Tap to Drums:
Extending Monophonically Tapped Rhythms to Polyphonic Drum Pattern Generation

Peter Clark
{peter.clark01@estudiant.upf.edu,
peterjosephclark1@gmail.com}

Supervisors:
Daniel Gómez-Marín {daniel.gomez@upf.edu}
Sergi Jordà {sergi.jorda@upf.edu}

Thesis for the Completion of a Master's of Science in Sound and Music Computing
Multimodal Music Interaction Lab (MMI) - Music Technology Group (MTG)
Universitat Pompeu Fabra (UPF)
Roc Boronat 138, Barcelona, Catalonia, Spain

Table of Contents

Abstract……...3
1. Introduction……4
2. State of the Art……7
2.1. Rhythm Perception……..7
Auditory Perception and Motor Coordination……10
2.2. Rhythm Similarity……11
Monophonic Rhythm Similarity……11
Polyphonic Rhythm Similarity……..16
2.3. Symbolic Sequences and Rhythm in EDM……17
2.4. Rhythm Spaces……..18
Representing Similarity through Perceptual Spaces……18
Representing Rhythmic Similarity……20
Benefits of a Rhythm Space Interface……..21
3. Experiment 1: Rhythm Flattening……22
3.1. Flattening Algorithms……22
Naïve Flattening Approach……23
Instrument-Based Flattening Approaches……..23
Frequency-Based Flattening Approaches……..24
Onset Density and Metrical Flattening Approaches……..24
Discrete, Semi-Continuous, and Continuous Representations...……...26
3.2. Evaluation of Flattening Algorithms……29
4. Experiment 2: Rhythm Tapping……..31
4.1. Materials and Procedure……..32
Participants……..32
Materials……..33
Procedure……33
4.2. Rhythm Tapping Experiment Results……35
4.3. Comparison with Flattening Algorithms……42
5. Discussion……45
6. Conclusion……46
Acknowledgements……47
Appendix……...48
Bibliography……...50

Abstract

In this paper, we explore the literature surrounding rhythm perception to develop algorithms that extract a monophonic rhythm from a polyphonic drum pattern. We develop machine learning models for those algorithms to predict the pattern's location in a polyphonic similarity based 2-d latent rhythm space. Following that we have 25 subjects tap along to polyphonic drum patterns to explore the behaviors of reproducing complex rhythms. The model was able to reasonably predict the location of a monophonic rhythm in the rhythm space (MAE=0.039, SD=0.057). Subjects tapped more accurately to an intended velocity as they became more experienced with the system. The model failed to predict the location of the subject-tapped monophonic rhythms (MAE=0.4580, SD=0.076), highlighting the need for a more thorough subject-rated investigation into refining a tap→polyphonic drums pipeline.

Keywords: *rhythm, rhythm perception, rhythm similarity, tapping, rhythm space*

1. Introduction

When we hear a series of musical sounds, we do not merely become aware of the relative pitch of the sound or the timbral qualities, but become aware of the temporal relations between the sounds. This is the basis of perceiving a mental construct called rhythm. Rhythm is an inextricable property of music, plays an integral role in shaping our auditory perception and experience of music. Long capturing the interest of musicians, cultural researchers, and cognitive scientists alike, the study of rhythmic perception can provide us with many benefits in the fields of music analysis, generation, reproduction, and well as neuroscience and sociology. One particular aspect of rhythm deemed a "largely uncharted territory" (Georgi, Gingras, & Zentner. 2023) is the study of how we perceive and reproduce the rhythm of complex polyphonic patterns, like a full drum kit. Although polyphonic rhythm can be thought of as a series of overlapping layers of sound, it is not clear if this is the actual way in which we perceive it. Therefore, research into the relationship of rhythm between monophonic and polyphonic patterns lends itself to aspects in music cognition, composition, performance, production, and specifically music generation. If we have a good understanding of how people will extract the underlying rhythm from a complex musical piece, we can try to model this relationship and take advantage of it in compositional and generative scenarios.

The field of music generation is proliferating at great speeds, with many generative models being released in the last 5 years including MusicVAE (Roberts, et al., 2018), DrumNet (Lattner & Grachten, 2019), Magenta Studio (Roberts et al., 2019), RhythmVAE (Tokui, 2020). These systems are not yet easily integrated into music production software (Magenta is an exception), are hard to train, data-heavy, and prone to copy-right claims. They are not designed for easy customization or live performances (Gómez-Marín, 2018; Roberts et al., 2019). For example, MusicVAE was trained on 1.5 million unique MIDI files. All of these factors reduce accessibility to the growing world of generative music and production, and provide motivation to create light-weight systems that are accessible and customizable. An avenue of approach is to represent aspects of rhythmic similarity in latent spaces, such as Drum Rhythm Space (Gómez-Marín et al 2020), and R-VAE (Vigliensoni et al. 2022). These

spaces can be navigated, with similarity relations represented by distance, allowing intuitive exploration of drum rhythms. R-VAE is implemented in the Max for Live programming language used in the digital audio workstation Ableton and allows users to create latent space models for custom datasets. The Drum Rhythm Space constructs a latent space based on subject similarity ratings and a novel interpolation algorithm. These two methods allow for easy exploration of generated rhythms, but lack a method to target a location in the latent space through a musical input. The process of selecting a desired output pattern involves selecting a location in the space. An input method, like tapping a rhythm, would act as a sort of seed location for the something similar to the desired polyphonic drum patterns, as opposed to needing an exploration of the space to find a similar pattern.

This paper presents an exploration towards the study of how people tap a single rhythmic stream to complex rhythms, the relationship between monophonically tapped patterns and the original complex polyphonic pattern, and modelling their tapping behavior. This modelling is interesting because it facilitates exploration of similar drum rhythms in 2D rhythm space, based on an accessible and intuitive tool for polyphonic drum pattern generation, that is started with a few simple taps. This project is furthering research on how to analyze, process, and generate drum patterns done by the Multi-Modal Music Interaction Lab at the Music Technology Group at UPF.

In order to illustrate our objects, a diagram is presented in figure 1.1. Currently, the rhythm space system is able to process a folder of MIDI drum patterns and project them into a 2-D space, where patterns are organized by similarity ratings. That process is highlighted by the dashed blue outline. Our main objective is:

"To what extent can a human tapped monophonic pattern capture the rhythmic structure of a symbolic polyphonic drum pattern and reliably predict the original pattern's position within a rhythm space?"

Our specific objectives are:

- Develop algorithms for flattening polyphonic drum patterns into a single channel that represents the rhythmic structure of the original polyphonic pattern, taking into account cognitive/perceptual attributes. (Fig. 1.1a)
- From a flattened pattern, create a model to predict a relevant drum pattern from a polyphonic drum rhythm space, organized by rhythmic similarity. (Fig. 1.1b)
- Compare artificially flattened patterns with human-tapped patterns, to study flattening algorithms capacity to predict human-tapped patterns. (Fig. 1.1c)
- From a tapped pattern, predict a relevant drum pattern from the rhythm space. (Fig 1.1d)

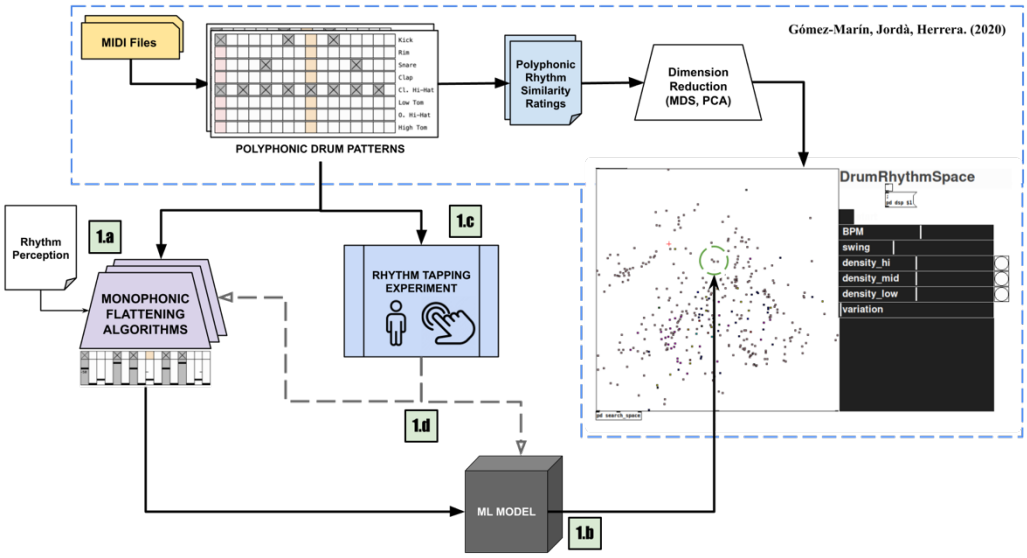


Figure 1.1: Project Roadmap.

We will begin by exploring the literature surrounding rhythm perception and similarity. Followed by an experiment for the creation of algorithms that output a monophonic pattern representing the rhythm of a polyphonic pattern, and predicting their location in the rhythm space from the monophonic pattern. Next, we will run an experiment to get subject tapped patterns. Lastly, we will analyze and discuss the results of both experiments.

2. State of the Art

2.1 Rhythm Perception

Rhythm

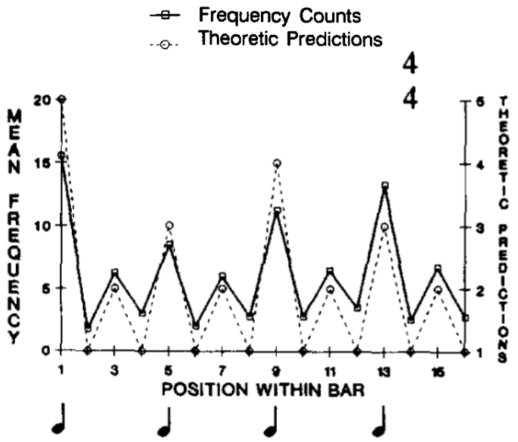
Rhythm refers to the temporal organization of music; the movement of a sequence of sounds characterized by regular recurrence or alternation of the timing, duration and stress of the sounds within the sequence (Merriam-Webster, 2023). We refer to the rhythm as a pattern of onsets, in congruence with the Gestalt approach. This Gestalt approach to rhythm places sounds (as notes) as the figure against a background of rests (Cooper & Meyer, 1960). The main defining characteristic of establishing a rhythm is the time in between the onset of each successive sound, what is called the inter-onset interval (IOI). Different rhythms sound similar if they share their IOIs, even if the note lengths are drastically different (Clarke, 1984; Povel, 1984). However, inter-onset intervals are not the only the characteristic to consider when perceiving a rhythm. Upon hearing a sequence of sounds, a human emergent mechanism projects the sounds onto a structure containing strong and weak accented beats called the *meter* (London, 2012). When presented with rhythms without a strong implied meter, sequences of notes are just perceived as a set of contiguous onsets, separated by the background of rests. In these scenarios, sequences of consecutive onsets separated by rests are referred to as perceptual groups (Povel & Essens, 1985). Rhythms with different perceptual groups are easier to discern than those with the same (Ross & Houtsma, 1994).

Meter

The regular strong positions in a meter are called the *pulse*. Typically, the pulse is when one would tap their foot when listening to a rhythm (Handel, 1989). The meter provides a structural framework for expectations of occurrences of sounds. Like grammar, the metrical principle of dividing a sequence into strong and weak accents recurses down to the lowest level of information, creating a hierarchical tree of rhythmic information. (Johnson-Laird, 1991). The lowest level of information is the temporally shortest division necessary to describe all the sounds and rests in the sequence. Unlike grammar, uncovering meter is a finite set of operations, and calls for cycling through only a handful of levels and sub-levels

of the hierarchical tree of rhythm information (Large & Kolen, 1994, London, 2004). Our cognitive system does not have access to infinite working memory however, and is prone to making partitions and thus handle perceptual meters of 2, 3, or 4 beats per measure, as they can be easily perceived without the need for counting (Miller, 1956). When a rhythm can be sufficiently *chunked*, applying a meter to a sequence of sounds reduces cognitive demands. Humans will even impose a meter on perfectly isochronous sounds (Fraisse, 1982) (sounds spaced out equally in time). Becoming attuned to a meter is manifested in many ways, like the common practice of assigning a strong accented beat on the first beat of a measure. Sounds that occur on the strong metrical positions are said to reinforce the pulse. We continuously use these cues to confirm the pulse and the meter of a sequence (Hannon et al., 2004; Large, 2008). As demonstrated in the Generative Theory of Tonal Music (GTTM) (Lerdahl & Jackendoff, 1983), how important a sound in the sequence is in relation to its meter is given by its height in the hierarchical tree of rhythm information (Longuet-Higgens & Lee. 1984; Palmer & Krumhansl, 1990) with the strongest pulse reinforcing sounds inhabiting the top of the tree. Height in tree is calculated by number of branching points from the top of the tree down to the note (see Fig. 2.1(1)).

(1)



(2)

Figure 2.1: (1) Hierarchical Trees of Two Simple Rhythms. (Longuet-Higgens & Lee, 1984).
(2) Frequency of note occurrence in note sequences in 4/4 meter. (Palmer & Krumhansl, 1990)

It has been suggested that even non-musicians have a strong ability for organizing beats into a meter, implying that metrical organization of notes in a sequence is an essential mental process, for which our cognitive systems excel at (Shaffer et al., 1985). As it can be seen, there is an interdependency between the perception of pulse, meter and rhythm: the nature of a rhythm relies on its meter, but we infer the meter of a sequence from its rhythm (Povel & Essens, 1985; Palmer & Krumhansl, 1990).

Syncopation

When we assume a meter is a hierarchical imposition on regions of a rhythm, a syncopated note is one whose onset is on a metrical unit of less importance than the one occurring prior to the next onset. The closer the syncopated onset is to a more metrically important position, and the metrically important position has no onset, the greater degree of feeling of anticipation or tension is provided. This feeling is called syncopation. Syncopation is a natural consequence of meter and its division of strong and weak beats. How syncopated a sound is relies on its location in the rhythm with respect to its meter and the important units within it (Cao et al., 2014). Note onsets occurring on pulses reinforce the meter while syncopations disrupt it (Volk, 2008). The departure from the implied metrical framework of a sequence of sounds shows the interplay between regularity and deviation in defining a rhythm, and the presence of syncopated sounds can greatly change the movement and energy of a musical piece. This is in line with the nature of perceptual groupings and the different IOIs possessed by reinforcing and syncopated sequences. In Fig. 2.1(2), a sound would be syncopated if it appeared at position 2 in the bar, and there were no notes in position 3.

