SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Executive Summary of the Thesis

A Study on the Effect of Dynamic Accents on the Perception and Measurement of Rhythm Complexity

Laurea Magistrale in Music & Acoustic Engineering

Author: Gabriele Maucione

Advisor: Prof. Massimiliano Zanoni

Co-advisors: Alessandro Ilic Mezza, Dr. Luca Comanucci

Academic year: 2022-2023

---

1. Introduction

Rhythm refers to the arrangement of sounds in time and is among the most influential musical attributes for what concerns the aesthetic, emotional, and behavioral response of the listener. Rhythm not only dictates when notes are played but also determines their duration and intensity. Furthermore, in a music performance, other more subtle factors can affect the perception of rhythm, such as timbre, envelopes, and the register of an instrument.

Rhythm complexity is considered a significant semantic descriptor of music content, influencing neural activities related to attention, reactivity, and excitement. Thus, it is a topic of interest for Music Information Retrieval (MIR), as it may be instrumental for recommender systems, music genre classification, and database navigation, to name a few.

In the past few decades, several studies aimed at finding a quantitative model of this markedly subjective attribute of music. However, the existing corpus of literature only consider onsets in a binary fashion, i.e., either present or absent, and disregard their intensity altogether. As such, existing works do not take into account the role of dynamics.

To the best of our knowledge, the present work is the first to investigate this aspect. In particular, we explore the effect of dynamic accents on the perception of rhythm complexity, considering both the monophonic and polyphonic case. Such an analysis is conducted by means of two subjective listening tests, where rhythms with different dynamics are compared. Additionally, we discuss the robustness of existing quantitative methods of measuring rhythm complexity when dynamics is present. Namely, we evaluate the performance of six well-established complexity measures originally proposed for binary rhythmic patterns.

Our analysis show that varying the onset intensity affects the subjective perception of complexity, and that existing measures prove reliable in capturing the complexity of rhythms with dynamic accents, especially in the case of polyphonic patterns.

The remainder of the manuscript is organized as follows. In Section 2, we present six rhythm complexity measures from the literature. Section 3 is dedicated to the illustration of the design principles for the subjective listening tests. In Section 4, we analyze the experimental results. Finally, Section 5 concludes this work.

2. Rhythm Complexity

In this section, we present the complexity measures considered in our study. In the literature, different families of rhythm complexity measures exist. In practice, however, all measures process rhythms expressed as binary sequences, where a 1 indicates an onset and a 0 a rest, and yield a scalar representing the level of complexity that a human listener would perceive. In the following, we introduce six well-established measures [5].
Belonging to the *rhythmic syncopation* family, **Toussaint's Metrical Complexity** and **Longuet-Higgins & Lee (LHL) Complexity** rely on a metrical hierarchy where each pulse is associated with a predetermined weight in order to find syncopation pulses. **Pressing's Cognitive Complexity** belongs to the *pattern matching* family, which is based on assigning sub-pattern archetypes a certain degree of complexity. **Weighted Note to Beat Distance (WNBD)** relies on measuring the distance between a pattern and a reference rhythm. **Off-Beatness** belongs to the *mathematical irregularity* family, which is characterized by a geometric interpretation of rhythm to find their irregularities. **Inter-Onset Intervals Information Entropy** belongs to the family of the *inter-onset intervals* (IOI) measures and computes the uncertainty of the period of time between the onsets. Finally, the **Grouped-Voice Polyphonic Measure** by Mezza et al. [4] is the only example of complexity measure that can deal with polyphonic patterns.

3. Subjective Listening Test

This section illustrates the design of two listening tests aimed at assessing the effect of dynamic accents on the perception of rhythm complexity, as well as the modeling capabilities of the metrics discussed above when dealing with non-constant intensity patterns. Specifically, the first test pertains to monophonic patterns, whereas the second concerns polyphonic ones. Participants were asked to rate on a scale of 1 to 5 the rhythm complexity they perceived while listening to a number of audio stimuli. Previous studies [2] adopted an alternative approach by instructing participants to reproduce a given rhythm, or to tap its pulse. However, this methodology aligns with a more specific interpretation of rhythm complexity, as the challenges associated with a rhythm's reproduction may not necessarily mirror those related to its conceptualization. Moreover, tackling the reproduction of a polyphonic pattern assumes a prerequisite familiarity with drumming techniques, thereby introducing a difference in the psychological problem addressed by the test, with respect to the one of monophonic rhythms. Conversely, asking participants to quantify the perceived complexity ensures the results from both the tests to be comparable.

