via any distance measure between histograms so it potentially could produce real and continuous values (as opposed to the discrete results of the edit distance). But on practice results for Euclidean distance between syncopation histograms are discretized around certain specific values.

To illustrate the behavior of some of the metrics, a simple experiment was carried out. All possible rhythmic patterns of 16 steps were organized according to their distances to the son (1 0 0 1 0 0 1 0 0 0 1 0 1 0 0) and to a metronome (1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0) rhythm. The results are presented on figure 5. The *y* axis of each space is the amount of onsets present on the patterns; the *x* axis is the distance to the son or the metronome. The color of the elements is proportional to the distance values, white being the closest and black the farthest.

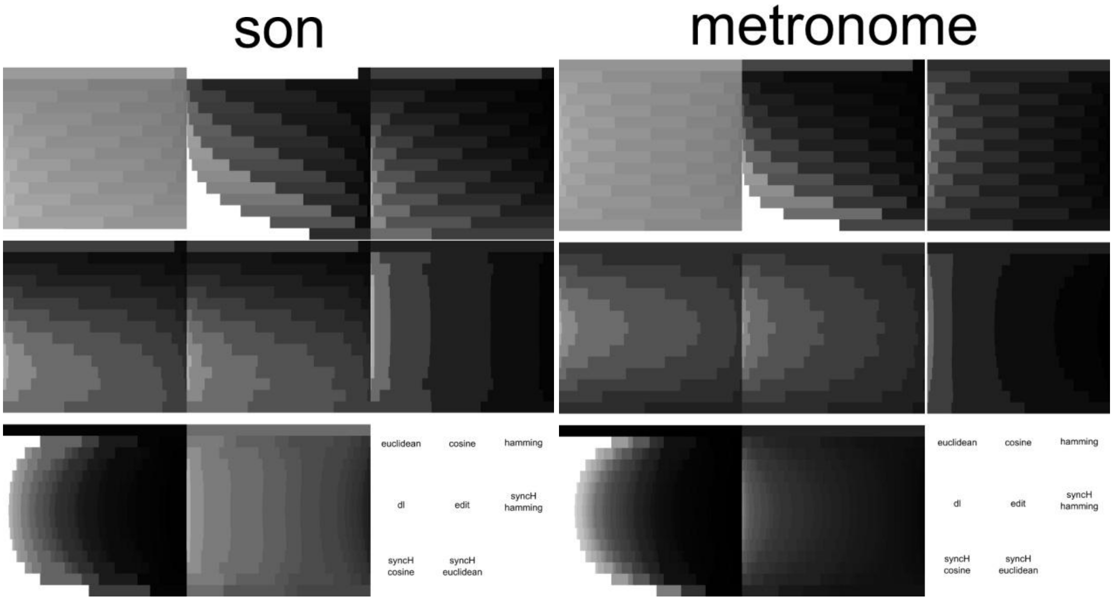

Figure 5. Distribution of all possible patterns in a 2D space given their distances to the son (left) and to the metronome (right). The distances used were Euclidean, cosine, hamming (top), Damerau-Levenshtein, edit, syncopation histogram hamming (center), syncopation histogram cosine, syncopation histogram Euclidean (bottom)

The graphics show that the Euclidean, cosine and Hamming distances (top of Figure 5) have fewer groups to cluster the rhythms than the syncopation distances measured with cosine or Euclidean algorithms (bottom). An analysis of the Euclidean syncopation distances to the son and to the metronome, reveal that all patterns are grouped in 42 different values. This means that the 65.334 possible patterns are scattered in 42 groups.

This is the largest number in which any of the metrics can segregate patterns. To approach the implementation of the interactive application defined in this work, it is an ideal metric. If it is used to extract patterns from a big set this would return the smaller sets of data, which would make further use of those resulting sets easier. For the rest of this work, the syncopation distance will be measured using Euclidean distance between the syncopation histograms and will be referred as *syncopation distance*.

4.2 Defining a general space for rhythmic analysis: edit distance vs. syncopation distance

To continue the analysis of all-possible rhythmic pattern spaces, the correlation between the edit distance and the syncopation distance was explored. In this case, the distances between all possible patterns and the metronome were measured. The results are plotted in a 2D plane where the color scale is the amount of patterns present at a given point. If the two metrics were orthogonal among them a shade of an even color would cover the space, meaning that all the points had a similar amount of patterns clustered. If the two metrics were completely correlated a white line in its diagonal would appear. As can be seen on Figure 6, none of these two extreme cases is present. Instead what we see is a semi correlation on the first lower third of the graphic and sort of orthogonality on the rest. On the right and top of the graphic, the band distribution can be observed showing that the upper half of the space contains most of possible patterns in a quasi normal distribution.

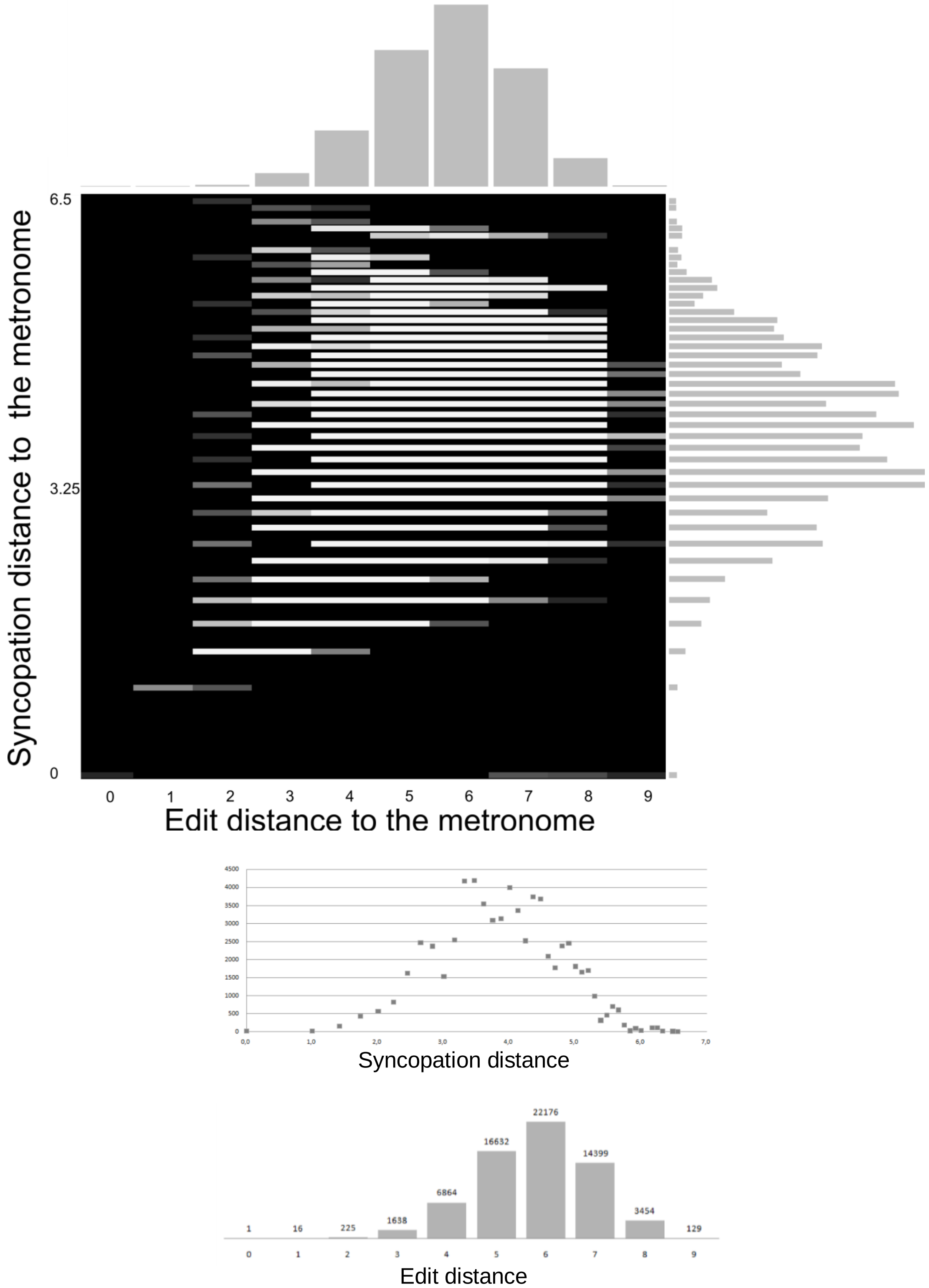

Figure 6. Distribution of all possible patterns on a plane limited by the edit distance to the metronome and the Euclidean distance to the syncopation histogram of the metronome (top). Distribution of patterns by the syncopation distance (center). Distribution of patterns by the edit distance (bottom).