Density

Density refers to the number of note onsets in sequence the interplay between density of notes and their positions within the meter affect the perception of a rhythm. Gabrielsson (1973) used note onset density as a metric to describe rhythmic similarity. Gómez-Marín et

al. (2020) confirm that density of note onsets of particular frequency ranges is a differentiable factor in polyphonic rhythms. The more complex and denser a sequence of sounds is, the less rhythmic value each individual sound will have overall. Rhythm reproduction worsens as the density and rhythm size increase (Milne & Dean., 2021) Vice versa, within sparser sequences, each sound is more rhythmically relevant as there are less auditory onsets to impose a meter on.

Frequency

A typical set of rhythmic instruments in dance music is a set of drums (Collins., 2013), which consists of a collection of different individual percussive elements. When dealing with polyphonic rhythms, or rhythms produced with more than one timbre and pitch, there are additional factors that affect rhythmic perception. Hove et al. (2014) highlights how the low frequencies have a larger influence in defining the rhythm than higher frequency sounds, but are not the only influence. Fuji et al (2011) show that snares are highly considered in metrical perception, and are used in other rhythm studies (Câmara et al., 2022; Frühauf et al. 2013).

As described within this section, attributes of rhythm perception like pulse, meter, syncopation, density, frequency, and grouping dynamics all come together to define human perception of rhythm and can be used to distinguish one rhythm from another analytically.

2.1.1 Auditory Perception & Motor Coordination

Exposure to rhythm and metrical entrainment is known to generate synchronous body movement on the perceived pulses (Grahn & Brett. 2007; Benedetto & Baud-Bovy. 2021) as well as expressing different metrical levels with different parts of the body (Burger et al. 2014). Thus, using tactile modalities such as tapping has been widely used to study the coupling of an acoustic signal with auditory perception. In previous studies of finger-tapping with regard to rhythm, it was found that there are innate relationships between the auditory perception of rhythm and the resulting motor-system reproduction. In particular, Grahn (2009) found that the mean velocity of taps (indirect measure of tapping force) was higher

on taps that coincided with pulses of the meter. Benedetto & Baud-Bovy (2021) confirmed this, going so far as to say that rhythm organization is directly encoded in the force profile of finger taps, fine-grained to multiple levels of tapping force relative to sounds position in the perceived meter. It is important to note that subjective rhythmic perception is highly dependent on one's musical experience and cultural background. In a spontaneous rhythm tapping experiment, Nistal et al. (2017) were able to identify subjects with musical experience compared to musically *naïve* subjects. Outside of level of musical experience, participants still differed substantially in their ability to perceive and discriminate the musical cues relevant to rhythm perception as well as their accuracy in rhythm reproduction (Milne, 2021). A tapped input signal is presumed to encode important information related to the tapped rhythm.

2.2 Rhythmic Similarity

2.2.1 Monophonic Rhythmic Similarity

When comparing monophonic rhythms, there are two main approaches to define rhythmic similarity among them. One is information-based approaches (Toussaint, 2004) and the other is perceptual-cognitive approaches (Johnson-Laird 1991; Cao et al. 2014). An information-based similarity metric between strings of information is the edit distance between them. This can be calculated through an algorithm that counts the number of transformations of its elements through insertion, swapping, or deletion that must be performed on one pattern to become another. In some cases, the edit distance has been seen to be correlated with reported rhythmic similarity (Guastavino et al., 2009; Post & Toussaint, 2011), and demonstrates reasonable results with types of cyclical rhythms like Afro-Cuban or Middle-Eastern rhythms. Some shortcomings of edit distance may be that it considers all three types of transformation to have the same weight (Toussaint et al., 2011) and that it and perceptual groups do not take into account for meter, and therefore syncopation (Post & Toussaint, 2011). While edit distance has been used in other aspects of cognitive psychology, such as deductive reasoning (Ragni et al., 2013), this approach however, fails to take into account perceptual and cognitive factors that affect human perception of rhythm. These factors

include salience of the meter (Palmer & Krumhansl, 1990), syncopation (Longuet-Higgens & Lee, 1984), or the induced meter and pulse (London, 2012).

Syncopation

Rhythmic patterns have different amounts of syncopation, defined by the relationship between the implied pulses of the rhythm, and the note onsets within the rhythm (Fitch & Rosenfeld, 2007). Longuet-Higgens & Lee (1984) derived a metric of rhythmic syncopation for rhythms with common time (4/4) measures. Every location within the rhythm is assigned a perceptual weight based on the importance of its location within the meter. The highest importance of a note is the first metrical unit in the measure (starting note) and has a perceptual weight of 0. The note location that starts the second half of the measure is weighted -1. The notes that start the second and fourth pulses in a measure are weighted -2 (Fitch & Rosenfeld, 2007). Through this metric, a syncopation value can be determined for the whole rhythm through addition of individual notes within the rhythm, and one that is cognitively and perceptually relevant. Gómez-Marín et al. (2015ab) showed that this type of monophonic syncopation metric is useful for subjective rhythmic similarity reports.

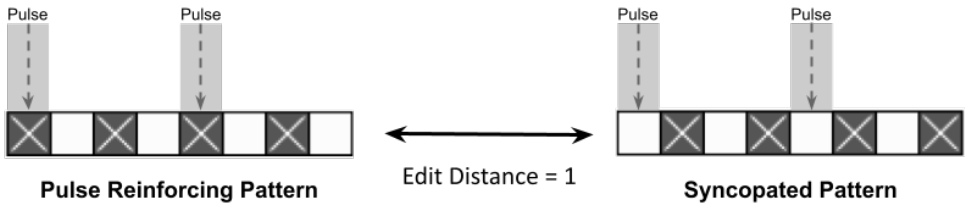


Figure 2.2: A pulse reinforcing pattern and a syncopated pattern with an edit distance of 1. Note onsets are represented by dark gray boxes with white *x*'s, rests are white boxes, pulses are indicated by gray rectangles.

In figure 2.2, we see an example of how a low edit distance doesn't always equate to high similarity. Despite the edit only consisting of one swap, a sequence with 4 pulse-reinforcing notes will sound noticeably different to a fully syncopated pattern due to metrical perception, something which edit-distance as a measure is agnostic towards.

Rhythm Families

The concept of *rhythm families* for use as a metric of rhythmic similarity was proposed by Cao et al. (2014). Rhythm families take into account identical regions or sub-regions of notes and syncopation of the notes of the sequence. The meter of a rhythm separates notes into three categories relative to how they relate to the metrical structure: *N* – note reinforces meter, *S* – note is syncopated, *O* – neither reinforcing or syncopated. How the sequences of notes are arranged is referred to as *syncopation* families. When referring to *identical regions*, they are the sub-regions of a pattern of notes that contain the same sequence of *NSO* notes as present in another pattern of notes, although not necessarily in the same place in the pattern.

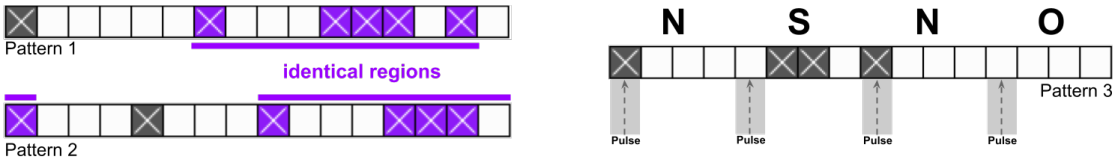


Figure 2.3: (L) Identical regions between two monophonic patterns in purple and (R) syncopation families highlighted in a 16-step pattern. N=Pulse Reinforcing. S=Syncopated. O=Neither/None/Rest. (Cao et al, 2014)

Identical regions existing within different locations and the presence of shared syncopation families increases reported similarity sensations between two rhythms (Cao et al., 2014). Patterns 1 and 2 on the left side of Fig 2.3 share the same 9-note region, while the location of the region within the meter is different for each. The relationship between identical regions and perceived similarity was further explored and confirmed by Gómez et al. (2015ab) in situations where the meter was induced, a form in which most rhythmic music is perceived. It was also shown that syncopation families are useful in rhythmic similarity predictions when the pulse is induced (as most western music is experienced) (Gómez-Marín et al, 2015ab). This is consistent with small edit distance similarity research from Toussaint (2013), as edit distance between two syncopation families reflects the similarity between the

two.

Evenness and Balance

Circular representation of rhythm in music has existed for a long time. As far as 13th century musician Safi al-Din al-Urmawi depicting rhythmic cycles (Toussaint, 2013) to modern study of music cognition (London 2004; Shepard 1964), due to the often-periodic nature of rhythm. The onsets of notes are spread around the unit circle, with the angular position of the note representing the temporal onset of the note, modulo the period of the rhythm. Having the rhythm represented as points on a circle allows exploration of different metrics for rhythmic similarity, specifically *balance* (Milne et al., 2017) and *evenness* (Milne & Dean, 2016). *Balance* and *evenness* are, respectively, the zeroth and first coefficients of the discrete Fourier transform (DFT) of a rhythmic pattern encoded onto the unit circle (Milne & Herff, 2020). The discrete Fourier transform represents a one-to-one relationship between the time-domain, and the frequency domain, and has relevance to human perception of sound. The human ear, specifically the cochlea, processes sounds through a Fourier-like transformation to identify the frequencies present.

The pattern of onsets of notes in a rhythm can be numerically represented by fractions over the period (length of pattern), and then multiplied by $e^{2\pi i X_p}$ to end up at the Argand vector of that pattern, Z_p (Milne et al., 2017). Take a 16-step pattern p with a kick drum on the first beat of every bar, and on the second beat of the last bar:

$$p\text{: } [1000100010001100]$$
$$\rightarrow X_p\text{: } \left\{\frac{0}{16}, \frac{4}{16}, \frac{8}{16}, \frac{12}{16}, \frac{13}{16}\right\}$$
$$\rightarrow Z_p\text{: } \{\, e^{2\pi i \frac{0}{16}}, e^{2\pi i \frac{4}{16}}, e^{2\pi i \frac{8}{16}}, e^{2\pi i \frac{12}{16}}, e^{2\pi i \frac{13}{16}}\}$$

The precise location of the note on the circle can be described by an Argand vector, a vector consisting of complex numbers. The temporal resolution of the vector is given by the sample rate, and can easily be up-sampled. As there are only as many terms in the Argand vector as there are onsets, it is said to be a *sparse* representation. It is a known fact that *mental features*, simplified representations of complex phenomena, are cognitively utilized for memory and

statistical learning (Turk-Browne & Scholl, 2009; Rohrmeier & Rebushcar, 2012), and the potential is there for human temporal rhythmic/sound processing uses these Argand vector-like mental representations (Feldman, 2016). A visual representation of pattern *p* can be seen in Fig. 2.4, on the right hand side.

Balance in this context refers to the proximity of the rhythm's *center of mass* to the center of the unit circle upon which its Argand vector is represented. It is normalized between [0,1], where 1 is a perfectly balanced rhythm whose center of mass is at the center of the circle, and a 0 the center of mass is on the circle edge. One can intuitively imagine it like a spinning coin; the coin would not spin properly if the center of mass is not at the center of the coin. The balance of a rhythm B_{zp} is calculated by taking 1 minus the magnitude of the zeroth coefficient of the discrete Fourier transform taken on the Argand representation of the rhythm Z of length K (Milne & Herff, 2020).

$$B_z = 1 - \frac{|\sum_{k=0}^{K-1} Z_k|}{K}$$

$$B_z = 1 - \frac{|\mathcal{F}(Z)_0|}{K}$$

A rhythm's *evenness* is the measure of lack of variance in the inter-onset intervals within the rhythm and like *balance*, is normalized between [0,1]. If the note onsets in a rhythm are all equally spaced out temporally (like a metronome), then this rhythm is maximally even and would have an evenness of 1. All perfectly even rhythms are perfectly balanced, but not all perfectly balanced rhythms are perfectly even (Milne et al., 2017). Evenness can be expressed as the concentration of the angular deviations when comparing the rhythm's onsets and the uniform distribution of steps in the rhythm (Milne et al., 2015). Evenness, E_z, is described mathematically as the first coefficient of the discrete Fourier transform taken on the Argand representation of the rhythm Z of length K.

$$E_z = \frac{\left|\sum_{k=0}^{K-1} Z_k e^{-2\pi i \frac{k}{K}}\right|}{K}$$

$$E_z = \frac{|\mathcal{F}(Z)_1|}{K}$$

As only the magnitude of the zeroth and first coefficients are used for balance and evenness respectively, this means that both the *balance* and *evenness* metrics are phase and direction insensitive. Neither a rhythm's balance or evenness will not change if it is started in the middle or played backwards (Milne & Herff, 2020). Unbalance rhythms and rhythms with clearly distinguishable halves are easier to reproduce by tapping than even ones, and this may facilitate location-finding in the pattern (Milne & Dean, 2021). This can be seen in pattern *p* in Fig 2.4 (R); the set of two notes on the left side of the right circle precede the first note of the pattern, providing indication of location in the pattern and meter.

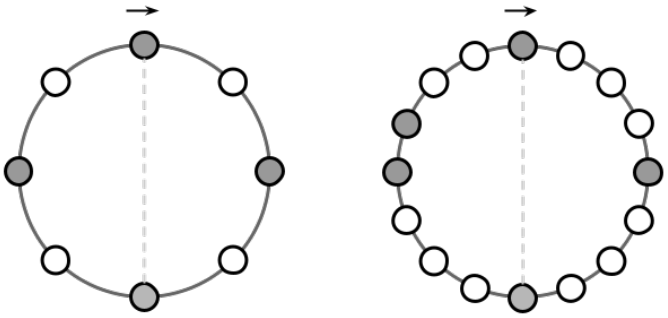


Figure 2.4: (L) A balanced and even rhythm (length 8). (R) Pattern *p*, a slightly unbalanced and uneven rhythm (length 16). Gray circles are note onsets, white are rests.

Syncopation (Gómez-Marín et al, 2015ab), similar regions of rhythm families (Cao et al., 2014; Gómez-Marín et al, 2015ab), and balance & evenness (Milne & Dean, 2021) have shown their usefulness as metrics for monophonic rhythm similarity judgements. They carry useful information about the rhythm and can be adapted or expanded upon for polyphonic similarity judgments, as it is a series of monophonic channels.