3.1. Monophonic Patterns

The first experiment concerns monophonic patterns. While the conclusions drawn by analysing monophonic patterns might not generalize well for what concerns complex musical composition, this approach aligns with much of the existing literature [5].
We use the Fitch & Rosenfeld's dataset [2], a collection of 30 hand-crafted monophonic binary patterns in 4/4 time signature, each lasting one bar with a 16-pulse resolution and containing four to five onsets.
To investigate the impact of dynamic accents on complexity perception, we synthesize four different versions of each pattern, resulting in a total of 120 stimuli. Each version is obtained by assigning different velocity[1] values to the onsets in each MIDI pattern: *Constant* (no intensity variation), *Hierarchy* (velocity is assigned according to the Lerdahl & Jackendoff's pulse hierarchy [3]), *Random* (random velocity values), and *Performed* (velocity from a human performance). The stimuli were synthesized as 8-second audio files, including two pre-count metronome bars to provide a reference meter to the listener, followed by two measures of the target pattern. Unlike previous studies that used the click sound of a metronome [1] for the target pattern, we favor a realistic drum snare sound to capture the diverse range of effects that intensity has on musical performance, encompassing loudness, timbre, and pitch variations.
In order to limit the duration of the listening test, the 120 stimuli were not all included in the same test session; instead, we split them in three groups. In order to do so, we randomly parti-

[1] Velocity is a MIDI parameter directly associated with the intensity of a note, and often mapped onto other attributes such as timbre.

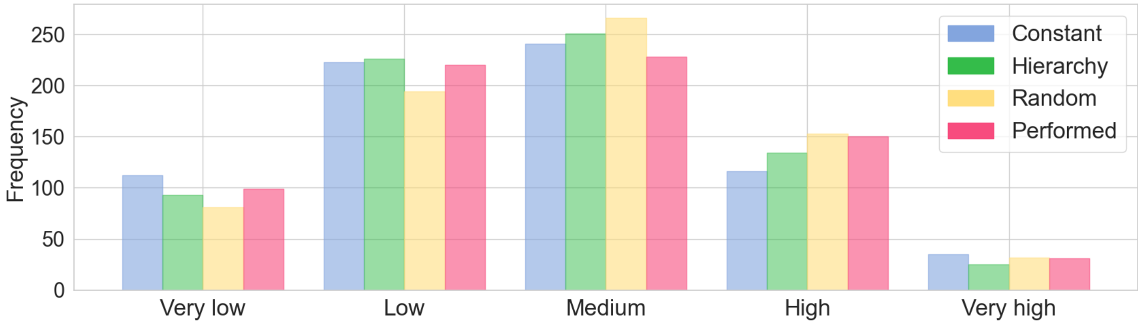

Figure 1: Scores distribution of the first listening test.

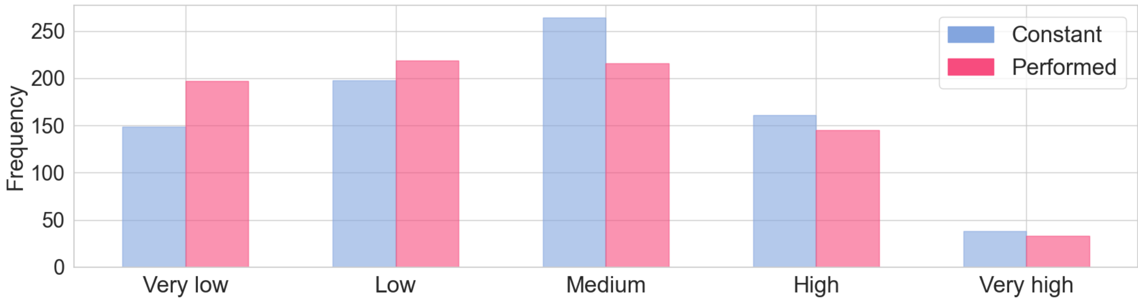

Figure 2: Scores distribution of the second listening test.

tioned the original dataset in three sets of ten patterns each, and included all four versions of the patterns. Thus, each test session presented the user with a likert-scale questionnaire on 40 rhythms. For the design of the test, we relied on the webMUSHRA software developed by AudioLabs.