This space shows that every point, except the 0,0 that has only one element (the reference), contains a rather big group of patterns, increasing towards the upper right section of the graphic. This fact affects the possibility of using this space for a simple interactive search and retrieve activity because at many patterns share the same edit distance and syncopation distance values.

As the current perspectives for interacting with rhythms via these two metrics is limited, there are two options. One is exploring deeper the perceptual phenomena in order to extract new and more accurate metrics that perhaps could lead to orthogonality. The other option is to use additional metrics, from the ones currently available, that allow searching in a meaningful way the dense zones of this 2D space. As the scope of this work is limited, the latter was chosen and is presented on the last chapter.

5. PERCUSSIVE SPACES FOR ANALYSIS AND CLASSIFICATION

5.1 Data analysis

This section presents the use of the two metrics discussed above to analyze the bass track of a collection of MIDI files and then to use that data to try to classify them. The MIDI files have been chosen to clearly resemble four different musical genres. Namely Latin Music, Rock, Pop and Hip-hop (see table 3). The aim is to test these rhythmic metrics as valid discriminators for high level aspects of music such as musical genres.

Using the different average values for each genre, several plots were generated to understand which of those values could be significant when trying to differentiate them. The criteria was to make 2D plots of all possible pairs of values and see which plots generated four dots (the data for each one of the four genres) the furthest apart.

a) Edit distances and syncopation distances

Dispersion

The Dispersion for both the edit distance and the syncopation distance seems to have a linear correlation (Figure 7, top). In detail, Hip-hop had an average close to three in both metrics, Pop almost 5 in both, Latin more than 5 in the edit distance but less than 5 in the syncopation distance and rock had almost 6 in both metrics. In general, these metrics seem to have a small correlation.

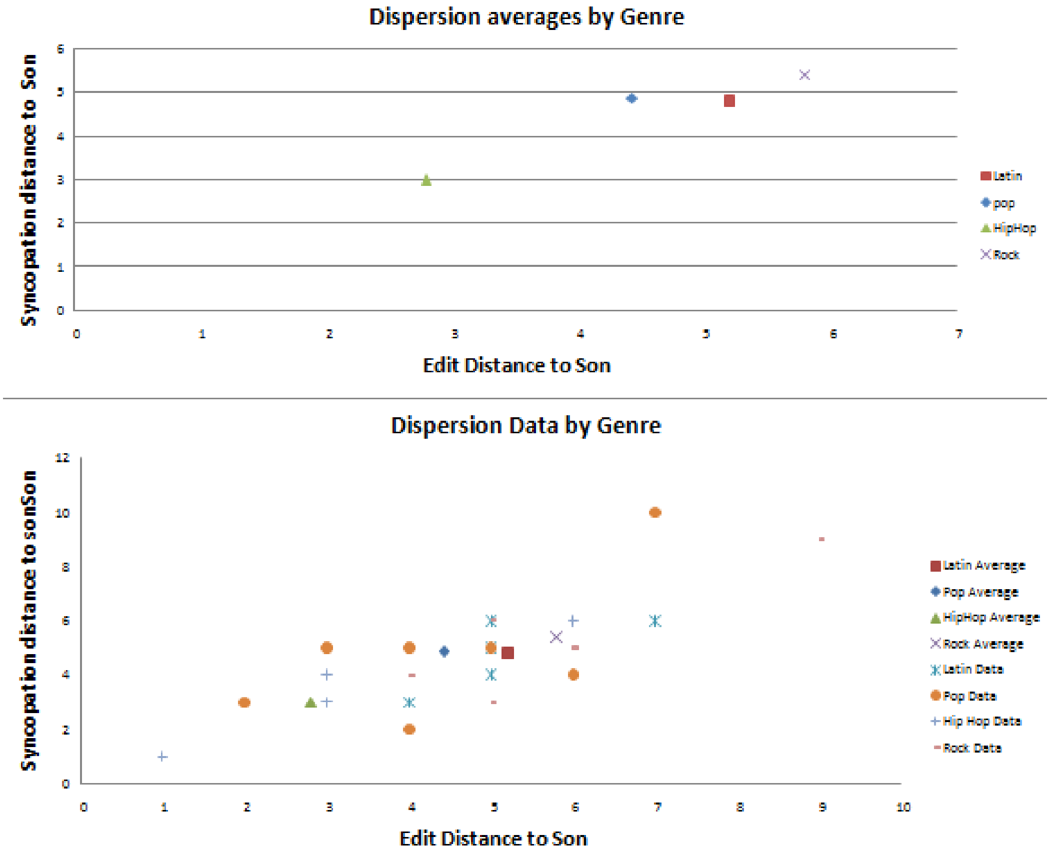

Figure 7. Dispersion averages of edit and syncopation distance for each genre (top) and dispersion data for all songs (bottom)

Flatness

The flatness averages of the edit and syncopation distances have a linear correlation. This means that both metrics counted the presence of different rhythms in a very similar way. In general, both metrics converge towards a pattern on the 2D space (see figure 8).

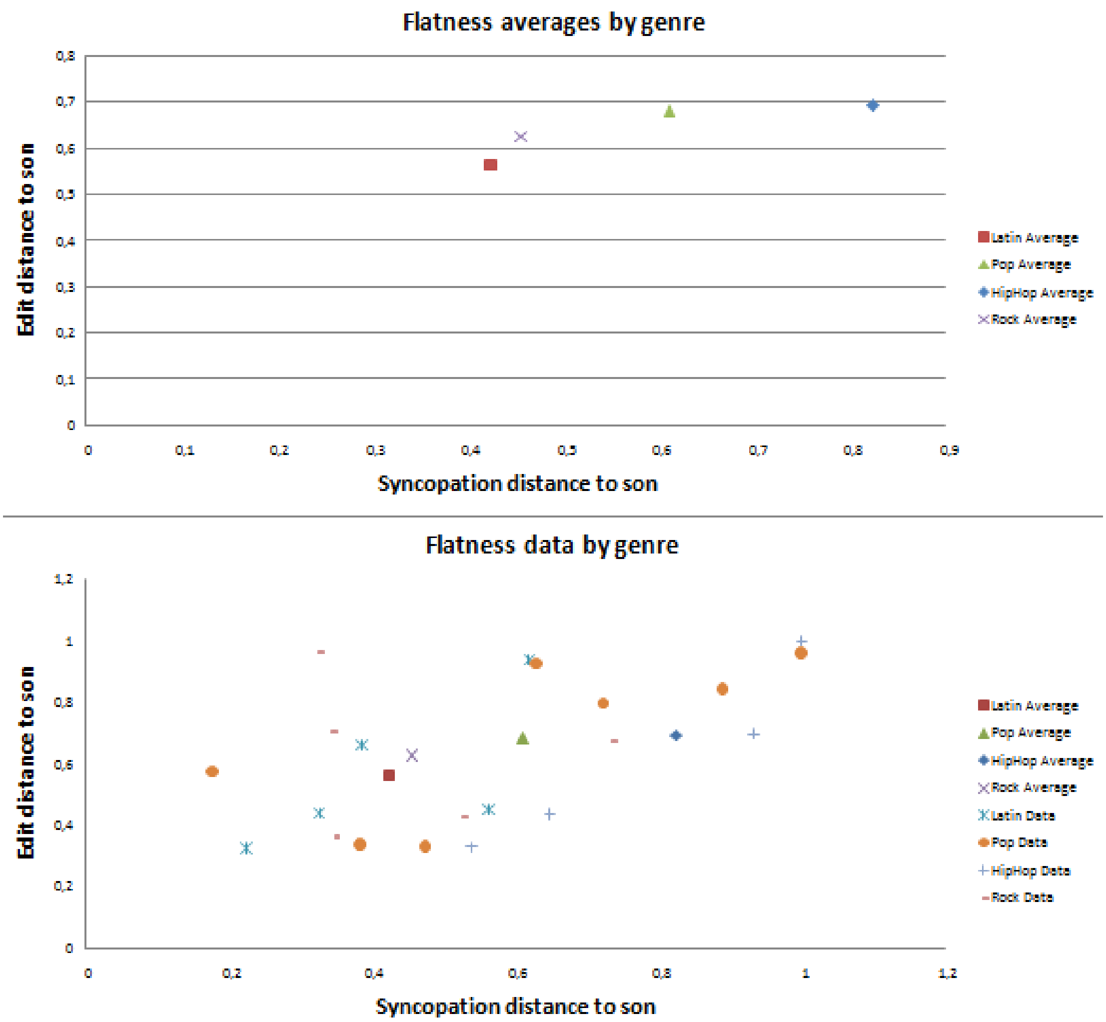

Figure 8. Flatness average for each genre (top), Flatness for each song (bottom)

Mean

The mean values for each category are plotted (See figure 9). Interestingly the mean values for each category show a strong correlation for the two metrics. A 1:2 correlation can be observed for every mean value of each category: 1.41 : 2.79 for the Latin genre, 2.01 : 4.21 in the pop genre, 2.01 : 4.22 for the hip hop genre and 1.9 : 4.2 for the rock genre. This roughly 1:2 proportion is clearly observed. A closer look to the data confirms this tendency for all the songs (Figure 4 bottom).