2.2.2 Polyphonic Rhythmic Similarity

With the focus on polyphonic rhythms, there are studied features that derive metrics for polyphonic rhythmic similarity. In Gabrielsson's (1973) experiments with polyphonic rhythmic similarity, he reported that similarity measurements were affected by five characteristics: (1) meter induced by drum pattern, (2) onset density of the drum sequence, (3) simplicity of drum pattern, (4) number of different instruments, and (5) movement

character of the rhythm. Other studies take into account some human perceptual characteristics. Bouwer et al. (2014), Burger et al. (2017), Witek et al. (2014), and Hove et al. (2014) showed that the most prominent frequency of a drum pattern has the power to confirm or disturb the meter, and affects similarity judgements. Witek et al. (2014) derived a polyphonic syncopation metric used to study desire to dance and listening pleasure in music. These four studies also demonstrated that lower frequency instruments have a higher ability to affect the establishment of a pulse within the metric as compared to higher frequencies. Hove et al (2015) derived a polyphonic syncopation metric by separating the drums within the pattern by their spectral centroid into three bins, consisting of low, middle, and high frequencies. Separating into the three percussive ranges for drums is akin to how the most energetic frequency bands of a sounds affect the way rhythm is processed (Gómez-Marín et al. 2020). Using these polyphonic similarity metrics and multi-dimensional scaling, Gómez-Marín et al. (2020) identified a set of descriptors of a rhythm that influence polyphonic similarity measurements the most (Table 2.1):

Descriptor	Definition
midD	Mid Density: Ratio of number of note onsets in mid-frequency channel to pattern length in steps
hiD	High Density: Ratio of number of note onsets in high-frequency channel to pattern length in steps
hiness	High Relative Density: Ratio of number of note onsets in high-frequency channel to total number of note onsets
lowsync	Low Syncopation: Syncopation value of low-frequency channel
hisyness	High Syness: Ratio of Syncopation value of high-frequency channel to number of note onsets in high-freq. channel.

Table 2.1: Relevant Descriptors for Polyphonic Similarity Measurements (Gómez-Marín et al. 2020)

2.3 Symbolic Sequences and Rhythm in EDM

The context of this study involves electronic dance music (EDM), an overarching genre that evolved from disco, dub, hip-hop and more in the late 1970's, becoming more popularized with the rise of accessible synthesizers and drum machines (Russ, 2019). The main aspect of rhythm in EDM is expressed through drumming (Butler, 2006). While symbolic

representations of sequences of musical notes don't fully encapsulate all the information about a sound, they are an effective representation that maintains the temporal and onset information. Symbolic representations are a standard way to store, process, and manipulate musical information (Collins, 2013; Russ, 2019). As Gabrielsson (1979) mentioned, timbral aspects of the sounds used affect similarity perception and thus, removing the variability of the sounds within the pattern, we can focus on the symbolic interpretation of rhythm. For the representation, we will use the common MIDI (Musical Instrument Digital Interface) protocol for the drum patterns, as is used in most digital audio workstations (DAW) for music production (Collins et al, 2013).

2.4 Rhythm Space

2.4.1 Representing Similarity through Perceptual Spaces

Mapping human perceptual knowledge to a two or three-dimensional space has scientific backing as a beneficial and utilizable system for understanding semantic memory organization. This feature is present in many diverse domains of human knowledge, such as color (Shepard, 1964), timbre (Grey, 1977), and textures (Hollins, Bensmaïa, Karlof, & Young, 2000). A perceptual space consists of a set of stimuli in some sensory domain along with a set of similarity relationships (Zaidi et al., 2013). In these perceptual spaces, similarity of two elements is represented by geometric proximity, where Euclidean distance between elements implies the level of perceptual similarity between them (Gärdenfors, 2000). Each dimension in a perceptual space represents a quantifiable characteristic measurable on all elements within the domain. One way of obtaining perceptual spaces is derived from subject-based dissimilarity comparisons, and provide an opportunity to understand the geometrical relationships of the elements in the space, be it motion effect in virtual reality (Han et al., 2022), wine (Ballester et al., 2008) or vowels (Pols et al., 1969).

As perceptual spaces derive from sensory categorization of elements, their use as a framework for categorization in auditory perception has led to novel insights in the domain of sounds and music. These spaces have been used to geometrically arrange timbre (Grey, 1977) and tonality (Krumhansl, 1979) and are proposed to be similar to the organization of

these representations in our minds. For example, Grey (1977) discovered that three metrics could be used to categorize perception of timbre similarity in sounds: the attack time, the centroid of the spectrum, and the spectral fluctuation. He suggested that with these quantifiable properties of a sound, we are able to make predictions of how similar a set of sounds will be, as perceived by humans. The organization of sounds in these structures can serve as an interface that is optimized for human perceptual understanding. This has spurred many studies of multi-dimensional analysis in timbre research (McAdams et al., 1995; Hourdin et al., 1997). Further research supports the idea that auditory perception of timbre and visual imagery of the timbre space access similar cognitive representations of timbre (Halpern et al., 2004), highlighting the potential benefits of using representative perceptual spaces in the study of music and sound.

With multi-dimensional descriptions for our data, we need a way to organize and visualize it in ways that are faithful to the original relationships within the data. There are a few techniques of dimensional reduction that are available to us to accomplish this. Multi-Dimensional Scaling (MDS) is a method for organizing pairwise similarity ratings of an arbitrary number of objects and mapping them onto a lower-dimensional space, usually two or three-dimensional (Mead, 1992). This method lends itself naturally to the analysis of similarity ratings comparing two rhythms, like in Gabrielsson (1973) and Gómez-Marín et al. (2016). MDS techniques are also widespread in the study of cognitive sciences and were used to discover fundamentals about the understanding of timbre (Grey, 1977) and pitch (Krumhansl, 1979). Other methods of dimensional reduction include t-distributed Stochastic Neighbor Embedding (tSNE) which works on principles of clustering like data, Uniform Manifold Approximation and Projection (UMAP) which works similarly but with Riemannian manifolds, and Principal Component Analysis (PCA). PCA works by remapping the coordinate system of the high dimensional data onto lower dimensions while still preserving relationships between the data. PCA does so by calculating the feature with the most effect, and then the subsequent feature with most effect once the first is removed, and so on (Jolliffe & Cadima, 2016). Like MDS, they are all used to visualize the

relationships between data that has a large set of features, otherwise known as high dimensionality.

2.4.2 Representing Rhythmic Similarity in Low-Dimensional Spaces

The principal contributor to using perceptual spaces in the study of rhythm is a set of papers by Gabrielsson (1973). In these, experimental procedures for researching rhythmic similarity, specifically polyphonic music. As a result of the subject's similarity ratings, he was able to arrange polyphonic drum patterns in two and three-dimensional spaces, with distances relating to the level of similarity between the two patterns. Desain and Honing (2001, 2003) used a three-dimensional perceptual space for organizing simple 4-note rhythms, with each spatial axis representing each inter-onset interval between the 4 notes of the rhythm. This work aimed to study human rhythm perception through a cognitive perspective, and found that participants could largely and reliably identify different rhythmic categories revealed by a *chronotopic time clumping* spatial map.

Gómez-Marín, Jordà, & Herrera (2016, 2020) explore the creation and validation of use of polyphonic drum rhythm spaces based on human music cognition and music interaction such as production, and considers previous low-dimensional rhythm/music spaces (Gabrielsson 1973; Dowling & Tighe, 2014).

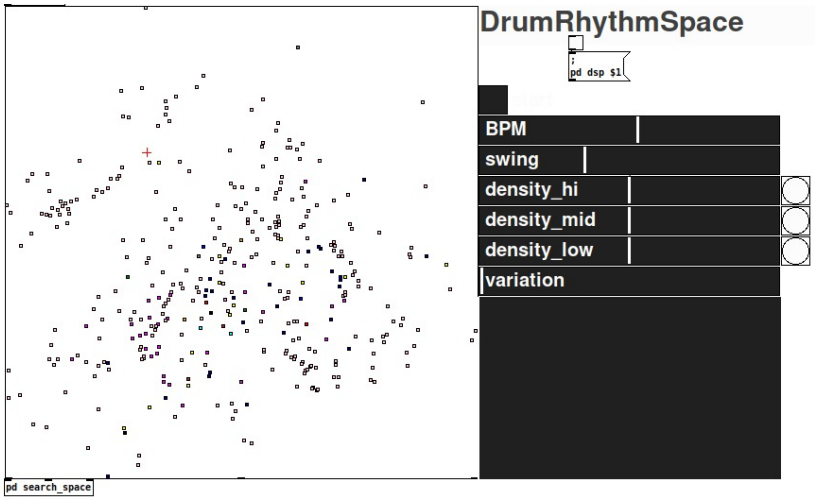


Figure 2.5: Polyphonic Drum Rhythm Space (Gómez-Marín et al. 2016, 2020). Each dot is a polyphonic drum pattern. Similarity between patterns measured by distance.

They use symbolic drum patterns, as they are represented in popular DAWs like Ableton or Logic, and reduce the expected effect of timbral differences between different drum sounds using the MIDI drumkit standard. The 1513 drum patterns used cover many genres within the grander electronic dance music field, including breakbeat, Chicago house, techno and more. With the exploration of how we process polyphonic drum patterns, they developed their polyphonic rhythm space with findings from polyphonic similarity ratings from participants, polyphonic perception (Cao et al, 2014), polyphonic distances (Gómez-Marín, Jordà, & Herrera., 2015a/b), the relevance of frequency range of instruments (Bouwer et al., 2014; Hove et al., 2014; Witek et al., 2014; Burger et al., 2017), and the polyphonic syncopation measure. Through PCA and MDS dimensional reduction techniques, Gómez-Marín et al. were able to arrange the rated drum patterns into a 2D space, finding the set of relevant descriptors for polyphonic similarity (see Table 2.1). Additionally, in this resulting two-dimensional cartesian space, they were able to extend a discrete space (filled with the rhythms the participants rated) to a continuous space through a novel interpolation algorithm for these drum patterns. The interpolation allows for users to explore combinations of rhythms as they move from one to the next. Rhythm spaces have also been used for higher dimensional data, such as in the field of Music Information Retrieval. However, due to high dimensionality, these rhythm spaces are not perceptually relevant to us, but better suit conclusions from machine learning algorithms (Chen & Chen, 1998; Makris et al., 2017).

2.4.3 Benefits of a Rhythm Space Interface

In studying the benefits of using a 2-D timbre similarity-based organization of drum sounds over traditional scroll-list or alphabetization methods, Turqious, Herman, Gómez-Marín, and Jordà (2016) found that subjects felt *implicit guidance* and a sense of freedom with the timbre similarity organization, following a process of "random exploration to fine-tuning". This is in line with theories of creativity and computational tools that support creativity (Stein. 1956; Shneiderman. 2002). Representing conceptual differences as spatial distances through embeddings allows people to navigate said conceptual space more intuitively and efficiently. There is ongoing research from the Multimodal Music Interaction Lab at UPF about the benefits of using rhythm spaces for augmenting music production.

3. Experiment 1: Rhythm Flattening

As we learned above, there are many factors that affect rhythm perception, and their relative importance (Palmer & Krumhansl,1990; Longuet-Higgens & Lee 1984) as well as how those factors relate to perceived rhythmic similarity (Gabrielsson, 1979; Gómez-Marín et al. 2020), we can now begin to construct the algorithms necessary to reduce, or flatten a polyphonic rhythm to a monophonic rhythm. By reduce, we mean to represent a polyphonic drum pattern, which has many channels, as a one-channel monophonic representation while keeping its rhythmic attributes. We refer to this process as flattening. Flattened and monophonic will be used interchangeably in this paper to describe the one-dimensional patterns. The principal approach for these algorithms will be to calculate a value for each time step that represents the perceived rhythmic salience in the larger pattern. The manner of attributing values to each note will take into account the factors of rhythm perception explored above. Once the algorithms return monophonic patterns that are not trivial, we can proceed to refine them with results from the literature. The output of the rhythm flattening algorithms will be used to train mdoels to predict the monophonic pattern's parent polyphonic pattern's location in the rhythm similarity space provided by Gómez-Marín et al. (2020).

3.1. Flattening Algorithms

The process of designing an algorithm that can effectively use the myriad of factors that influence rhythm perception to flatten a polyphonic drum pattern is a complex one, but standing on the shoulders of previous research, we have an excellent place to start. First, let us define the type of data we are working with for the rhythm flattening. The input polyphonic patterns that we are working with are the same 1513 patterns used in the creation of the polyphonic rhythm space (Gómez-Marín et al. 2020). They are all in 4/4 common western meter and span a number of electronic music genres, including various forms of Chicago house, techno, breakbeat, and reggaeton. They are represented by a sequence of 16 steps at ¼ note fidelity, containing a channel for all drum notes represented in the MIDI protocol. The term *parent pattern* will be used interchangeably for polyphonic drum pattern.

The output representation will take the form of a 1 or 0 occupying 16 steps at ¼ note resolution. In attempts to reduce the effects of timbre and number of instruments on rhythmic perception of the parent patterns, the set of MIDI drums were reduced to 8-channels representing a simple drum kit consisting of: Kick drum, Snare drum, Clap, Low Tom drum, High Tom drum, Open Hi-Hit, Closed Hi-Hat, Rimshot. Each instrument used the same sound when listened to out loud for the duration of this exploration and later experiment.

Naïve Flattening Approach

The first aspect we focus on is the agnostic onset of notes. This simple naïve flattening of a polyphonic rhythm is to have the monophonic pattern contain a note, if at any step, a MIDI note of any kind appears in the polyphonic drum pattern. A simple flattened pattern provides a low level of information, failing to capturing distinguishing features between two rhythms unless the parent patterns were very sparsely filled and of different genres. When the parent patterns were complex and filled with a range of MIDI notes, the naïve flattening failed to provide any identifying information and were often maximally filled.

Instrument-Based Flattening Approaches

As specific instruments have been used in rhythm studies in the past, including a snare drum (Fuji et al. (2011); Câmara et al., 2022; Frühauf et al. 2013), a kick drum (Hove et al. (2014) or a clap (Velautham & Yoong, 2022). We constructed an instrument-specific algorithms where a flattened note was output if either a kick, snare, and/or clap, or a simultaneous combination of them were present at that time-step. This method, both visually and auditorily demonstrated a good relationship between the parent pattern and the resulting flattened rhythm, however suffers from the fundamental flaw of being limited to polyphonic drum tracks containing a kick, snare, or clap. While most tracks in EDM have these elements in varied fashion, we do not wish to limit ourselves to just those specific instruments. The idea of using specific instruments relates to the broader idea of how different frequencies affect our perception of rhythm and sound (Krumhansl, 1990; Bouwer et al., 2014; Hove et al., 2014; Witek et al., 2014; Burger et al., 2017), and the studied effects of frequency ranges in

distinguishing aspects of polyphonic rhythmic similarity (Gabrielsson, 1973ab; Gómez-Marín et al. 2020).