3.2. Polyphonic Patterns

The second experiment extends the analysis to polyphonic patterns. In doing so, we focus on drums, to exclude the interplay with the harmonic component that other musical instruments would entail. Given the nature of drums and their interplay, it becomes crucial to extend research from monophonic to polyphonic patterns, to mark a progression in the application of rhythm complexity measures to real music.
The structure of the test is the same of the first one, with the only variation of a different set of stimuli. Groove MIDI Dataset (GMD) was adopted for this purpose. We selected ten single-measure rhythmic patterns from GMD by computing the following steps. We split all the MIDI files in measures and kept only the 4/4 time signature patterns; we discarded measures with less than three voices or less than eight onsets; we computed the expected complexity of each of the resulting patterns according to Grouped-Voice Polyphonic Toussaint's measure; we partitioned the rhythms in ten groups of increasing complexity; from each group, we select a random pattern among the ones whose velocity distribution had the highest variance. We also applied the same reduction mapping applied by the authors of [4], reducing the maximum number of voices in each MIDI file to nine.
For each pattern, two audio clips were synthesized: *Performed* (original velocity, recorded from human performance) and *Constant* (velocity set to 100). In this way, the test compares rhythmic patterns with and without expressiveness. We repeated each rhythm four times, and then synthesized audio files using Groove Agent by Steinberg.

4. Results

4.1. Monophonic Patterns

The first listening test was completed by 72 participants, which are all either experienced musicians, students from the Music & Acoustic Engineering MSc at Politecnico di Milano, or researchers in the field of MIR.
The analysis of the subjective scores reveals that the proposed rhythms were generally per-

	Toussaint	LHL	Pressing	WNBD	IOI Information Entropy	Off-Beatness
Constant	0.569570	0.584780	0.703710	0.664680	0.551650	0.552740
Hierarchy	0.442090	0.485310	0.615350	0.560720	0.543100	0.445330
Random	0.668360	0.707180	0.747770	0.771720	0.480080	0.611540
Performed	0.679810	0.700700	0.759780	0.732970	0.478920	0.664490

Table 1: Pearson correlation coefficients between mean test scores and metrics' scores in each velocity mode.

ceived as not highly complex (Fig. 1). Dynamics showed to affect listeners' perception, but in a different way for each compared intensity mode. Excluding "Very high" ratings, a comparison of velocity modes indicates that the *Random* (R) mode is generally perceived as more complex, with more votes in the "High" and "Medium" categories. Conversely, *Constant* (C) mode tends to have fewer scores in the high ratings. *Performed* (P) and *Hierarchy* (H) are placed midway, with a higher complexity perceived for P mode. Thus, when a single instrument is involved in the pattern, results showed that a variation in the intensity generally imply an increment in the perceived complexity.

Table 1 shows the Pearson correlation coefficients between the six measures described in Section 2 and the average subjective scores for each velocity mode.

Overall, WNBD and Pressing generally exhibit high correlation scores. Table 1, however, reveals not all the coefficients are equally high. In this regard, it is worth remembering that the experiment focuses on monophonic snare patterns, which may be unfamiliar to listeners. This could impact their perception and judgment.

A common trend for all the rhythm complexity measures - except IOI Information Entropy - is that they achieve the highest correlation in R mode, the second highest in P, the second lowest in C, and the lowest in H.

Further examination considers shifts in perceived complexity among velocity modes. H mode stands out with a divisive behavior, as the users equally split among the ones who found it more complex than C mode and the ones that perceived the opposite. Potentially, this diversity in the listener perceptions made it challenging for the metrics to capture its characteristics accurately. R and P modes showed to be perceived with less ambiguity, possibly due to their resemblance to the real-life dynamics of a human performance.

The results highlight the importance of considering listener perspectives and the inherent challenges of assessing metrics in the context of monophonic patterns. Actually, the absence of intensity in C mode does not necessarily diminish metric performance; instead, the listener expectations based on a majority of stimuli containing intensity variations determined a shift in their perception.