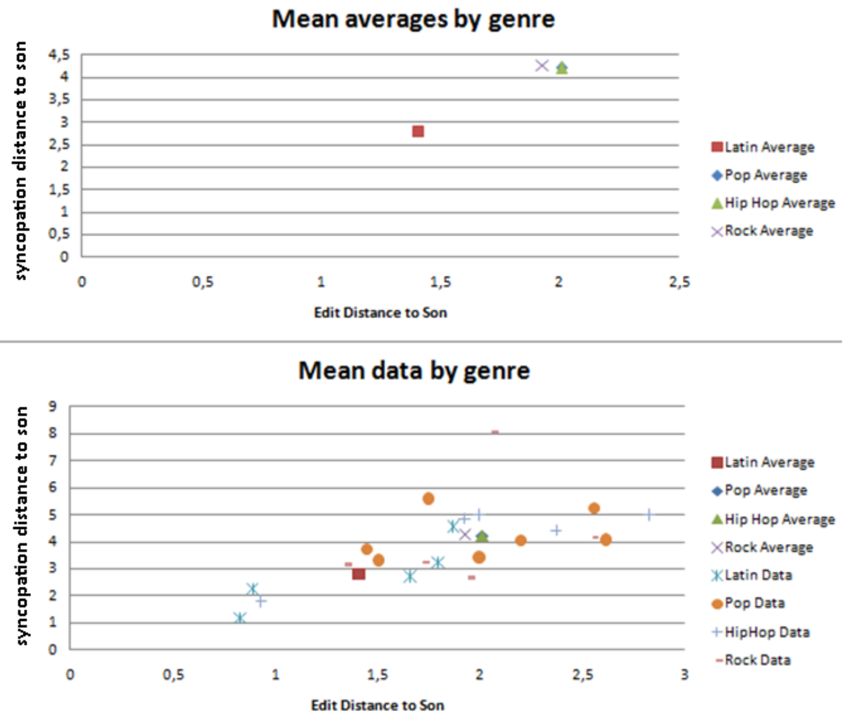

Figure 9. Mean values for all categories (top) and mean values for each song (bottom)

Standard deviation

The standard deviation of the edit distance and the Euclidean distance are also highly correlated. As can be observed in figure 10 (top) the relation between the means of each genre forms a very straight line. In general, the two distance metrics hold a strong correlation throughout the data set. Their means, standard deviation, flatness and dispersion are very much correlated.

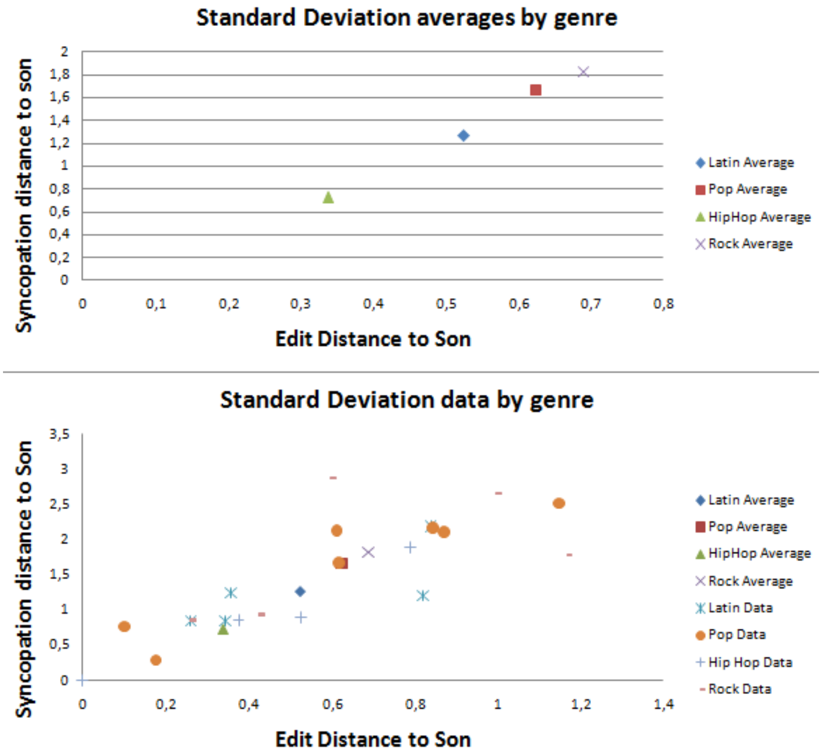

Figure 10. Standard deviations by genre (top), and by song (bottom).

b) Mixing variables

After checking for the relation of the same parameters of the two different distance metrics and noticing a similar tendency, the next step is to check if the different parameters can be a good source for segmentation. To do this, the combinations that are left are plotted. Each combination is analyzed searching for pairs of variables able to separate the averages of the genres within the space. Potentially, the combinations that spread the data the most would lead to a good segmentation algorithm. As can be seen in Figure 11 (bottom right), the best combination is the dispersion of the edit distance vs. the Standard deviation of the syncopation distance. This combination clearly spreads the averages of each genre, but most of all separating hip hop and rock.

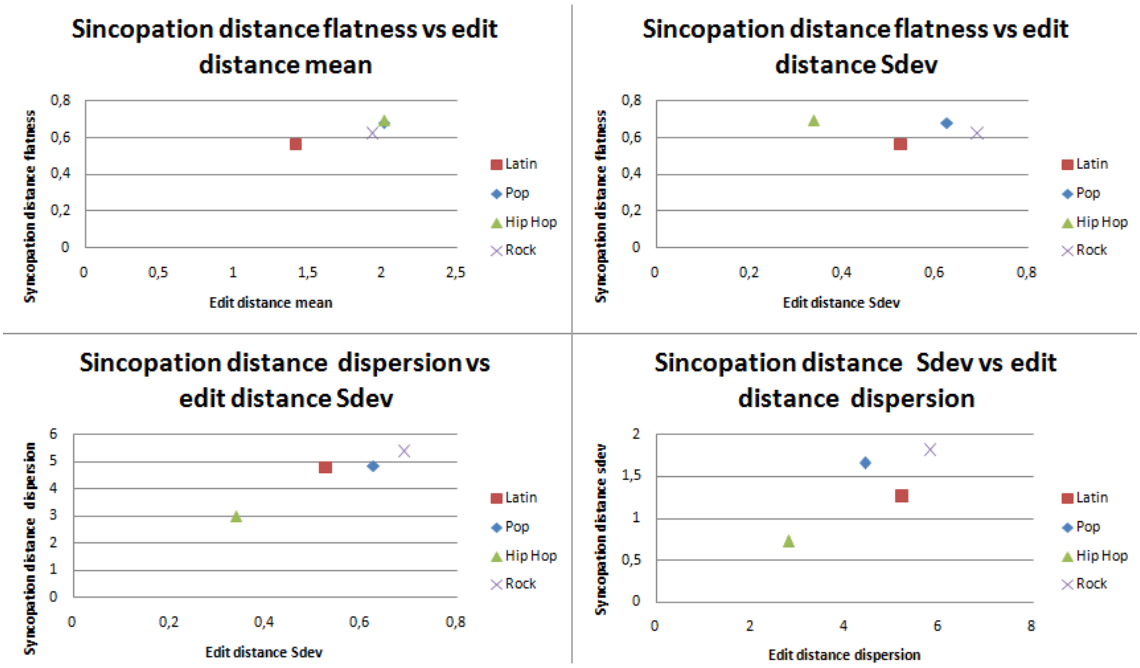

Figure 11. Data combination possibilities: syncopation distance flatness vs. edit distance mean (top left), syncopation distance flatness vs. edit distance standard deviation (top right), syncopation distance dispersion vs. edit distance standard deviation (bottom left), edit distance dispersion vs. syncopation distance standard deviation

In order to confirm the previous assumption, the data of all the songs is plotted for these two values. The result can be seen on Figure 12. A higher dispersion of the data can be observed. Pop songs are all over the place specially overlapped with Latin songs. The only possible discrimination from this pair of dimensions is Hip Hop from Rock. The result from eliminating the pop songs and the Latin songs is presented in Figure 13. With a function such as the one described by the black line on figure 8, all rock songs could be separated from 3 out of 4 Hip Hop songs.