Frequency-Based Flattening Approaches

To extend the instrument-based algorithms to the desired more generic applications, we moved onto separating all of the MIDI drum sounds into three frequency ranges (*low, mid, high*) by their spectral centroid before considering rhythmic relevance of any note. Therefore, we are making another interim reduction from a full polyphonic drum pattern to a three-channel polyphonic pattern, based on its frequency. This allows us to consider all types of drums. Splitting into three frequency ranges will also allow us to examine how rhythmic factors such as syncopation affect specific groups of instruments that share perceptual rhythmic similarity. We know that low channel sounds contribute heavily to the rhythm, followed by the mid channel (Hove et al. 2014), and often times in 4/4 music the kick drum is used to reinforce the meter (Butler, 2006). With this information, an approach we take establishes the hierarchy of sound importance by frequency channel, with low channel being the most important. We explore two flattening algorithms with variations of this approach. The first, a flattened note only is output if a) a note in the low channel exists, b) a note in mid channel exists & none in the low channel and c) a note in the high channel exists & in no other channel. The second only considers if a) a low note exists, or b) if notes exist in mid and high channel simultaneously. This frequency approach over a specific instrument one takes into account much more of the notes occupying the middle and higher frequency ranges, such as toms, hi-hats, shakers, maracas and so on. The higher range of instruments often appear as *rolls*, or repeated short length notes, and often are placed to fill the gaps of the lower frequency sounds (Collins, 2013), and so will have notes populating more the flattened rhythm.

Onset Density and Metrical Flattening Approaches

With a focus on the rhythmic flattening, we have to consider the main factors of rhythm perception shared in section 2.1/2.2: the effect of the meter, the effect of syncopation, and the effects of note onset density. The frequency approach above indirectly makes us consider

the density of notes both overall and in each frequency band, but with an inverse affect. The more notes in the parent pattern that appear in each frequency channel, the more likely they would be to appear in the flattened pattern. This is contrary to research about sparser patterns being easier to replicate (Milne & Dean, 2021). An approach considering the inverse of relative density of note onsets in each frequency channel to the total amount of note onsets in the pattern takes that into consideration. A sparser pattern will have its individual note onsets weighted higher than a dense pattern.

Longuet-Higgens & Lee (1984) provide us with a manner of calculating metrical strength of each note for 4/4 patterns, and Palmer & Krumhansl (1990) provide a updated metrical weighting for those calculations.

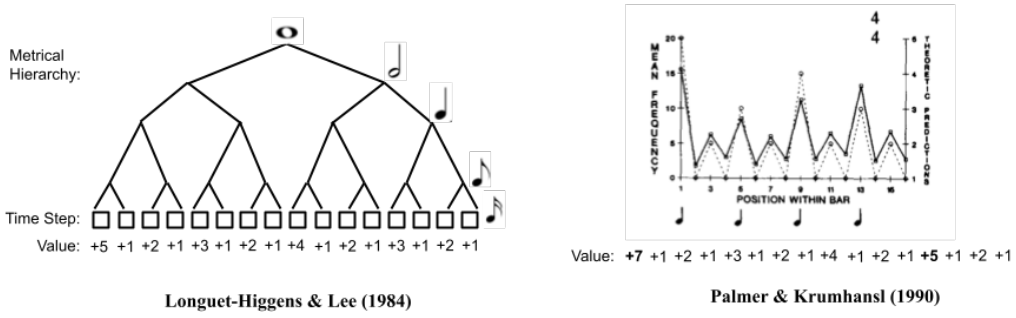


Figure 3.1: Metrical Strength Calculations by Position

In a variety of parent patterns from different genres, just considering whether a note reinforces the meter ignores the role of syncopation. In the cases where a note is syncopated, we shift the values of the metrical strength one time position to the right to get syncopation strength (in the cases where a note is syncopated: has a rest of higher metrical value preceding note onset). We call this the syncopation salience for a note. As all notes have a value, a note is output in the flattened pattern if it is higher than the lowest calculated value. In order to avoid arbitrary thresholds, deciding if a note is metrically important enough to include in the flattened rhythm should be based on the polyphonic drum pattern it comes from.

3.1.1 Discrete, Semi-Continuous, & Continuous Representations

Discrete	The velocity value of notes that are present in the flattened rhythm are 1, if they are above or equal to the normalized mean, otherwise their value is 0.
Semi-Continuous	If the value of notes that are larger than the normalized mean, they appear in the flattened rhythm at their calculated velocity value. If they are below the normalized mean, their value is 0.
Continuous	All calculated values of notes appear in the flattened rhythm.

Table 3.1 Three different representations types of flattened rhythm.

With just a discrete representation of a note being included or not, we began to notice that there wasn't enough information in the flattened rhythms to separate their respective rhythmically distinguishable polyphonic patterns. Different fidelities of representations have different capabilities in the amount of information they contain, so we consider three different levels. In this way we are making the flattened rhythm a bit more complex by adding another dimension. However, this dimension will only be physically represented by the volume (velocity[1]) of the same monophonic output note. This is important as we can evaluate if included the added information from the different representations improves the predictive capabilities of the flattening algorithms.

To provide a threshold for inclusion from each parent pattern that is not arbitrarily chosen, we first devise a weight for each note that we can then apply the metrical strength to. A way to do this is based on the relative note onset density for each frequency channel in the polyphonic drum pattern. The method we used to arrive at a self-normalized salience for notes of each frequency channel is as follows:

- Find number of notes in each frequency channel.
- Divide number in channel by total number of notes in a polyphonic pattern to find relative channel density.
- Divide 1 / relative channel density to get the salience for notes in their respective channels.

[1] Velocity is the intensity of a MIDI signal, interpreted as volume / strength of signal.

- Normalize salience by dividing by sum of salience for all channels.

Consider a simple polyphonic drum pattern with 8 notes in total. 2 in the low channel, 2 in the mid channel, and 4 in the high channel.

	Relative Note Density	Salience	Normalized Salience
High	4 / 8 (0.5)	1/0.5 (2)	2/10 (**0.2**)
Mid	2 / 8 (0.25)	1/0.25 (4)	4/10 (**0.4**)
Low	2 / 8 (0.25)	1/0.25 (4)	4/10 (**0.4**)
	#notes (channel/total)	4+4+2=10= Σ(Salience)	Salience/Σ(Salience)

Table 3.2 Normalized Salience per Note per Channel Example Calculation

We now have a concrete value for the weights of each note, we can use the method of calculating metrical strength from Palmer & Krumhansl (1990) and our derived syncopation strength metric to arrive at a value that represents the rhythmic salience of a note in the polyphonic pattern. When we have this value for every note in the parent pattern, we can move on to calculating a threshold. At each time step, sum the notes in that time step, and calculate the mean rhythmic salience over the pattern. Time-steps with an equal or greater rhythmic salience value than the normalized mean will appear in the pattern, and those below will not. We normalize the mean once more to keep the range of the values within zero and one[2]. In this way, the decision to include a note in the flattened rhythm is self-regulated for each individual pattern, and no arbitrary thresholds are chosen.

An example of the three different representations of the same pattern is shown in Figure 3.2. All time-steps in the pattern that are gray are played at their respective velocities, and all time-steps in white are silent.

[2] As we will be passing the note values to an audio output, normalizing the values allows them to serve as a volume or velocity value.

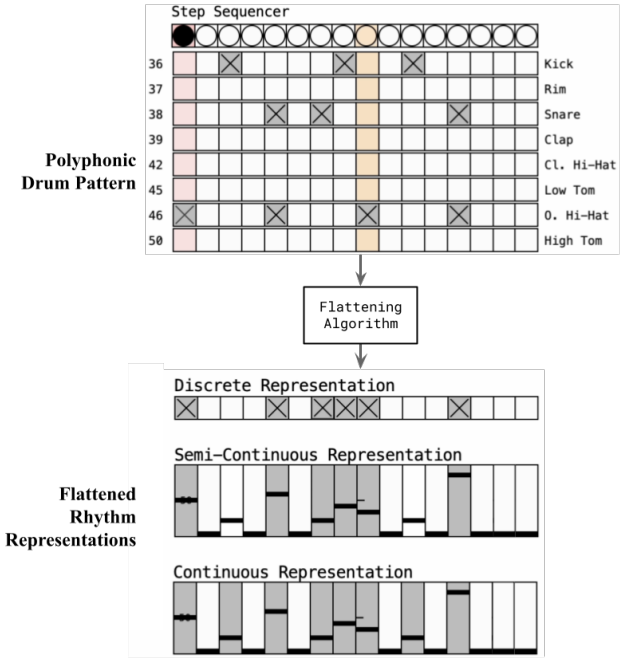


Figure 3.2 Three different representations types of an example flattened rhythm

With the set of representations for each flattened rhythm and the exploration into methods for calculating strength at each time step, we can test a set of flattening algorithms to identify which algorithms are the most promising. Each of the selected flattening algorithms will be tested for each representation type, for a total of six flattening algorithms. The subset of flattening algorithms which we will test consider:

1. Note Onset Density (as shown in the calculation above) and the syncopation weights if note onsets are syncopated.
2. Note Onset Density (as shown in the calculation above), syncopation weights if note onsets are syncopated, metrical strength weights for note onsets (Palmer & Krumhansl., 1990).

3.2 Evaluation of Flattening Algorithms

Having several flattening algorithms, an experiment was devised. The goal is to measure by which account a flattened representation can be used to find the location of its parents in the rhythm space. The algorithms that best predicts the position of the polyphonic patent pattern will be considered the best algorithms. Five types of machine learning regressions were used: {Linear regression, k-Nearest Neighbors, Decision Tree, Support Vector Regression, & Deep Learning}. The mean absolute error between predicted and real positions was chosen as the error metric because we are searching for how far away our coordinate prediction is compared to the known location of the polyphonic pattern, and having an accuracy scale between 0-100 provides an intuitive understanding of performance of the network.

The inputs to the models are algorithms' flattened rhythm representation of a polyphonic drum pattern, which are a sequence of 16 numbers ranging 0-1. The outputs are a set of 2-D coordinates in the polyphonic rhythm space. The regression types were chosen as they reflect the nature of the problem and input-output relationship. Linear regression models the relationship between the variables by fitting a linear equation to the data. We expect this to show the non-linearity of our data. k-Nearest Neighbors (kNN) is a clustering algorithm based on similar outcomes. Decision Trees recursively split the data into a tree of decisions that classify the outcome. Support Vector Regressions (SVR) are a regression algorithm for multi-dimensional data, and are good at minimizing error. Deep learning is field of machine learning that uses hidden layers to approximate the relationship between the variables.

Using Python and the TensorFlow (Abadi et al. 2016) package, we are looking to build a light-weight model that's minimally resource intensive in training, that can provide us consistent results with sufficiently minimal distance to the original patterns coordinates. This is especially the case as the process needs to be completed for each algorithm and representation type. For the deep learning analysis, we use a fully-connected feed-forward neural network. To accomplish this, a grid search was done alongside some final manual hyperparameter tuning to navigate the hyperparameter space and optimize our network. The

best scoring neural network over all the algorithms was selected as our prediction model. The following hyperparameters were selected from the optimization:

Input Layer	Hidden Layer 1	Hidden Layer 2	Hidden Layer 3	Output Layer	Batch Size	Training Epochs	Learning Rate
16	32	16	8	2	32	200	0.001

Table 3.3 Neural Network Hyperparameters

As mentioned, the input layer is list of numbers with the length of the monophonic pattern, and the output layer being the *x* and *y* coordinates in the rhythm space. The hidden layer sizes are all multiples of 4, chosen to reflect the 4/4 common meter of the patterns we are examining. A ReLu[3] activation function was used in between every layer. An increase in training epochs led to overfitting across all algorithm types and hyperparameter selections. Likewise, an increase in learning rate led to overfitting even within a relatively small number of epochs. Vigliensoni et al. (2022) also found in their latent space rhythm models that overfitting routinely to happen at around 250 epochs, even with a larger model. After the structure was selected, we train and test the model 10 times with an 80/20 train-test split so we can see how the model performs on average. The results of the regressions hints at some leading candidates for predictability of the patterns location in the polyphonic rhythm space. Mean absolute error was used as a metric here, as we are trying to optimize for a distance in the Euclidean-natured rhythm space (Gómez-Marín et al. 2020). A trend is starting to emerge related to the representation levels of the algorithms (see Table 3.5). As we move from the discrete algorithms to the continuous ones, all regression results reduced the mean absolute error, which suggests that this pattern will likely be followed when tested with neural networks.

[3] Rectified Linear Unit

Regr. Type ↓	1. Density & Syncopation			2. Density & Syncopation & Meter		
	Discrete	Semi-Cont.	Continuous	Discrete	Semi-Cont.	Continuous
Linear	0.433 (0.018)	0.380 (0.017)	0.343 (0.018)	0.443 (0.019)	0.378 (0.015)	0.356 (0.017)
kNN	0.395 (0.037)	0.274 (0.024)	0.110 (0.017)	0.562 (0.024)	0.351 (0.026)	0.114 (0.018)
Dec. Tree	0.289 (0.022)	0.141 (0.015)	0.103 (0.013)	0.383 (0.022)	0.217 (0.016)	0.089 (0.014)
SVR	0.407 (0.021)	0.298 (0.022)	0.132 (0.020)	0.421 (0.021)	0.363 (0.019)	0.322 (0.021)
Deep Learning	0.122 (0.107)	0.070 (0.090)	0.039 (0.057)	0.141 (0.122)	0.099 (0.105)	0.041 (0.058)

Table 3.4 Regression Type vs Algorithm Type Results [Mean Abs. Error (std)]

The discrete→continuous improvement trend seen in the regressions is also present here, indicating that the extra information provided by the continuous representation is not extraneous. The best performing model was Density & Syncopation (Continuous) followed closely by Density, Syncopation & Meter (Continuous) with an honorable mention to Density & Syncopation (Semi-Cont.). While the best model was able to earn a low average MAE of 0.0392, it's standard deviation (0.0575) points to a slightly wider distribution of prediction errors and more outliers. The same pattern follows for all algorithms, where the standard deviation is close to or as large as the MAE.