Notably, the IOI Information Entropy measure differs from others. Indeed, it achieves the highest correlation with C mode, and the lowest with P mode. A possible explanation is that this measure has a unique approach with respect to the other metrics, focusing on onset distance rather than absolute position, and thus challenging the common association of syncopation with dynamics.

4.2. Polyphonic patterns

In this section, we discuss the results of the second subjective listening test. A total of 82 participants completed the test.

Fig. 2 depicts the distribution of 1640 collected ratings. It can be observed that the test participants perceived P mode as generally less complex than C mode. Indeed, P collected a higher number of votes in the "Medium" and "Low" classes, while C mode showed a tendency to receive higher scores.

The perceived complexity difference between P and C modes can be attributed to the synthesis process. Indeed, the original patterns included ghost notes, also known as muted notes, that are distinguished for being played with a very soft dynamic between the main notes, but are fundamental to add depth to the groove of a drum performance. The C mode velocity profile transformed completely the role of ghost notes, making them main events and resulting in denser patterns, which have been reportedly perceived

	Toussaint	LHL	Pressing	WNBD	IOI Information Entropy	Off-Beatness
Constant	0.836070	0.600332	0.812444	0.722569	0.679872	0.592801
Performed	0.967446	0.828665	0.923137	0.811797	0.916034	0.547831

Table 2: Pearson correlation coefficients between mean test scores and polyphonic metrics in each velocity mode.

as more complex.
Similarly as it was done for the first test, we used the collected responses from the listening test as a reference to evaluate the presented measures. Since these measures only refer to monophonic patterns, we adopted the Grouped-Voice algorithm from [4] to derive the polyphonic versions of each of them. For each of the so obtained polyphonic measures, we computed the Pearson correlation coefficients with the average subjective scores assigned by the users, which are presented in Table 2.
Generally, these coefficients are higher than those in Table 1, referring to the monophonic case. This suggests that the measures are reliable in the analysis of realistic drum patterns, especially in the context of polyphonic rhythms. The increase in correlations may be attributed to the familiarity of the analyzed objects to listeners, resulting in more consistent judgments with respect to the first test.
Toussaint's metrical complexity measure consistently exhibited the highest correlation coefficient, reaching an absolute highest score of just above 0.96. This strong linear relationship between Toussaint's and the test scores suggests its effectiveness in capturing perceived complexity, as also showed by the high coefficient of determination ($R^2 = 0.93$) of the linear regression model shown in Fig. 3.
Pressing also achieved high correlation scores, consistently ranking as the second-best option.
WNBD and LHL obtained quite high correlation coefficients, although not as high as the ones of Toussaint and Pressing.
Also, IOI Information Entropy showed higher coefficients than the ones obtained in the first test, exhibiting on of the highest correlations in the P mode; still, the correlation with the average users' scores in the C mode was pretty low in comparison.
However, the lowest correlation score was obtained by Off-Beatness in relation to the P mode (reaching just above 0.54), indicating its lesser

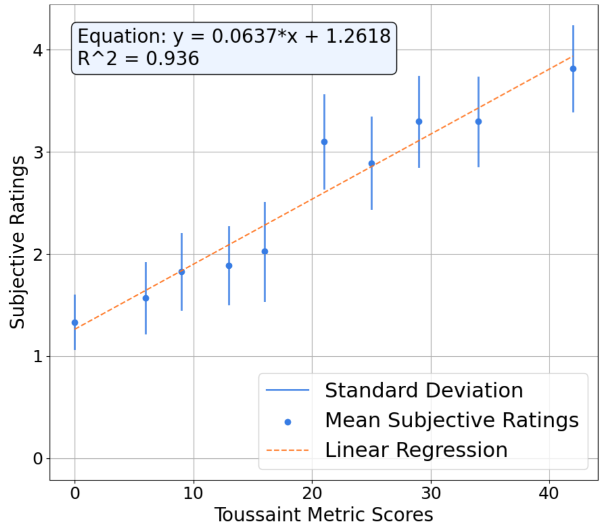

Figure 3: Linear regression model between Grouped-Voice Toussaint's Metrical Complexity and average test scores.