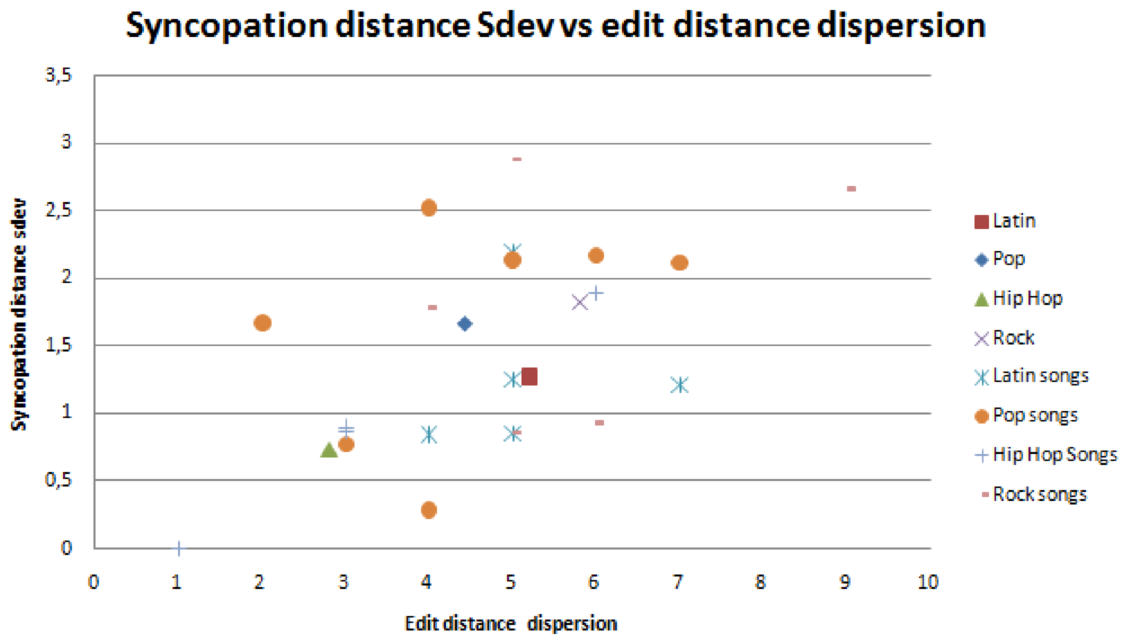

Figure 12. Syncopation distance standard deviation vs. edit distance dispersion for all songs and their genre averages.

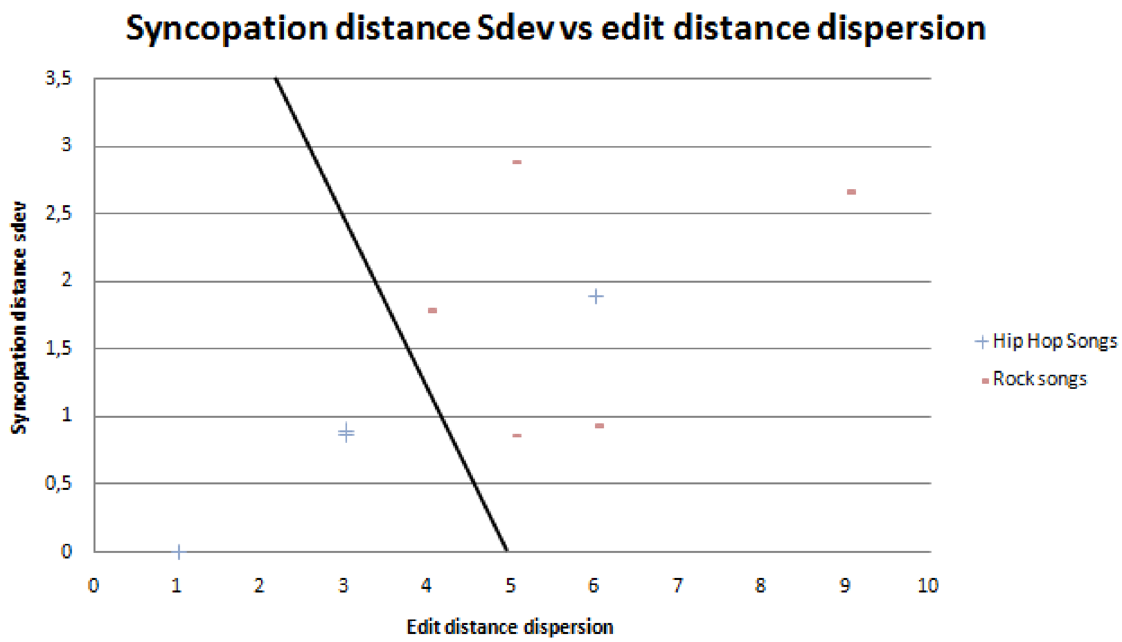

Figure 13. Syncopation distance dispersion vs. edit distance standard deviation for rock and Hip-hop songs.

c) Comparing distances and density

The next step is to compare values of the distance metrics with values of the density metric. Density is the amount of notes present on each pattern. For this measure, the mean, standard deviation, minimum, maximum, dispersion and flatness was calculated for each song. Then, the average of these values is taken as a summary for each genre.

By exploring data from density and plotting them against the distances, an interesting pair emerged. That is the dispersion of the edit distance vs. the maximum density for all songs. As can be seen, this pair is useful for segmenting Hip-hop songs away from Latin and Rock songs (figure 14 bottom right). It could also be used to separate pop songs from Latin songs (Figure 14 bottom left).

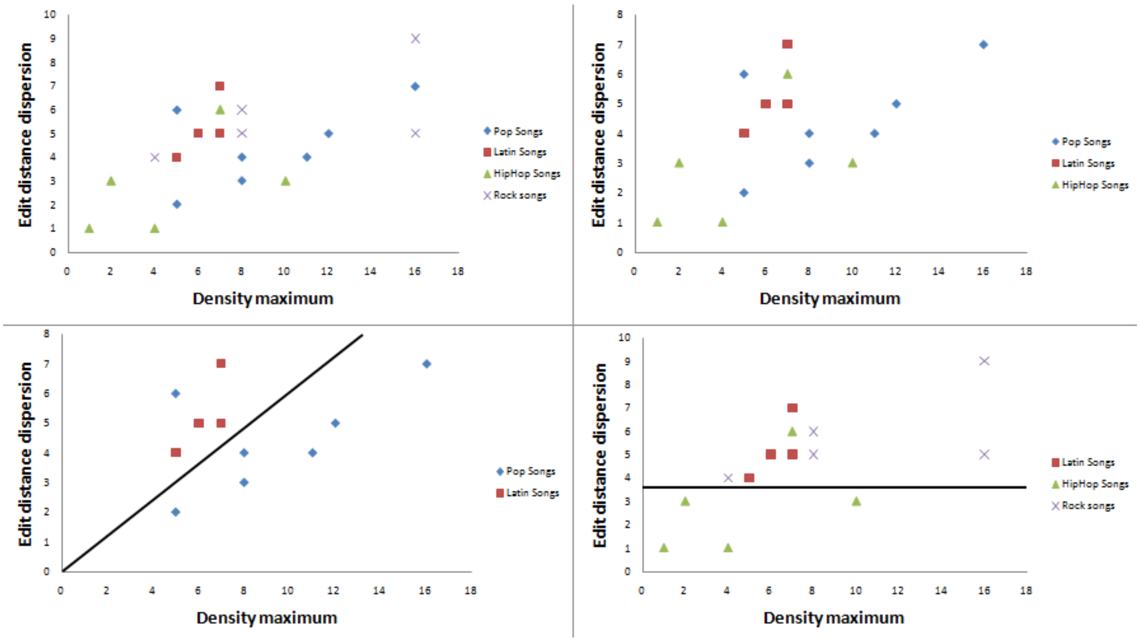

Figure 14. Edit distance dispersion vs. Density maximum for all songs. Top left all, top right Pop, Latin and Hip-hop songs. Bottom right Latin, Hip-hop and Rock songs. Bottom left, Pop and Latin songs.