This machine learning analysis is by no means conclusive, exhaustive or final. The intent of this analysis is to show that with a lightweight model, we can achieve one of the principal goals of this investigation: to reasonably extend a monophonic pattern to polyphonic rhythm space and compare the performance of the different flattening algorithms.

4. Experiment 2: Rhythm Tapping

Now that we have some research-backed predictions derived from the previous experiment for what a flattened rhythm should look like when attempting to locate the position of a

polyphonic drum pattern in a rhythm space, we can move onto testing how subjects would tap to a polyphonic pattern. Tapping in the context of this experiment refers to tapping on a pressure sensitive physical interface. In out experiment, we used an Ableton Push 2 instrument, a music production device that contains velocity-responsive drum pads. Polyphonic drum patterns are presented to subjects maintaining the same synthetic timbres across all patterns. These factors are chosen to reduce the intrusion of timbre in subject's rhythmic assessment. These experiments with symbolic drum patterns are also suitable as they resemble normal conditions of composition in EDM (Collins et al., 2013), as well as composition with the Ableton Push 2 itself. The goal of this experiment was not to one-off test subject's responses to a polyphonic drum pattern, but instead to observe their ability to submit a monophonically tapped rhythm that they felt was an accurate representation of a parent pattern. Instead of measuring an immediate response, we hope that ability to re-tap a pattern will help to eliminate some of the expected individual differences or potential errors in tapping due to the varied levels of musical experience of the subjects (Nistal et al. 2017; Milne & Dean. 2021).

Based on previous research, we hypothesize that participants will be able, to a certain degree, to tap within velocity/pressure ranges. We also hypothesize that the best scoring flattening algorithm in the previous experiment will bear the most similarity to the mean tapped rhythm from all subjects in this experiment.

4.1 Materials and Procedure

4.1.1 Participants

Twenty-six subjects participated in this experiment. Subjects came from varied background, including university students and staff at UPF. The mean age of the participants was 34.5 years old (SD=12.7). 16 subjects identified as male and 7 as female. Participants have spent 3.5 years (SD=1.4) playing an instrument, 2.9 years (SD=1.8) studying music, 2.6 years (SD=2.1) performing musically, and 2.5 hours (SD=1.6) interacting with music daily.

4.1.2 Materials

The experiments were created with Python and PureData (Puckette, 1996), a visual programming language. The equipment used included a MacBook Air laptop upon which the experiment was run, Marshall Major IV headphones (wired connection), and an Ableton Push 2 (drum pad) used for its velocity sensitive drum pads.

4.1.3 Procedure

Subjects were allowed to position the drum pad wherever was most comfortable for them to use it. This experiment was divided into two tasks: Tapping a drum-pad consistently aiming for a target velocity/intensity range, and trying to tap along to what they perceive is the rhythm of the polyphonic drum pattern. Both tasks are repeated-measure and all subjects participated in reproducing all of the provided polyphonic drum patterns. Before the experiment began, subjects were given a set of instructions, an informed consent form. The informed consent covered subject's data rights under the GDPR and their rights over the data collected in the following experiments. When the form was returned signed, a set of questions about the subjects' musical abilities were asked. They concerned the amount of time a subject has played an instrument, studied music, performs musically, and interacts with music on a daily basis. This was followed by repeating the task instructions for the subject, alongside answering any questions a subject may have before the trials begin.

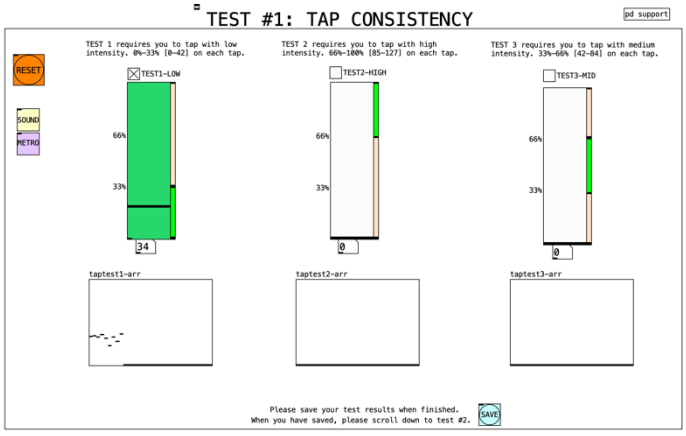


Figure 4.1 Task 1: Tap Consistency Task View

Task 1:

The first task consists of three subtasks. Each one of these tasks asks the subject to tap on the provided drum pad as consistently as possible in the three different target ranges: $[0, \frac{1}{3}]$, $[\frac{1}{3}, \frac{2}{3}]$, $[\frac{2}{3}, 1]$, referred to as *low*, *mid*, and *high* ranges respectively. The range covers what the Push drum pad can register on a tap, from a MIDI value of 0-127. The tests were taking the L-R order as shown in Fig. 4.1. Each subject tapped 32 times for each range. Figure 4.1 shows the visual feedback for a successful tap in the target range, auditory feedback as volume reflecting the intensity of the tap, as well as the results of the subject's previous taps in that subtask below. Subjects could clear the results of all of their tests if they wished, but then they must start again from the beginning. This was allowed as there were subjects not familiar with the equipment. When a subject completed all three subtasks, they could press *save* and move onto the next task. The purpose of this first task is familiarizing subjects with the equipment they will be using in the main task of this experiment, as well as provide a benchmark for use of this equipment in further studies. This is because subjects do not all use the same single drum pad out of the selection of 64 (8x8) on the Ableton Push, nor are they necessarily familiar with said drum pad or any forms of sound or music production, or tapping with consistency.

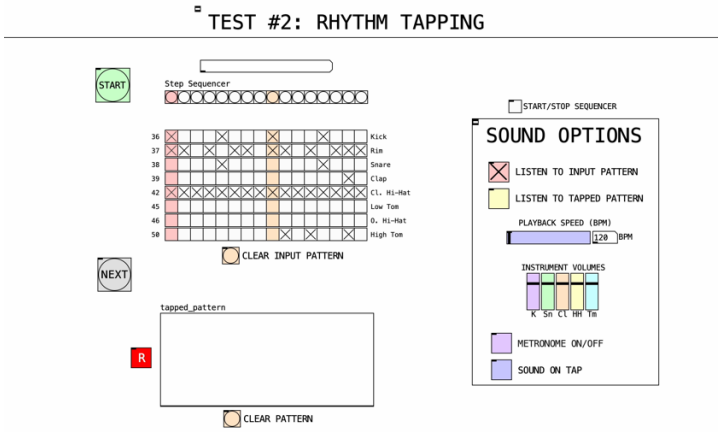


Figure 4.2: Task 2: Rhythm Tapping Experiment View

Task 2:

In the second task, participants were presented with a screen showing an eight-channel drum sequence with the instruments mentioned in section 3.1. Below subjects could see their inputted pattern from the drum pad, working on the same step-sequencer/clock. When they were ready, the start button was pressed to begin the trials, see fig 4.8. The subjects were instructed to tap along to what they perceive the underlying rhythm of the polyphonic drum pattern was. Subjects could control the volume of each instrument, the playback speed, whether just the input or the tapped pattern was playing, and if they wanted a metronome. When a subject was content with their tapped rhythm, they press *next* and repeated until there are no more rhythms. Each subject reproduced 18 rhythms, 16 of which are unique and 2 that are repeated once for a control measure. The rhythms were delivered in a random order for each subject, intending to minimize the effect of training on the results. The polyphonic drum patterns were chosen for their positioning throughout the rhythm space (Gómez-Marín et al. 2020) and their variety of genres. The patterns names are listed in the appendix, are shown in Figure 4.3 with a number identifier.

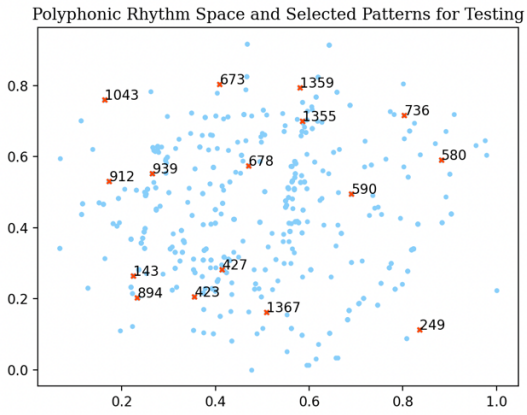


Figure 4.3 Simplified Rhythm Space and Selected Patterns

4.2 Rhythm Tapping Experiment Results

Subject Data Processing

By having control patterns, we have a way to compare how consistent a subject is at tapping the same pattern vs the overall mean from all subjects. This can give us insight about how

an individual performs on all patterns, and can identify outliers. Our method for removing outliers is as follows: Take the mean tapped pattern for each control pattern, by taking the average tapped value at each time step. Compare the subjects two attempts at the control pattern and calculate how the subject differs from the mean tapped pattern. If the subjects' summed error is outside a threshold of [Mean Summed Error + 1.5 Std. Dev.] for either control pattern, they are considered an outlier and are removed from the data pool (see fig 4.4).

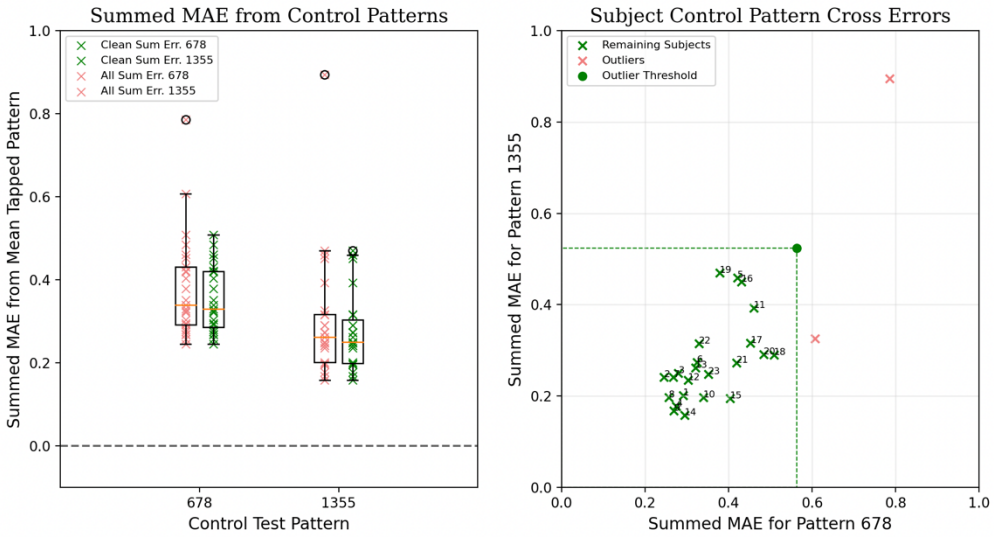


Figure 4.4: Subject summed MAE in control patterns by pattern. (L) is the distribution by pattern. (R) compares a subject against both control patterns. Green subjects are kept for further analysis, red discarded.

From the original set of 26 subjects, through this filtering method 2 outliers were identified who showed a large inconsistency when tapping the control patterns. It is known that subjects will have a large variance in results, and that between subjects there will also exist discrepancies. (Nistal et al. 2017; Milne & Dean. 2021), so it is reasonable to assume there will be outliers of this sort in experiments that concern tactile rhythm reproduction.

Task 1 Results

After removing outliers, we can now look at the first task in the experiment, tapping consistency. In Figure 4.5, we can see how all subjects performed over the 32 taps required

for each tapping range. In purple, we have the high tapping range, in blue the mid tapping range, and green for the low range. The solid lines indicate the mean tapped value across all subjects for each range at that tap. We can see that the mean tapped value for the high tap range has more fluctuations that the other two tap ranges, and that initially, subjects tended to tap at lower values than the high tap range. To explore this, an analysis of variance (ANOVA) was conducted on the results split into 4 quarters by tap order. The results of the ANOVA between the first, second, third and fourth quarters of each tapping task are: F(3,19)→ F=6.995 (p=0.00037). This indicates that Tukey's Highly Significant Difference (HSD) test was needed to identify the specific pairwise differences between the quarters.[4]

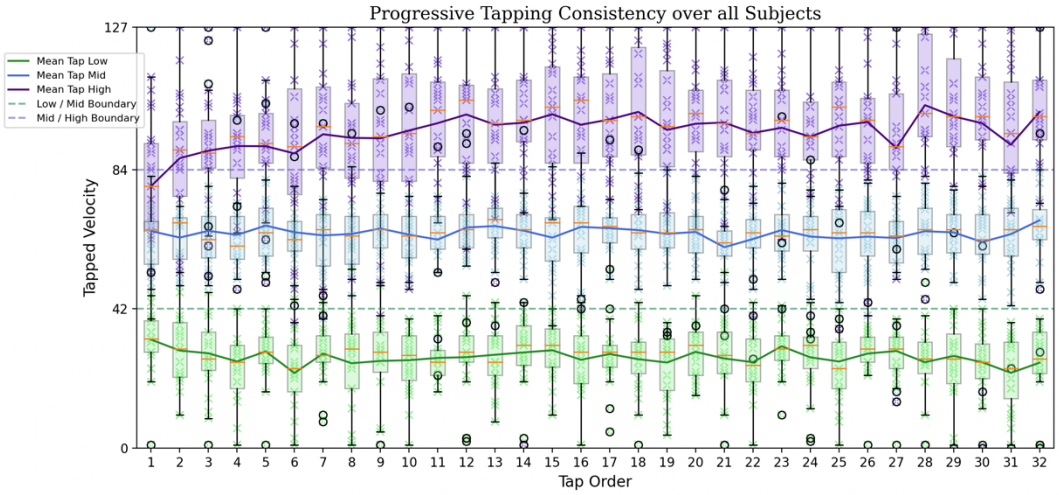


Figure 4.5: Aggregated tap consistency results over all subjects. Tap range task by color.

Results of Tukey's HSD between Q1, Q2, Q3, Q4:

- The mean tapping consistency for Q1 (M=0.1441, SD=0.0711) is significantly *lower* (MD=-0.0544, p=0.0140) than for Q2 (M=0.0896, SD=0.0513).
- The mean tapping consistency for Q1 (M=0.1441, SD=0.0711) is significantly *lower* (MD=-0.0699, p=0.0009) than for Q3 (M=0.7742, SD=0.0392).
- The mean tapping consistency for Q1 (M=0.1441, SD=0.0711) is significantly *lower* (MD=-0.0667, p=0.0014) than for Q4 (M=0.0773, SD=0.0249).