effectiveness in dealing with polyphonic patterns with dynamic accents.
Analyzing the C and P intensity modes in Table 2, it becomes apparent that the correlation scores of the P mode are higher than those of the C mode. Indeed, except Off-Beatness - which also obtained the lowest correlations in both modes and thus is excluded from the following considerations - all the metrics showed to respect this trend. Unexpectedly, this suggested that the quality of rhythm complexity measurement was higher on rhythms with realistic intensity profiles, as the ones played by actual musicians. Existing metrics are thus seemingly more reliable when applied to rhythmic patterns with intensity variations, even though they process the patterns as binary sequences.
Low correlations in Table 2 may be justified by high variance in users' responses. At this regard, we conducted a further analysis to verify the presence of distinct listeners populations, considering their ratings in relation to specific patterns and examining how the intensity mode influenced them.

The users revealed to quite consistently perceive the P version of each rhythm as less complex than its counterpart from C class.
However, the analysis of pattern-wise scores variance in the C mode revealed mixed opinions among users, especially for rhythms at the extremes of the proposed complexity scale. In particular, in the ratings distribution of these patterns, two distinct populations are observable. This contrasts the single-valued nature of scalar metrics, which cannot capture the divisive nature of these rhythms, thus negatively affecting the correlations. Instead, the ratings assigned to the rhythms of P mode showed less sparse distributions and lower standard deviations, indicating less indecision and more consistent judgements.
Thus, considering users' perspective proves fundamental when interpreting correlation coefficients. The unfamiliarity associated with constant-intensity rhythms led to varied opinions and compromised internal judgment scales, resulting in inconsistent evaluations. The comparison allowed by Table 2 aligns with the notion that users' internal representation of rhythm inherently includes intensity information, making metrics more effective for rhythms with intensity variations.

5. Conclusions

In this study, we investigated the novel topic of the influence of dynamic accents on the perceived complexity of rhythms through two experiments involving subjective listening tests, with a significant participant involvement. The first experiment focused on monophonic rhythmic patterns, while the second examined polyphonic rhythms. Both experiments compared variations of rhythmic patterns with different dynamics. We used the collected data from these tests to gain insights into human rhythmic perception and its relationship with rhythmic accents. Additionally, we evaluated six existing rhythm complexity measures using the test responses as a reference.
The results showed generally high correlations between the complexity measures and the test results, especially in the polyphonic case, indicating the reliability of these measures for real music performances and applications in Music Information Retrieval.
The comparison of the two experiments highlight that a too general interpretation on the cognitive phenomena of rhythm complexity is hard to find. Indeed, in the monophonic test rhythms with constant intensity were perceived as less complex, while the opposite happened for the polyphonic rhythms. These differences stress the importance of context, as well as influential factors of memory and familiarity, in the context of perception.
The study also extends prior work by applying the Grouped-Voice algorithm to develop polyphonic versions of six rhythm complexity measures originally designed for monophonic patterns, and confirming its validity.
While the analysis focused on dynamic accents, there is room for more in-depth exploration in future research. Comparisons of cases with different intensity levels for the same accents or exploring variations in introducing intensity at different metrical positions could provide further insights. Additionally, exploring aspects like polyrhythms, polymeters, and triplet rhythms could complement the study of rhythm complexity, which, thus far, has only been studied in the context of binary rhythms.

References

[1] Fleur L. Bouwer, John Ashley Burgoyne, Daan Odijk, Henkjan Honing, and Jessica A. Grahn. What makes a rhythm complex? the influence of musical training and accent type on beat perception. *PLoS ONE*, 13, 2018.

[2] W. Tecumseh Fitch and Andrew J. Rosenfeld. Perception and production of syncopated rhythms. *Music Perception: An Interdisciplinary Journal*, 25:43–58, 2007.

[3] Fred Lerdahl and Ray Jackendoff. *A generative theory of tonal music.* The MIT Press, Cambridge. MA, 1983.

[4] Alessandro Ilic Mezza, Massimiliano Zanoni, and Augusto Sarti. A latent rhythm complexity model for attribute-controlled drum pattern generation. *EURASIP Journal on Audio, Speech, and Music Processing*, 2023.

[5] Eric Thul. Measuring the complexity of musical rhythm. 2008.