Another quite interesting pair is the Standard deviation of the density and the dispersion of the syncopation distance. Figure 15 shows the plots for these variables.

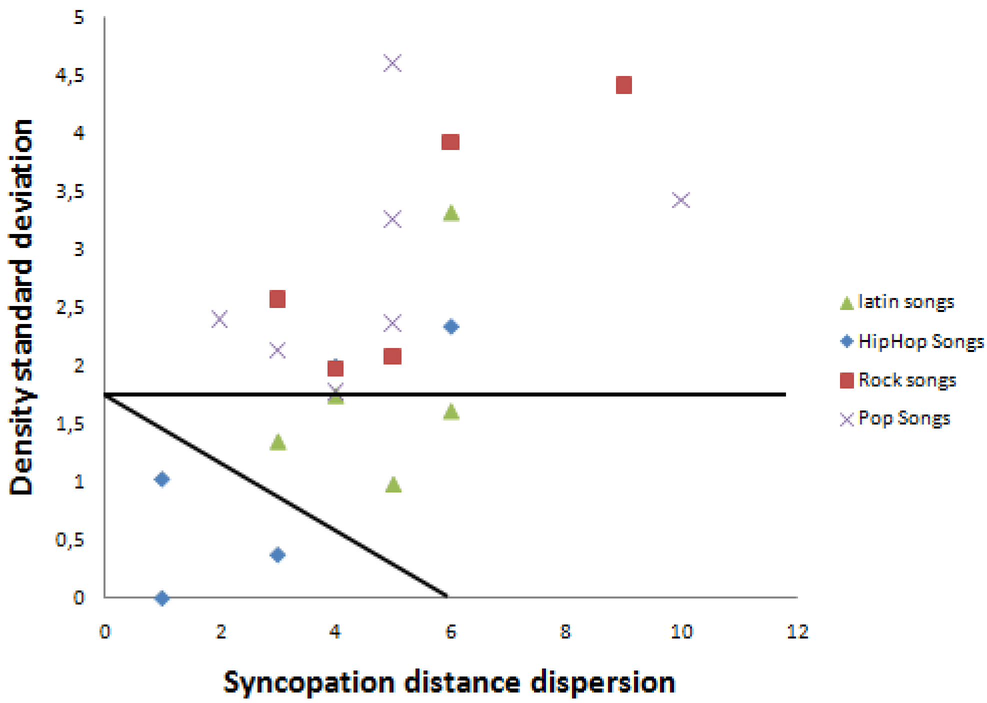

Figure 15. Standard deviation of the density and the syncopation distance dispersion for all songs.

d) Segmentation algorithm

As can be seen on figures 13, 14 and 15 there are some combinations of variables that spread the data in such a way that it seems auto organized by genres. From them some successful algorithms were derived. As is explained on each of those figures by the continuous lines, simple linear functions can be defined in order to segregate the disperse data into their genres. Those linear functions can

- Differentiate Hip-hop from Rock with a 75% of accuracy (3 out of 4)
- Differentiate Pop from Latin with an 85% of accuracy (6 out of 7)
- Differentiate Hip-hop from Latin and rock with 80% of accuracy (4 out of 5)
- Differentiate Hip-hop from the rest with 60% of accuracy (3 out of 5)
- Differentiate Latin from the rest with a 80% of accuracy (4 out of 5)

In general, functions like these would be of interest to classify automatically songs by their genre. This means, by analyzing the bass of a song we can guess what genre it belongs to. Due to the limited size of this work and the small size of the database, the functions that could be derived from this analysis are not expected to work for new songs.

5.2 Conclusions

The bass instrument is said to be the link between the percussion and the melodic instruments on a musical ensemble. As such it plays an important role for both the harmonic and melodic aspects of the music as well as for the rhythmic development of the pieces. Being this useful for the structure of music, it is not outrageous to expect that it can carry information about the genre that a song belongs to.

The genres presented here are very broad and the selected elements that represent them have been chosen both by subjective ideas of the genres and also by the availability of the MIDI files. This limits the aim of this work, because of disagreements on the items on all the genres.

The metrics used in this work are still under theoretical development. This can be seen in their difficulty to differentiate patterns sharing the same notes, edit distance and syncopation distance values. Nevertheless, as these are only scientific tools available for working with rhythms, they were used to test their proficiency in addressing musical tasks from in a broader sense than the discrete rhythm experiments.

The density or amount of notes present in a pattern was found useful in this classification task. This could point its relevance in the creation of bass patterns on certain types of music.

6. PERCUSSIVE SPACES FOR INTERACTION

6.1 Introduction

As it was mentioned on Chapter 4, the available metrics used to predict perceptual similarity of rhythms tend to group many patterns under the same values. This complicates the generation of new patterns by controlling their similarity values from a reference pattern. In some cases the result is not a pattern but a group of patterns and at some other points times the result is no pattern at all. To improve this scenario two problems must be solved: First, the problem of having big clusters in some regions of the 2D space composed by the edit distance and the syncopation distance, and then, the problem of having no elements in some regions of the space.

6.2 Avoiding clusters: Using density as an extra variable

To address the first problem, the best way is to use additional variables when generating rhythms. This would obviously add one dimension to the algorithm, but it would also help reduce the possible parameters present on a cluster. The implications of using new variables would be explored below.

To add new variables and improve the search algorithm, a variable that proved to be useful on the analysis experiments presented on the previous chapter was the amount of notes present on a pattern or density. Although it is not regarded in the literature as a main perceptual or cognitive factor for measuring the distance of two rhythmic patterns, it is a very simple and straightforward way to extract information from a pattern. It is also a valid way to compare between patterns.

Having selected density as a starting point, the possibilities of the dataflow of an interactive system used to generate new rhythms are presented. Density could help disperse the patterns and distribute the zones that contain many elements into smaller subsets (see Table 2). On the best case scenario, it would help individualize patterns according to the three variables used. The latter case would be that setting the three variables leads to only one pattern and not to a group of patterns.

There are different options to setup an algorithm that uses these three variables to search within the all-possible-patterns space (see Figure 16). The first alternative is to have all variables independent from each other and search throughout the space in parallel (Figure 16 top). The downside of this alternative is that a great amount of the space contains no elements as data is clustered around some specific regions as it is seen on Figure 6. So having all variables independent from each other, moving along all its range would cause certain dead spots very unattractive for real time interaction. These dead spots would make turning the knobs and interacting with the system into an uncertain procedure as a physical action could lead to a zero response from the algorithm.

Additionally, the order has a huge impact on the processing time. For example, having all the processes in parallel would imply that all 65344 patterns should be searched and ordered three times, once for each metric every time any of the parameters is changed.

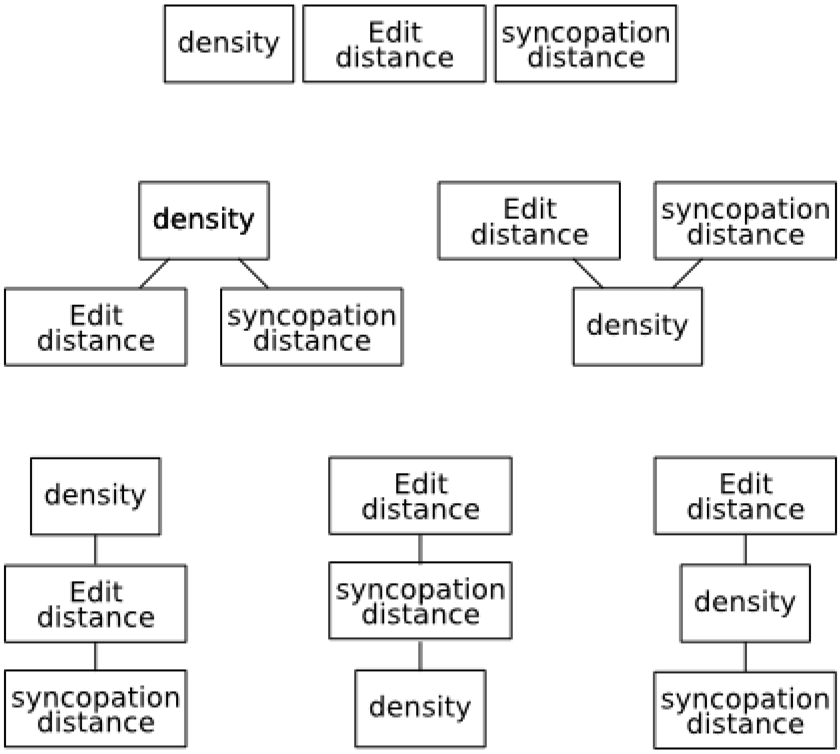

Figure 16. All data flow possibilities: Parallel on top, mixed on the center, in series on the bottom.