[4] M=Mean. SD = Standard Deviation. MD=Mean Difference. p=p-value.

- The rest of the pairwise interactions between the quarters were all *not significantly different*: Q2-Q3 (p=0.8125), Q2-Q4 (p=0.8912), Q3-Q4 (p=0.9978).

From the pairwise comparisons, it becomes clear that subjects performed worse as a whole in the first quarter of the tapping consistency tasks as compared to the rest of the task, when considering all ranges. This lends itself to the idea that people will learn how to better convey their tapping intentions with increased training.

In Figure 4.6, we can see a subject's mean absolute error and spread for each tap range. Immediately apparent is the larger spread and higher number of subject responses outside of the target range in the high tap range test condition (purple boxes). For example, subject 12 was consistently in the target range for the low and mid tap range tasks, but not so in the high tap range task.

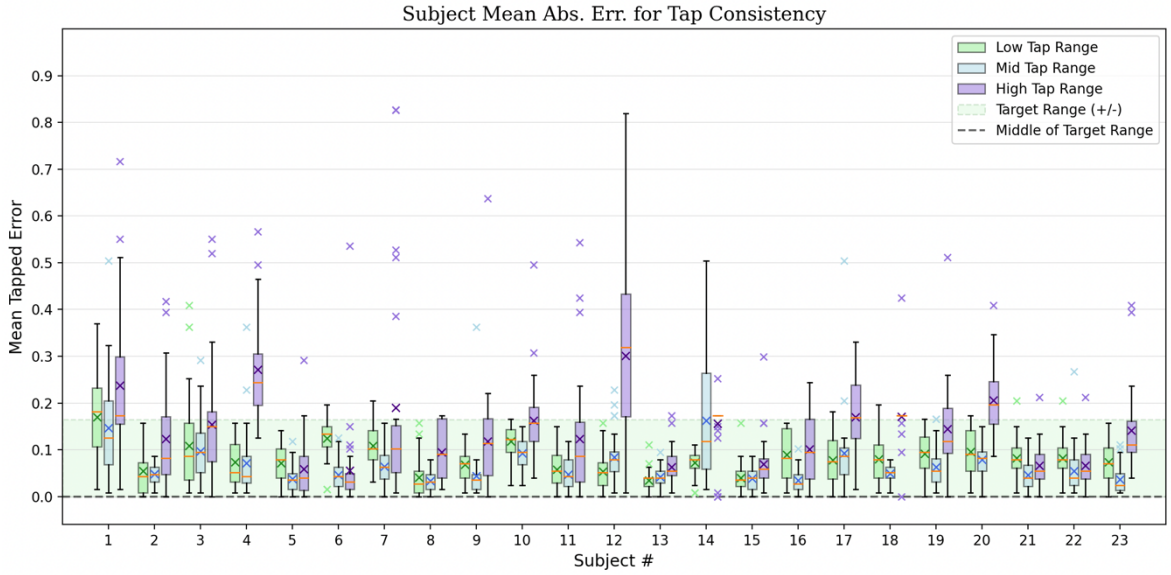


Figure 4.6: Subject mean abs. error by test type. Results within the green rectangle were within the target range in the task.

An (ANOVA) was conducted on the results from three tapping ranges: low, mid and high. The results of that ANOVA are: $F(2,66) \rightarrow F=16.0611$ ($p=0.000002$). This indicates that there is a significant difference between the mean absolute error of the different test types; to figure that out, a Tukey's HSD test was conducted to identify the specific pairwise differences.

- The mean absolute error from the low tapping range test (M=0.0813, SD=0.0301) was significantly *lower* than from the high tapping range (M=0.1411, SD=0.0667). Tukey(low, high) → (MD= 0.0598, p=0.0002)
- The mean absolute error from the mid tapping range test (M=0.0.0657, SD=0.0338) was significantly *lower* than from the high tapping range (M=0.1411, SD=0.0667). Tukey(mid, high) → (MD= 0.0754, p<0.000001)
- The mean absolute error from the low tapping range test (M=0.0813, SD=0.0301) was *not significantly different* than from the mid tapping range (M=0.0657, SD=0.0338). Tukey(low, mid) → (MD= -0.0156, p=0.51)

These results indicate that subjects had the most errors in the high tap range task as compared to the low and mid tasks, while the low and mid tasks performed similarly well. The large amount of spread mostly seen in the high tap range result in purple (Fig. 4.6) visually reinforce the results from the pairwise comparison.

Task 2 Results

With the results of the tapping consistency task done, we can now observe at how subjects reproduced polyphonic drum patterns by tapping. Since every subject tapped to every pattern, we can get a mean tapped value for each time step in each of the subjects' tapped pattern. In Figure 4.7, we see how participants responded to 4 of the 16 patterns in the test. The solid red line is the mean subject-flattened rhythm for the polyphonic pattern. The rest of the patterns are included in appendix II. We will use the average tapped value for each pattern as the baseline to compare the individual subjects to. By doing so, we can see how the tapping range of a subject compares to the whole, whether subjects had the tendency to tap stronger or weaker than the average, and the range of a subject's taps in general. With this average from subjects, we can show what can be thought of as a rhythmic profile for each polyphonic pattern.

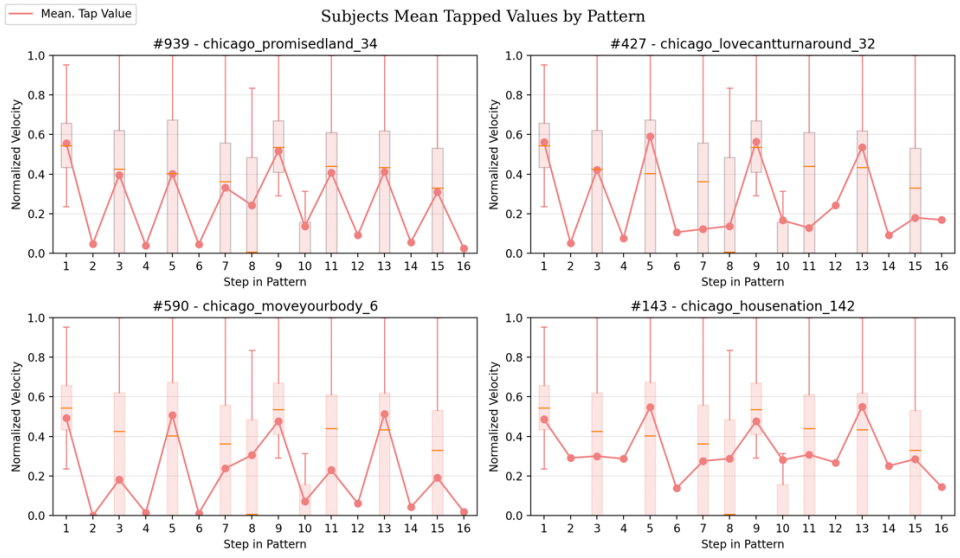


Figure 4.7: Mean Tapped Velocity per Pattern. (4 of 16 patterns shown)

In figure 4.8 we see individual subject's results comparing to the baseline mean tapped rhythms for each pattern. The graph on top shows us the mean absolute difference, which gives us an idea of how far off each subject was from the mean tapped rhythm. The dashed line reflects the mean differences of all participants (M=0.190, SD=0.049). The graph on bottom shows us the direction of that difference, with negative values in this graph meaning that the participant tapped softer compared to the mean. The dashed line in the bottom graph and the solid line in the top are the same line. In the top graph, we can begin to discern a little bit about participants tapping styles. The spread of each subject's mean tapped error may indicate that they are using a larger range of tapping intensity than others. In the bottom graph, we can see whether participants had a tendency to tap softer or harder than others. Some participants, 2, 7 and 8, tapped very similar to the mean tapped rhythm that we extracted from all subjects, and the differences in their taps is relatively balanced between too soft and too hard. This suggests that there are participants tapping behavior that reflect the average with some degree of accuracy. Again, we expect to see a large variance between subjects tapping performance (Nistal et al. 2017; Milne & Dean. 2021).

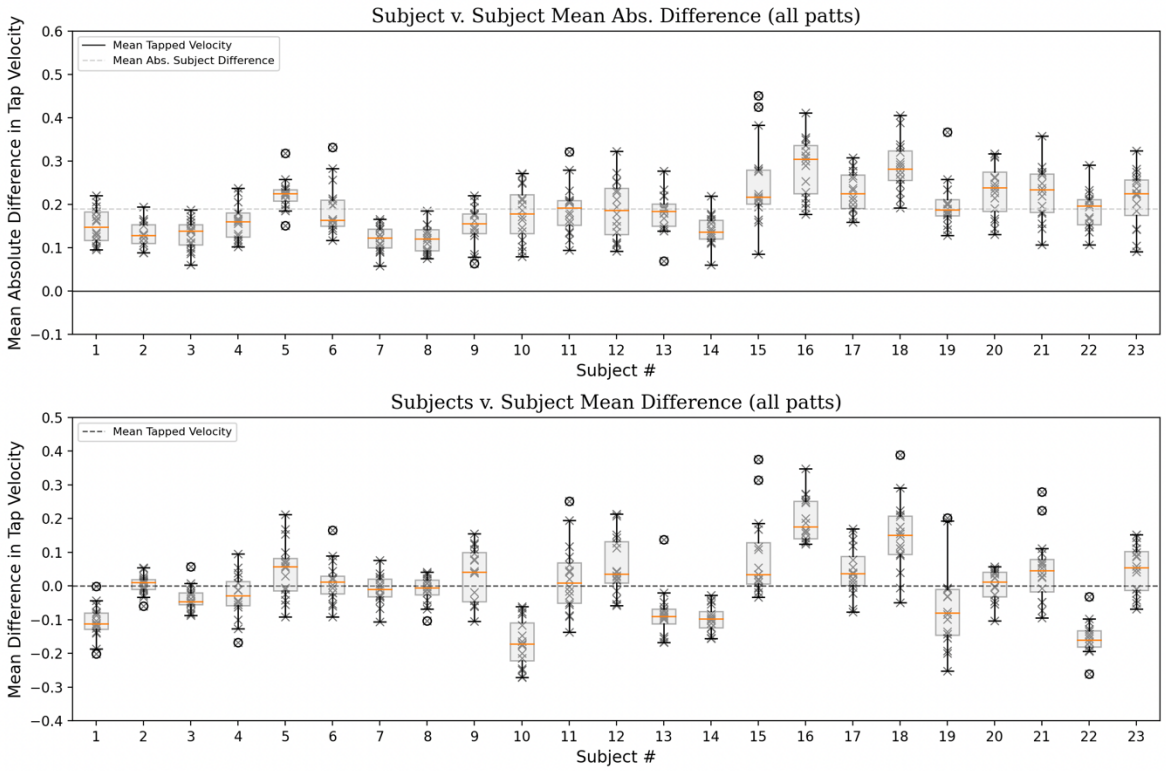


Figure 4.8: (Top) Subject Absolute Tapped Difference over all Mean Tapped Rhythms. (Bottom) Subject True Tapped Difference over all Mean Tapped Rhythms.

In order to make sure any individual patterns were not problematic; we look at the tapping behavior for each pattern. An ANOVA was run on the mean absolute difference to the mean tapped rhythm; $F(15,362) \rightarrow (F=13.17, p=7.64\text{x}10^{-34})$. Following this, a Tukey's HSD test was conducted to see if any particular pair of patterns different significantly. The results showed the none of the pairwise comparisons between the 16 test patterns were significant. Similar results arise when the ANOVA was run on the true mean difference to mean tapped pattern; $F(15,362) \rightarrow (F=21.17, p=3.43\text{x}10^{-51})$. A Tukey's HSD likewise showed no significant differences between the true mean differences of the test patterns. We interpret this to mean that subjects performed better in no individual pattern compared to another, but have plenty of variance within the patterns and their overall mean difference from the average tapped rhythm. Looking at figure 4.9, the mean absolute and true differences are show top and bottom respectively. While some patterns show considerably more spread (894 vs 143), no pattern had a significantly larger difference than another.

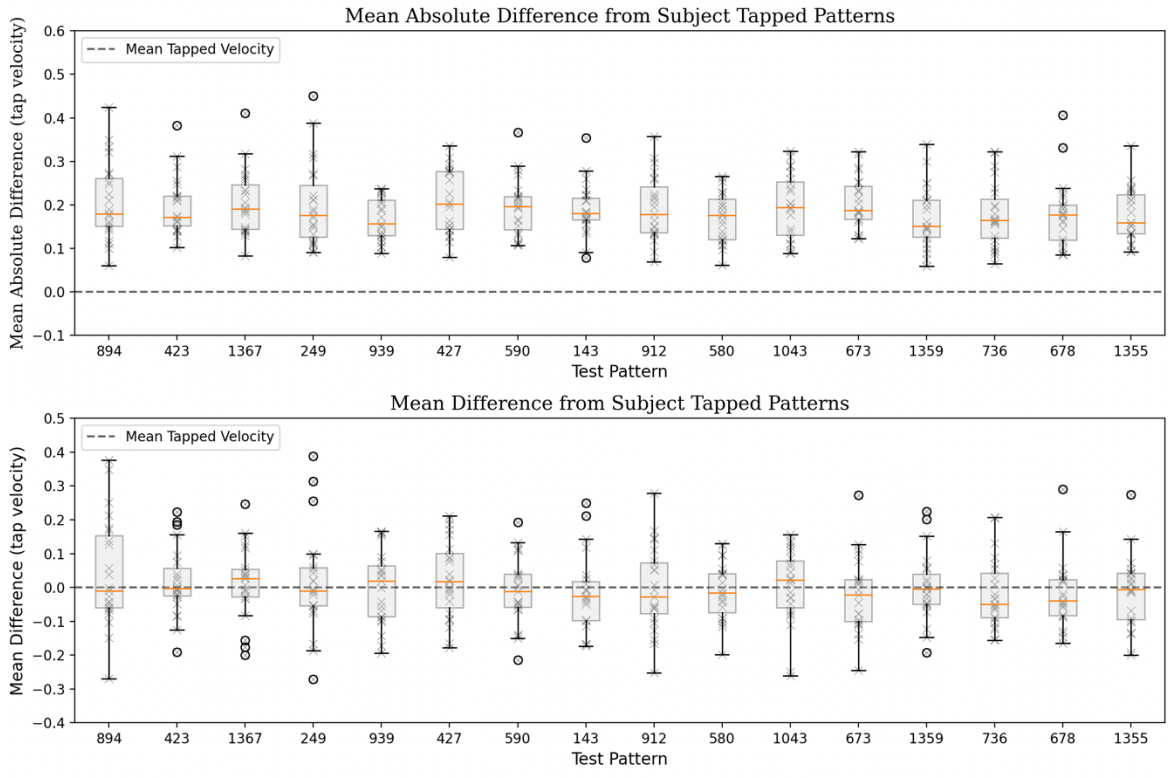


Fig 4.9: (Top) Pattern Absolute Tapped Difference over all Subjects.
(Bottom) Pattern True Tapped Difference over all Subjects.