The second alternative is to have the three variables in series (Figure 16 bottom). The process on the bottom left of Figure 11 was programmed as a prototype on Pure Data. Calculating the density as a first step reduces the computation, because the edit distance needs only to be calculated for a specific amount of onsets and not to the whole 65344 element set. Next, the syncopation distance is calculated only for the remaining elements that share the edit distance and the amount of onsets. The details of the dataflow can be seen on Figure 12.

6.3 Avoiding empty spaces: Analysis of possible resulting sets

To address the second problem, an analysis of the unpopulated zones of the space has to be done in order to know the points that contain no elements so they can be avoided. Specifically the idea is to remove the blank spots from the knobs so they can give a feeling of continuity. For user interaction this is major goal, ensuring that every action in this system has a response

The methodology is first to filter from the possible pattern universe those who share the amount of onsets set by the Density knob (see Figure 17). The edit distance from the resulting set of patterns to the reference pattern is then computed. Those results are ordered according to the edit distance. The edit distance values not present on that result are removed from the range of the edit distance knob so the motion of the knob at any point has a corresponding set. Then, the set of patterns that share the values of the Edit distance knob are ordered with the syncopation distance to the reference pattern. Again, the syncopation distance values not present on that result are removed from the range of the syncopation distance knob so the motion of the knob at any point has a corresponding set. The set of patterns that share the value present on the syncopation distance knob are output. The resulting set of patterns can contain one or more elements but would never be an empty set.

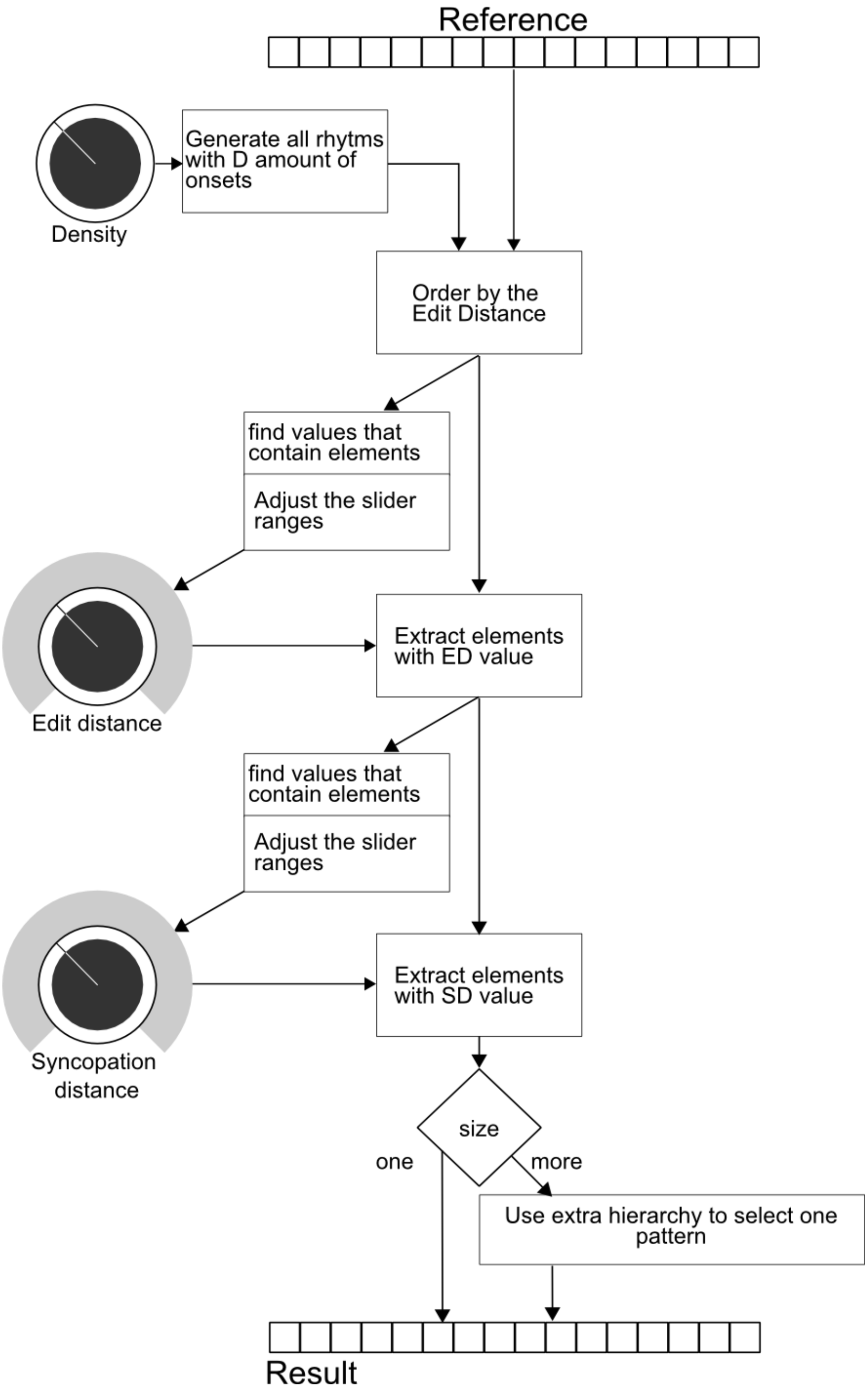

Figure 17. Dataflow of a serial system. The inputs are a rhythmic pattern, the density, the edit distance and the syncopation distance. The output is a new pattern.

Now the algorithm is improved for real time tweaking by avoiding variable setups that would yield no result. But still, this methodology leaves the door open for returning sets with more than one element. The following section deals with possible ways to overcome this situation.

6.4 Listening exercise of non unitary sets

After producing an application in Pure Data capable of performing the procedure described above, one listening activity was carried out in order to explore the results generated by the system. The activity consisted of extracting results in six different points of density and at specific edit and syncopation distance values from the reference pattern 1 0 1 0 0 0 1 0 0 0 0 0 1 0 0 0. Once the patterns were extracted they were reproduced. The aim was to understand the similarity among the patterns that were located on the same place in the system. The values of each point can be seen in Table 4.

Table 4. Materials for the listening exercise.

Set	Density	Edit distance	Syncopation distance	Number of results
A	6	2	0	6
B	6	2	1.732	9
C	6	4	0	19
D	6	5	2.236	206
E	6	7	3.162	20
F	6	8	3.3166	4

Table 5. Reference and result patterns from sets B and E.

Set	**Pattern**
Reference	**1 0 1 0 0 0 1 0 0 0 0 0 1 0 0 0**
B	1 1 1 1 0 0 1 0 0 0 0 0 1 0 0 0
	1 1 1 0 0 0 1 1 0 0 0 0 1 0 0 0
	1 1 1 0 0 0 1 0 0 0 0 0 1 1 0 0
	1 0 1 1 0 0 1 0 1 0 0 0 1 0 0 0
	1 0 1 1 0 0 1 0 0 1 0 0 1 0 0 0
	1 0 1 0 1 0 1 1 0 0 0 0 1 0 0 0
	1 0 1 0 0 0 1 1 0 1 0 0 1 0 0 0
E	1 1 0 0 0 0 0 0 1 1 0 0 0 0 1 1
	0 1 1 1 0 0 0 0 0 1 1 1 0 0 0 0
	0 1 0 0 1 0 0 1 0 1 0 0 0 0 1 1
	0 1 0 0 0 0 1 1 1 1 0 0 0 0 0 1
	0 1 0 0 0 0 0 1 1 1 0 0 0 0 1 1
	0 1 0 0 0 0 0 0 1 1 0 0 0 1 1 1
	0 0 1 1 1 1 0 0 1 1 0 0 0 0 0 0
	0 0 1 1 1 1 0 0 0 1 0 0 0 0 0 1
	0 0 1 1 1 1 0 0 0 0 0 0 0 1 0 1
	0 0 1 1 0 0 0 0 1 1 1 1 0 0 0 0
	0 0 0 1 1 1 0 1 1 1 0 0 0 0 0 0
	0 0 0 1 1 1 0 0 0 0 0 0 1 1 0 1
	0 0 0 1 0 0 0 1 1 1 1 1 0 0 0 0
	0 0 0 1 0 0 0 0 1 1 1 1 0 0 0 1
	0 0 0 1 0 0 0 0 0 1 1 1 0 0 1 1
	0 0 0 0 1 1 0 0 0 1 0 0 0 1 1 1

The reference pattern and the patterns from sets B and E are presented on Table 5. Set B is created by using a density value of 6 an edit distance value of 2 and a syncopation distance value of 1.732. Setting the density value in 6 (as are all sets in Table 5) shortens the amount of possible patterns set from 65.534 to 8.008 (Table 2). These 8.008 patterns of 6 onsets can have edit distance values from 2 to 8. This is because the least transformations to be done from a pattern with density 4 (the reference) and any pattern with density of 6 is 2. In other words, at least 2 transformations have to be done from a pattern with 6 onsets to become a pattern with 4 and vice versa. The highest value means that at the most you can perform eight transformations to the reference pattern for it to become a complementary pattern with 6 onsets. The total number of possible patterns with density 6 and edit distance 2 is 66.

With density set to 6 and edit distance is set to 2, the 66 different patterns are grouped under syncopation distance values of 0, 1, 1.414, 1.732, 2, 2.236 and 2.449. Other syncopation distance values generate empty sets. It is important to point that the distribution of the 66 patterns in these seven categories of syncopation distance is irregular and has not been analyzed in this work. Finally, setting a syncopation distance of 1.732 generates finally a set of 9 patterns, presented on Table 5 as set B. Reducing 8.008 patterns with density 6 to 9 patterns with edit distance 2 and syncopation distance 1.732 is considerable. Nevertheless, the functionality of an interactive application supposed to deliver only one pattern is not achieved.

Set E has a density of 6 and distances edit of 7 and syncopation 3.162 to the reference pattern. Setting the edit distance in 7 reduces the amount of possible patterns from 8.008 to 614. Setting the syncopation distance to 3.162 reduces the amount of patterns to 20. Again, the reduction is significant, but it is not the desired unitary set.

a) Setup for the listening exercise

The resulting equidistant patterns presented on Table 5 were reproduced using a player in Pure Data where any of the six different points/sets could be selected, and each of the resulting patterns could be selected too (Figure 18). A player reproduced the selected pattern with mute and volume controls and also the reference with mute and volume controls. The timbre of the reference and the result were different and produced with a simple amplitude modulation synthesizer. The pitch of the reference pattern was 400Hz and the result pattern had a pitch of 800Hz. The duration of each sound was 40 milliseconds.

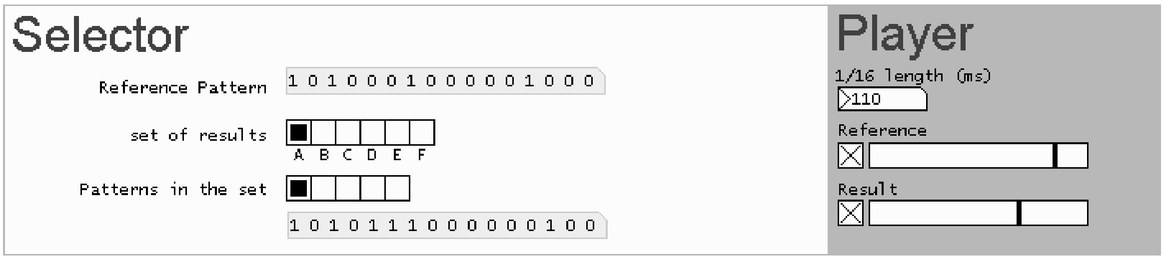

Figure 18. Selector and player for the listening exercise.

b) Observations from the listening exercise

In order to explore the coherence of elements in the resulting sets and seeking for a strong perceptual coherence, the listening test was carried out. Each element of a set was reproduced in sequence several times.

Patterns within a given set do not seem to have a very strong perceptual relation when listening to them. It is clear that numerically they share the same distances to the reference, but perceptually a musically trained ear is not able to predict that those patterns belong to a same group. Nevertheless an evolution can be perceived within the group, as the elements are progressively selected with the scroll ratio on the GUI, the onsets progressively shift from left to right. This is due to the way the universe of possible patterns is constructed for the processing in the software as explained on section 3.3 a. It is precisely this shift is what generates the perceptual incoherence of a group: The place of the onsets within the pattern is a highly relevant feature of a rhythmic pattern and a clear means of defining its similarity in a listening scenario.

On the other hand, when listening progressively to one element of each set moving from the closest set to the farthest one (from A to F), it is very clear that the similarity is shifted away from the reference pattern. This is illustrated by a compilation of one pattern extracted from each set presented on Table 6. When the edit distance is in lower values (sets A, B and C), there are obviously more coinciding onsets in places of high beat reinforcement such as the first onset and the 8^{th} onset (if present). Progressively, the onsets of the resulting patterns begin to shift away from the onset places of the reference pattern, slowly grouping in the places that originally belong to silences on the reference.

In its turn, an increase on the syncopation distance, as performed between set A and set B (both sets share the same density and the same edit distance), shows a clustering of the onsets towards the first quarter of the pattern. This change eliminates the high reinforcement of the beat on pattern A, generated by having an onset on the first step and a silence on the second step, which is a reinforcement of -4 according to the L salience profile (see section 3.2 a). Also, when placing an onset on the fourth step removes the reinforcement of the beat caused by having an onset on the third step and silence on the fourth on pattern A. The resulting syncope of the first quarter of the sequence is shifted from -5 to 2. It goes from a strong reinforcement to syncope. In terms of the elements in the syncopation histogram, it goes from having 1 syncope of -4, 2 syncope of -1 and one syncope of -2 (syncopation histogram 1 0 1 2 0 0 0 0) to having one syncope of -2, one syncope of -1 and one syncope of 2 (syncopation histogram 0 0 1 1 0 1 0 0)

Table 6. Selection of one pattern from each set of the listening exercise.

Set of origin	Pattern
Reference	**1 0 1 0 0 0 1 0 0 0 0 0 1 0 0 0**
A	1 0 1 0 1 1 1 0 0 0 0 0 1 0 0 0
B	1 1 1 1 0 0 1 0 0 0 0 0 1 0 0 0
C	1 0 1 1 1 0 1 0 0 0 1 0 0 0 0 0
D	1 1 1 1 0 1 0 0 0 0 1 0 0 0 0 0
E	1 1 0 0 0 0 0 0 1 1 0 0 0 0 1 1
F	0 0 0 1 0 0 0 0 1 0 0 1 0 1 1 1

The exercise of rhythmic similarity apparently can be done in macro or micro levels of detail. Selecting one element from the list defined by edit and syncopation distances is the macro detail level. Selecting within the elements of a set with the same distance values could be considered the micro detail level.

6.5 Establishing hierarchies within resulting non-unitary sets

As the observation exercise suggests, an extra mechanism to create a hierarchy among the results with same density, edit distance and syncopation distance is reviewed in this section. Following a variation of the rhythmic families methodology (Cao, Lotstein, & Johnson-Laird, 2014), an algorithm to quantify the distance between the resulting elements of a rhythmic set and the reference pattern is proposed. The main idea behind this procedure is to select the element that is closest to the reference from the resulting set. As an additional result, the rhythmic family similarity between all the patterns in the resulting set is measured to the reference so all patterns receive a value.

The rhythmic family similarity distance is based on the syncopation salience measure. Families are established defining if each section of a given pattern reinforces the beat, challenges the beat or neither. As described in section 3.2 b, once the family has been defined, different distance measures can be established between the rhythmic family of the reference and the rhythmic family of any given pattern. Namely, distances can be measured using Euclidean, Hamming, cosine or edit algorithms. Using the convention of using a 1 for a section where syncopation is found, -1 for a section with beat reinforcement and 0 for none, the rhythmic family vector is defined. The application developed for this task uses Euclidean distance to compute the distance between rhythm family vectors. There is no argument for the selection of Euclidean above the other distances as it was previously done for the syncopation distance on section 3.1. The suitability of this metric above the others is an important issue that should be addressed further.