With the patterns not showing any noticeable issues in subjects' performances and the distribution of patterns selected from the rhythm space (fig 4.3), we can extrapolate and say that participants most likely will not have any issues with a particular region in the rhythm space.

4.3 Comparison with Flattening Algorithms

For every polyphonic pattern, we now have a subject-flattened rhythm to compare with the flattening algorithm predictions. We will look at a few error metrics to get an idea which of the flattening algorithms is most similar to the subject-flattened rhythm. Those are the mean absolute error, mean squared error (MSE), root mean squared error (RMSE), and the R^2 metric (RSQR).

	1. Density & Syncopation			2. Density & Syncopation & Meter		
	Discrete	Semi-Cont.	Continuous	Discrete	Semi-Cont.	Continuous
MAE	0.1924	0.1424	0.1176	0.1924	0.1362	**0.1145**
MSE	0.0704	0.0385	0.0336	0.0704	0.0344	**0.0298**
RMSE	0.2654	0.1963	0.1832	0.2654	0.1854	**0.1726**
RSQR	-0.3911	0.2386	0.3368	-0.3911	0.3211	**0.4111**
CV / CV [subjects]	2.009	2.2278	**1.5346**	2.2907	2.6425	**1.915**

Table 4.1 Error Metrics and Coefficient of Variance (CV) by Algorithm Type

For all error metrics except one, the continuous version of the density, syncopation, and meter algorithm performs the best. We calculate the coefficient of variation (CV) to assess the stability of the mean differences to the tapped pattern across the algorithms. In the table, we take CV / CV of the mean tapped pattern, to compare the amount of variability in the algorithms predictions versus the subject average. The continuous density and syncopation algorithm had the least variability in its mean difference rates. A Tukey's HSD test was conducted on the algorithms and the pattern means. The only significantly different algorithm compared to the mean was discrete density and syncopation (MD=-0.1076, p=0.0118), ruling it out of future consideration without modification.

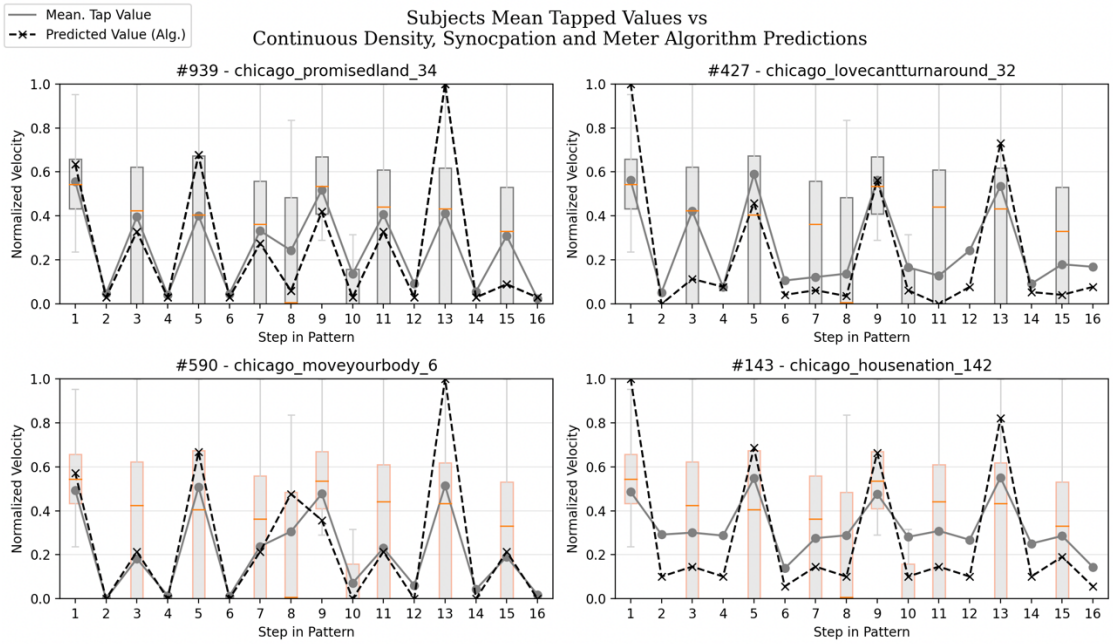


Figure 4.10: Comparison of Continuous Density, Syncopation and Meter Algorithm predictions and Mean Tapped Value for 4 patterns. Same patterns located in fig. 4.7.

In Figure 4.10, we plot the predictions of the continuous version of the density syncopation and meter flattening algorithm. From a visual standpoint, the algorithm (green) appears to trace the general movement of the tapped pattern (red), or express similar movement at the very least. This leads us to believe the design approaches in the algorithm are sound, but still need refinement to closer predict the human tapped values. Since the algorithm normalizes its prediction values to work as a MIDI input, it has one position that always has the value of 1. This is not the case with the average human tapped behavior, but was so for some individual subjects.

Possessing the subject-tapped patterns, we can feed them into the algorithm's prediction model from section 3 and compute the error in predicted position in the rhythm space. Despite the visual similarities, the model performed poorly on predicting the subject-tapped rhythm's location in rhythm space (MAE_{all}=0.4580, $SD_{subject}$=0.0766, $SD_{pattern}$=0.0353). Full results of the model's predictions are in appendix III.

Future research for tapping to polyphonic drum patterns study should include the pipeline: polyphonic drum pattern presented to get subject tapped pattern, predict location and find polyphonic drum pattern in rhythm space, listen and rate quality of prediction.

5. Discussion

Two explorations were conducted to study the relationship between monophonically tapping complex rhythms, our predictive ability of the tapped rhythm, and its predictive ability to guess the original or similar polyphonic pattern in a rhythm space. We developed flattening algorithms that output a monophonic patten representing the rhythm of a polyphonic drum pattern. We demonstrated that we can reasonably predict the location of that polyphonic pattern from the algorithm-output patterns (MAE_{best}: 0.039). We set up an experiment to study how people tapped to polyphonic patterns and observe patterns of tapping behavior. Lastly, we briefly compared the results with the flattening algorithm predictions.

By showing that we can reasonably predict a location in a polyphonic drum rhythm space with a monophonic pattern and studying how people tap their perceived rhythm of polyphonic drum patterns, we hope to demonstrate the viability of a tapped input for these latent drum spaces. Subjects were able to tap to patterns from all regions of the drum rhythm space, and progressively increase their ability to tap on the equipment as intended. Looking at tapping behavior as a whole in this context serves to add to the study of tapping behavior in complex rhythms. A tapped input is now able to now lead to a full polyphonic drum pattern output, and with customization tools and the natural navigable ability of the drum rhythm space, we are on our way to creating a real time tool for music production. However, there were a few limitations of this study. Notably, only one piece of production equipment was used in testing (Ableton Push 2). The results for the tap consistency task (figs 4.5, 4.6) could indicate tendencies of subjects to not use the full tapping range for a variety of reasons, being operational comfort or a physical aspect of the Push 2 equipment. Similar tests on other drum pads are required, and potentially customization based on individual tapping behavior. We did not account for participants experience specifically with digital music production.

It is important to note that this study and related research are only considering a subset of the wide range of metrical structures in western music. There is evidence that metrical weighting hierarchies from the GTTM do not apply to Sub-Saharan music (Toussaint, 2015). And so more research is needed and further exploration into the reproduction of complex rhythms in different metrical structures.

The results of the flattening algorithm exploration in conjunction with the subject tapped results provide hope for refining the predictive capabilities of such algorithms to properly reflect the nature of human rhythm perception. The predictions of the continuous representation of Onset Density, Syncopation Strength, and Metrical strength were visually similar to the behavior of the subject tapped patterns. Interspersing human tapped results alongside predicted results could help improve the predictive abilities, but would require testing on a larger subset of the patterns that the rhythm space is built on. It remains clear that there is a lot more nuance in the rhythmic perception of polyphonic drum patterns to be explored if we hope to predict it. For the goal of moving forward in the prediction of rhythm perception, we can confidently contribute the research above.

6. Conclusion

Over the course of this research project, we have studied the literature about rhythm perception, similarity and representation in latent spaces. We have developed algorithms for flattening polyphonic drum patterns into a single channel that represents the rhythmic structure of the original polyphonic pattern. From those flattened pattern, we created a model to predict a relevant drum pattern from a polyphonic drum rhythm space. We conducted an experiment to explore human rhythm tapping behavior and compare artificially flattened patterns with human-tapped patterns. What's left is to complete the system to be able to go from a tapped pattern and predict a relevant drum pattern from the rhythm space. Researchers and musicians alike stand to benefit from a deeper understanding of our relationship with rhythm, and its use in the real world. We hope the insights gained from this project help in

the field of music cognition and its real-world applications in computational music creativity.

Acknowledgements

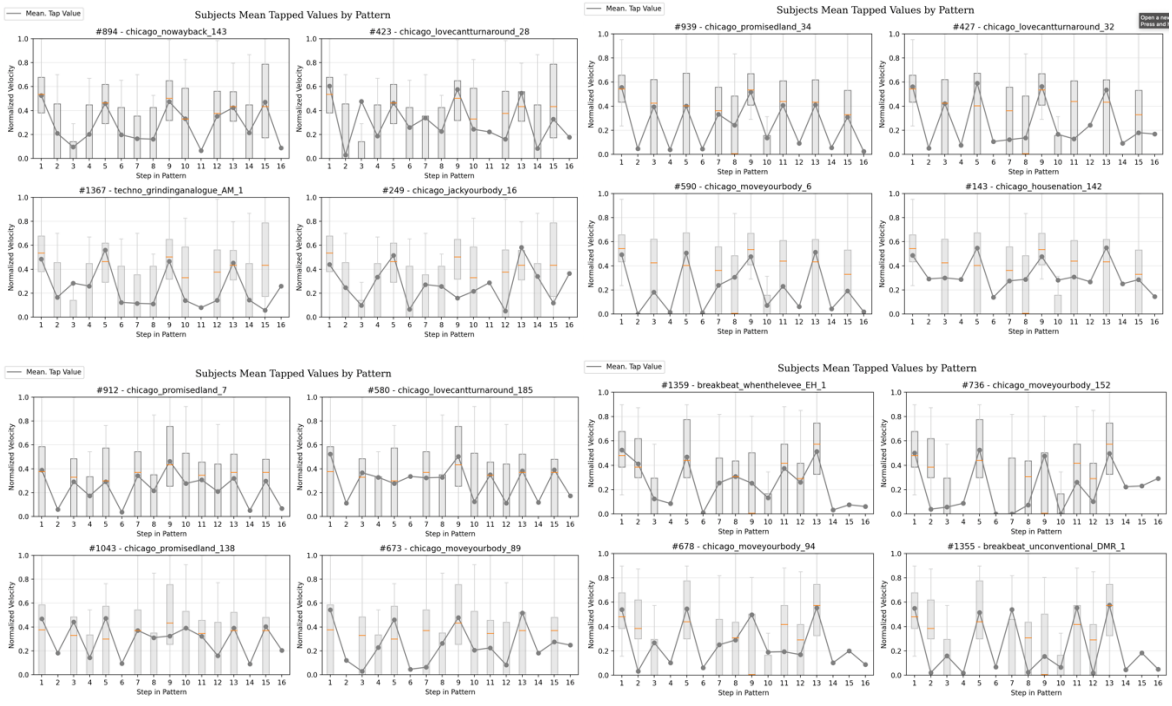
I would like to thank Daniel Gómez-Marín for his unwavering support and infectious enthusiasm about this project.

Appendix

I. Selected Pattern Names and Index Numbers

- 143: *Chicago_housenation_142*
- 249: *Chicago_jackyourbody_16*
- 423: *Chicago_lovecantturnaround_28*
- 427: *Chicago_lovecantturnaround_32*
- 580: *Chicago_lovecantturnaround_185*
- 590: *Chicago_moveyourbody_6*
- 673: *Chicago_moveyourbody_89*
- 678: *Chicago_moveyourbody_94*
- 736: *Chicago_moveyourbody_152*
- 894: *Chicago_nowayback_143*
- 912: *Chicago_promisedland_7*
- 939: *Chicago_promisedland_34*
- 1043: *Chicago_promisedland_138*
- 1355: *Breakbeat_ unconvential_DMR_1*
- 1359: *Breakbeat_whenthelevee_EH_1*
- 1367: *Techno_grindinganalogue_AM_1*

II. Mean Tapped Velocity for Test Patterns.



III. MAE for continuous density syncopation meter model predictions vs subjects / patterns.

Subject	MAE Subject	MAE Pattern	Pattern
1	0.39358323	0.48484201	894
2	0.38832215	0.51370967	423
3	0.35054897	0.41809661	1367
4	0.38542534	0.51463798	249
5	0.53228652	0.47781285	939
6	0.4148635	0.52483981	427
7	0.40765995	0.42638968	590
8	0.40525963	0.43606062	143
9	0.41100792	0.45602239	912
10	0.43575109	0.40700196	580
11	0.45904801	0.46600478	1043
12	0.54822024	0.45566988	673
13	0.3871414	0.45021026	1359
14	0.36749881	0.44382862	736
15	0.55174406	0.43217048	678
16	0.63666603	0.42062438	1355
17	0.51512241		
18	0.61406627		
19	0.50314806		
20	0.44060358		
21	0.48069082		
22	0.42901865		
23	0.4762112		

Bibliography

Abadi, Martín, Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., … others. (2016). Tensorflow: A system for large-scale machine learning. In *12th $USENIX$ Symposium on Operating Systems Design and Implementation ($OSDI$ 16)* (pp. 265–283).

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.

Benedetto, A., & Baud-Bovy, G. (2021). "Tapping Force Encodes Metrical Aspects of Rhythm". *Frontiers in Human Neuroscience*.

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

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.

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

Cameron, D., Potter, K., Wiggins, G., & Pearce, M. (2017). Perception of Rhythmic Similarity is Asymmetrical, and Is Influenced by Musical Training, Expressive Performance, and Musical Context, *Timing & Time Perception*, *5*(3-4)

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

Câmara, Guilherme Schmidt., Sioros, George., Danielsen, Anne. (2022). "Mapping timing and intensity strategies in drum-kit performance of a simple back-beat pattern". Journal of new music research, Vol.51 (1), p.3-26

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

Cooper, G., & Meyer, L. B. (1960). The rhythmic structure of music. Chicago, IL: University of Chicago Press.

Chen, J. C. C., & Chen, A. L. P. (1998). "Query by rhythm: An approach for song retrieval in music databases." *In Research issues in data engineering. Continuous-media databases and applications*. proceedings., eighth international workshop on (pp. 139–146).

Clarke, E. F. (1984). “Structure and expression in rhythmic performance.” In P. Howell, I. Cross, & R. West (Eds.), Musical structure and cognition (pp. 209–236) New York: Academic Press.

Dean RT, Bulger D, Milne AJ. (2021) On the Roles of Complexity and Symmetry in Cued Tapping of Well-formed Complex Rhythms. Music perception. 2021;39(2):202–25.

Desain P., Honing H., (2001) “Modeling the effect of meter in rhythmic categorization: Preliminary results” *Journal of Music Perception and Cognition* 7(2).

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

Dowling, W. J., and Tighe, T. J. (2014). “Psychology and music: The understanding of melody and rhythm.” Psychology Press.

Feldman, J. (2016). “The simplicity principle in perception and cognition.” *Wiley Interdisciplinary Reviews: Cognitive Science, 7*, 330–340.

Fitch WT, Rosenfeld AJ. Perception and Production of Syncopated Rhythms. Music perception. 2007;25(1):43–58.

Fraisse, P. (1982). “Rhythm and tempo.” In D. Deutsch (Ed.), *The psychology of music* (pp. 149-180). London, UK: Academic Press.

Frühauf, J., Kopiez, R., & Platz, F. (2013). Music on the timing grid: The influence of microtiming on the perceived groove quality of a simple drum pattern performance. *Musicae Scientiae*, *17*(2), 246–260.

Fujii, S., Hirashima, M., Kudo, K., Ohtsuki, T., Nakamura, Y., & Oda, S. (2011). Synchronization error of drum kit playing with a metronome at different tempi by professional drummers. *Music Perception, 28*(5), 491-503.

Gabrielsson, A. (1973). Similarity ratings and dimension analyses of auditory rhythm patterns: I & II. Scandinavian Journal of Psychology 14(3):161–176.

Gärdenfors, P. (2000). “Conceptual Spaces: The Geometry of Thought” (1st ed.). *Cambridge, Massachussetts: The MIT Press*.

Georgi, M., Gingras, B., & Zentner, M. (2023) “The Tapping-PROMS: A test for the assessment of sensorimotor rhythmic abilities.” *Auditory Cognitive Neuroscience*.

Guastavino, C., Gómez, F., Toussaint, G., Marandola, F., & Gómez, E. (2009). “Measuring Similarity between Flamenco Rhythmic Patterns.” (Vol. 38) (No. 2).

Grahn, J. A., and Brett, M. (2007). Rhythm and beat perception in motor areas of the brain. *Journal of cognitive neuroscience* 19, 893–906. doi: 10.1162/jocn.2007.19.5.893

Grahn, J. A. (2009). The role of the basal ganglia in beat perception: neuroimaging and neuropsychological investigations. *Ann. N. Y. Acad. Sci.* 1169, 35–45. doi: 10.1111/j.1749-6632.2009.04553.x

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

Gómez-Marín, D., Jordà, S., & Herrera, P. (2015a). "Pad and Sad: Two Awareness-Weighted rhythmic similarity distances." *In 16th International Society for Music Information Retrieval Conference (ISMIR).*

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

Gómez-Marín, D., Jordà, S., & Herrera, P. (2016). "Strictly Rhythm: Exploring the effects of identical regions and meter induction in rhythmic similarity perception." *In Music, Mind, and Embodiment: 11th International Symposium,* CMMR 2015, Plymouth, UK.

Gómez-Marín, D. (2018). Similarity and style in electronic dance music drum rhythms (Ph.D. Thesis, Universitat Pompeu Fabra). Retrieved 2021-03-12, from http://www.tdx.cat/handle/10803/543841 (Accepted: 2018-05- 10T10:44:33Z Publication Title: TDX (Tesis Doctorals en Xarxa))

Gómez-Marín, D., Jordà, S., & Herrera, P. (2020). "Drum Rhythm Spaces:From Polyphonic Similarity to Generative Maps. J New Music Res. 2020 Aug 24;49(5);438-56. DOI: 10.1080/09298215.2020.1806887

Halpern, A. R., Zatorre, R. J., Bouffard, M., Johnson, J. A. (2004). "Behavioral and Neural Correlates of Perceived and Imagined Musical Timbre" in *Neuropsychologia*, Vol.42 (9), p.1281-1292

Han, S., Lee, J., Yun, G., Han, S. H., Choi, S. (2022). "Motion Effects: Perceptual Space and Synthesis for Specific Perceptual Properties," in *IEEE Transactions on Haptics*, vol. 15, no. 3, pp. 626-637, 1 July-Sept. 2022, doi: 10.1109/TOH.2022.3196950.

Handel , S. (1989). "Listening: An introduction to the perception of auditory events." Cambridge, MA: MIT Press.

Handel, S. (1992). “The differentiation of rhythmic structure.” *In Perception and Psychophysics*, 52, 497-507.

Hannon, E. E., Snyder, J. S., Eerola, T., Krumhansl, C. L. (2004). “The role of melodic and temporal cues in perceiving musical meter.” Journal of Experimental Psychology: Human Perception and Performance, 30, 956-974. doi: 10.1037/0096- 1523.30.5.956

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.

Hourdin, C., Charbonneau, G., & Moussa, T. (1997). “A multidimensional scaling analysis of musical instruments’ time-varying spectra.” *Computer music journal*, 21(2), 40–55.

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.

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.

Jolliffe, Ian T., Cadima, J. (2016). "Principal component analysis: a review and recent developments". *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. **374** (2065): 20150202.*

Krumhansl, C. (1979). “The psychological representation of musical pitch in a tonal context.” *Cognitive Psychology*, 11(3), 346–374.

Krumhansl, C. (2000). “Rhythm and Pitch in Music.” In *Psychology*.

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

Large, E. W., & Kolen, J. F. (1994). “Resonance and the perception of musical meter.” Connection Science, 6, 177-208.

Large, E. W. (2008). “Resonating to musical rhythm: Theory and experiment.” In S. Grondin (Ed.), The psychology of time (pp. 189–231). Bingley, UK: Emerald.

Lattner, S., & Grachten, M. (2019). “High-level control of drum track generation using learned patterns of rhythmic interaction.” In Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA 2019).

Lee, C. S. (1991). “The perception of metrical structure: Experimental evidence and a model.” *In P. Howell, R. West, & I. Cross (Eds.), Representing musical structure* (pp. 59-127). London: Academic Press.

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

Longuet-Higgins, H. C., & Lee, C. S. (1984). “The Rhythmic Interpretation of Monophonic Music.” Music Perception: An Interdisciplinary Journal, 1(4), 424–441.

London, J. (2004). “Hearing in time.” New York: Oxford University Press.

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

Makris, D., Kaliakatsos-Papakostas, M., Karydis, I., & Kermanidis, K. L. (2017). “Combining LSTM and feed forward neural networks for conditional rhythm composition.” *In 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.

Mead, A (1992). "Review of the Development of Multidimensional Scaling Methods". *Journal of the Royal Statistical Society. Series D (The Statistician). **41** (1): 27–39. JSTOR234863*

Miller, G. A. (1956). “The magical number seven, plus or minus two: some limits on our capacity for processing information.” *Psychological Review*, 63, 81-97.

Milne, A. J., Bulger, D., & Herff, S. A. (2017). “Exploring the space of perfectly balanced rhythms and scales.” *Journal of Mathematics and Music, 11*(2–3), 101–133.

Milne, A.J., Dean, R. T. (2016). “Computational Creation and Morphing of Multilevel Rhythms by Control of Evenness*.” Computer music journal*. 2016;40(1):35–53.

Milne, A. J., Herff, S. A., Bulger, D., Sethares, W. A., & Dean, R. T. (2016). XronoMorph: Algorithmic generation of perfectly balanced and well-formed rhythms. *Proceedings of the 2016 international conference on new interfaces for musical expression (NIME 2016)* (pp. 388–393). Brisbane, Australia: Griffith University.

Milne, A.J., Herff, S.A. (2020). “The perceptual relevance of balance, evenness, and entropy in musical rhythms.” *Cognition*. 2020;203:104233–104233.

Milne, A., Dean, R., & Bulger, D. (2021). “Tapping to unfamiliar and highly syncopated rhythms: Modelling behavior and cognitive mechanisms.” 10.31234/osf.io/qaek6.

Moritz, M., Heard, M., Kim, H.-W., & Lee, Y. S. (2021). “Invariance of edit-distance to tempo in rhythm similarity.” *Psychology of Music*, *49*(6), 1671–1685.

Nistal J., Herrera P., Jordà S. “Exploring the spontaneous expression of human finger-tapping.” Paper presented at: 2nd Conference on Computer Simulation of Musical Creativity; 2017 Sep 11-13. Milton Keynes, UK.

Ó Nuanáin, C., Herrera, P., & Jorda, S. 2015. “Target-based rhythmic pattern generation and variation with genetic algorithms.” *In Sound and Music Computing Conference.*

Palmer, C., & Krumhansl, C. L. (1990). “Mental representations for musical meter.” *Journal of experimental psychology. Human perception and performance,* 16(4), 728–741.

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

Pols, L. C., van der Kamp, L. J., Plomp, R. (1969) “Perceptual and physical space of vowel sounds”. in *J Acoust Soc Am* 46:458–467, doi:10.1121/1.1911711, pmid:5804118.

Post, O., & Toussaint, G. (2011). The Edit Distance as a Measure of Perceived Rhythmic Similarity. *Empirical Musicology Review,* 6 (3), 164–179. Retrieved from https://kb.osu.edu/dspace/handle/1811/52811

Povel, D-J. (1984). “A theoretical framework for rhythm perception.” Psychological Research, 45, 315-337.

Povel, D-J., & Essens, P. (1985). “Perception of temporal patterns.” Music Perception, 2, 411-440.

Puckette, M. “PureData (Pd): real-time music and multimedia environment”. Accessed from: http://msp.ucsd.edu/software.html

Ragni, M., Khemlani, S., & Johnson-Laird, P. N. (2014). “The evaluation of the consistency of quantified assertions.” *Memory and Cognition*, 42, 53-66.

“Rhythm”. In *Merriam-Webster.com*. Retrieved Aug 14, 2023.

Roberts, A., Engel, J., Raffel, C., Hawthorne, C., & Eck, D. (2018). A hierarchical latent vector model for learning long-term structure in music. In Proceedings of the 35th International Conference on Machine Learning.

Roberts, A., Engel, J., Mann, Y., Gillick, J., Kayacik, C., Nørly, S., ... Eck, D. (2019). Magenta Studio: Augmenting creativity with deep learning in Ableton Live. In Proceedings of the International Workshop on Musical Metacreation (MUME).

Rohrmeier, M., & Rebuschat, P. (2012). "Implicit learning and acquisition of music." *Topics in Cognitive Science, 4*, 525–553.

Ross, J., & Houtsma, A. J. M. (1994). "Discrimination of auditory temporal patterns." *In Perception and Psychophysics,* 56, 19-26.

Russ, Martin. (2019). "Background." *Sound Synthesis and Sampling*, Routledge, London. ch1-p.83.

Shaffer, L. H., Clarke, E. F., & Todd, N. P. (1985). Metre and rhythm in piano playing. *Cognition*, 20, 61-77.

Shepard, R. N. (1964). "Attention and the metric structure of the stimulus space." *Journal of Mathematical Psychology*, 1(1), 54–87.

Shepard, R. (1999). Cognitive psychology and music. *In P. Cook (Ed.), Music, cognition, and computerized sound* (pp. 21-35). Cambridge, MA: MIT Press.

Shneiderman, B. (2002). Creativity support tools: a tutorial overview. In Proceedings of the 4th conference on Creativity & cognition (C&C '02). Association for Computing Machinery, New York, NY, USA, 1–2. https://doi-org.sare.upf.edu/10.1145/581710.581711

Stein, M. 1953. "Creativity and culture." The journal of psychology, 36(2), 311-322.

Tokui, N. (2020). Towards democratizing music production with AI-design of variational autoencoder-based rhythm generator as a DAW plugin. arXiv preprint arXiv:2004.01525

Toussaint, G. T. (2004). A Comparison of Rhythmic Similarity Measures. *In Proceedings of the 7th international society for music information retrieval conference* (ISMIR 2004)(pp. 242–245).

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

Toussaint, G. T. (2015). Quantifying Musical Meter: How Similar are African and Western Rhythm? *Analytical approaches to world music*. 2015;4(2).

Turk-Browne, N. B., & Scholl, B. J. (2009).” Flexible visual statistical learning: Transfer across space and time.” *Journal of Experimental Psychology: Human Perception and Performance, 35*(1), 195–202.

Turquois, C., Hermant, M., Gómez-Marín, D., & Jordà, S. (2016). “Exploring the Benefits of 2D Visualizations for Drum Samples Retrieval.” In Proceedings of the 2016 ACM on conference on human information interaction and retrieval (pp. 329–332).

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: An Interdisciplinary Journal, 32(2), 201 – 217.

Velautham L, Yoong RCS. (2022). “Can’t clap to a beat? How rhythmically challenged people experience and strategize keeping time to music.” *Psychology of music*. 2022;50(4):1254–66.

Vigliensoni, G. & McCallum, L. & Maestre, E. & Fiebrink, R., (2022) “R-VAE: Live latent space drum rhythm generation from minimal-size datasets”, *Journal of Creative Music Systems* 1(1). doi: https://doi.org/10.5920/jcms.902

Volk, A. (2008). “The study of syncopation using inner metric analysis: Linking theoretical and experimental analysis of meter in music.” Journal of New Music Research, 37, 259-273.

Zaidi, Q., Victor, J., McDermott, J. Geffen, M., Bensmaia, S., T. A. Cleland. (2013). “Perceptual Spaces: Mathematical Structures to Neural Mechanisms,” in *Journal of Neuroscience*. 6 November 2013, 33 (45) 17597-17602; DOI: 10.1523/JNEUROSCI.3343-13.2013