After applying this methodology, most of the sets can be split using rhythm family distance as a metric. Nevertheless, in some cases there are rhythms so similar that share the same density edit distance, syncopation distance, and rhythmic family distance. To illustrate this behavior, the results from the listening test were analyzed computing the distance from the reference’s rhythmic family to each element on the set.

The results are shown on last two columns of Table 7 where the use of the rhythmic family methodology is analyzed. The distances of every element on the resulting sets were analyzed to measure how many groups were found within the data. Additionally, the number of elements in the categories closer to the reference is reported.

Table 7. Materials for the listening exercise using additional percussive family distance measured using Euclidean distance to the reference. The additional column shows how many different family distance values are found inside the resulting group.

Set	Density	Edit distance	Syncopation distance	Number of results	Percussive family groups	Elements in closer group
A	6	2	0	6	1	6
B	6	2	1.732	9	1	9
C	6	4	0	19	2	14
D	6	5	2.236	206	7	7
E	6	7	3.162	20	3	5
F	6	8	3.3166	4	2	3

As can be seen, the use of the rhythmic family methodology helps differentiate the elements in the set in a small proportion specially leaving many elements in the closer group. For groups A and B, all elements share the same rhythmic family distance to the reference. Groups C and F have both its elements spread in two groups, with more than halve of the elements being in the closer group. Groups D and E have the most elements in their result and have more groups to locate them, 7 and 3 respectively. In the case of group D which is the most populated one, the reduction of possible elements is significant going from 206 to 7.

7. CONCLUSIONS

As seen on chapters 2 and 4, the edit distance and the syncopation distance are experimentally tested metrics that resemble human ratings of rhythmic similarity between percussive patterns. Both metrics are tested in rating experiments and correlate, up to a certain level, with the similarity expressed by human subjects. A comparison between both metrics as shown in section 4.2 shows that they are not orthogonal or correlated, but rather show a particular behavior that deserves further attention and experimentation. A possible way to improve the prediction of human rhythmic similarity ratings is to combine the metrics, which come from separate theories, in a unified algorithm.

As seen on chapter 5, the use of raw edit and syncopation distances and statistical metrics derived from them proved useful in genre classification tasks. Extrapolating the metrics (tested only in percussive contexts) to melodic scenarios such as bass melodies, helped differentiate genres in a small MIDI database. Although the database is quite small to claim the actual use of these metrics in more complex scenarios, the results prove that the metrics do not have a random behavior but instead they make possible for genre classifiers to go way beyond chance in accuracy. This result raises new questions on the role the bass as an instrument plays in conveying information about music genres in addition to timbre and tempo which are mainly used. Also suggests that the use of these metrics with more traditional genre classification techniques could reinforce actual results.

Development of tools for real time interaction with drum sounds and rhythmic patterns based on the cognitive space was not completely finished. The use of cognitive similarity distances, such as edit and syncopation and syncopation family, in a rhythm production environment is very promising; particularly when seeking for a space to drive rhythm design. Nevertheless the space proposed in this work, proved not to suffice for delivering one single pattern as output from the system. Further research must be carried out to refine the system to achieve this objective.

Implementation of the distances used in this work, have internal adjustments and variations that require making decisions at some point. These decisions such as picking the right metric used to measure distance between two syncopation or family vectors, selecting the resolution of the rhythm family vector and selecting the weights used to compute the syncopation level or the syncopation family affect the results. Distortions caused by those decisions are highly relevant but unfortunately not explored completely in this work. Further research could start from recognizing the impact of these configurations on more robust listening tests. Other research direction is to compare the results obtained when measuring the distance between syncopation families using different metrics than the Euclidean distance.

Other scenarios where the metrics used in this work could be explored are the evolution of rhythmic patterns in musical improvisations of both rhythmic and melodic instruments. Perhaps new algorithms could be designed to anticipate the next rhythmic variations of an ongoing improvised section such as a melodic solo or a percussive roll. Also, the relations of the different parallel elements of rhythmic ensembles could be explored with the same metrics used in this work to help decipher the mechanisms that make those ensembles and their patterns endure over time.

REFERENCES

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

Cha, S. H., & Srihari, S. N. (2002). On measuring the distance between histograms. *Pattern Recognition*, *35*(6), 1355-1370.

Deza, M. M., & Deza, E. (2009). *Encyclopedia of distances* (pp. 1-583). Springer Berlin Heidelberg.

Fitch,T. and Rosenfeld A.J.(2007) Perception and Production of Syncopated Rhythms. *Music Perception*, Vol. 25, No. 1. (September 2007), pp. 43-58.
Available online: http://homepage.univie.ac.at/tecumseh.fitch/wp-content/uploads/2010/08/FitchRosenfeld20071.pdf

Foote, J., Cooper, M. L., & Nam, U. (2002, October). Audio Retrieval by Rhythmic Similarity. In *ISMIR*.

Foote, J., & Uchihashi, S. (2001, August). The Beat Spectrum: A New Approach To Rhythm Analysis. In *ICME*.

Gómez, F., Thul, E., & Toussaint, G. (2007). An Experimental Comparison of Formal Measures Of Rhythmic Syncopation. Proceedings of the 2007 International Computer Music Conference, 2007, pp 101 - 104

Gustafson, K. (1988). The graphical representation of rhythm. *PROPH) Progress Reports from Oxford Phonetics*, *3*, 6-26.

Holzapfel, A., & Stylianou, Y. (2009, April). A scale transform based method for rhythmic similarity of music. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on* (pp. 317-320). IEEE.

Ladinig O. (2009) “Rhythmic complexity and metric salience,”in Temporal expectations and their violations (Olivia Ladinig, 2009),pp.29–47, Science Park 9041098XH Amsterdam: Institute for Logic,Language and Computation Universiteitvan Amsterdam, 2009

Lerdahl, F., & Jackendoff, R. (1983). A Generative Theory of Tonal Music. Cambridge: The MIT Press

Longuet-Higgins HC, Lee C (1984) The rhythmic interpretation of monophonic music. Music Perception 1: 424–440.

Orpen, K. S., & Huron, D. (1992). Measurement of similarity in music: A quantitative approach for non-parametric representations. *Computers in music research*, *4*, 1-44.

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

Post, O., & Toussaint, G. (2011). The Edit Distance as a Measure of Perceived Rhythmic Similarity. *Empirical Musicology Review*, *6*(3).

Povel, D. J., & Essens, P. (1985). Perception of temporal patterns. *Music Perception*, 411-440.

Shmulevich & Povel, D.-J. (2000) “Measures of temporal pattern complexity,” Journal of New Music Research, vol. 29, no. 1, pp. 61–9, 2000

Smith, L. (2010). *Rhythmic similarity using metrical profile matching*. Ann Arbor, MI: MPublishing, University of Michigan Library.

Smith, M. and Honing, H. “Evaluating and extending computational models of rhythmic syncopation in music,” in Proceedings of the International Computer Music Conference. International Computer Music Association, 2006, pp. 688–91.

Sung-Hyuk Cha, Sargur N. Srihari, On measuring the distance between histograms, Pattern Recognition 35 (2002) 1355–1370.

Thul, E., & Toussaint, G. (2008). Rhythm complexity measure: A comparison of mathematical models of human perception and performance. Proceedings of the International Symposium on Music Information Retrieval, 2008, pp. 663–8.

Typke, R., Giannopoulos, P., Veltkamp, R., Wiering, F., Van Oostrum, R. (2003) Using transportation distances for measuring melodic similarity, in: Holger H. Hoos, David Bainbridge (Eds.), Proceedings of the Fourth International Symposium on Music Information Retrieval, Johns Hopkins University, Baltimore, 2003, pp. 107–114.

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

Toussaint, G. (2010). Computational geometric aspects of rhythm, melody, and voice-leading. *Computational Geometry: Theory and Applications*, *43*(1), 2-22.

Tutzer, F. (2011). Drum rhythm retrieval based on rhythm and sound similarity. *Master's thesis, Department of Information and Communication Technologies Universitat Pompeu Fabra, Barcelona*.

Wiering, F., Typke, R., Veltkamp, R., (2004) Transportation distances and their application in music-notation retrieval, Computing in Musicology 13 (2004) 113–128.

Witek MAG, Clarke EF, Wallentin M, Kringelbach ML, Vuust P (2014) Syncopation, Body-Movement and Pleasure in Groove Music. PLoS ONE 9(4): e94446. doi:10.1371/journal.pone.0094446