İSTANBUL UNIVERSITY PRESS

Konservatoryum - Conservatorium
CONS 2024, 11 (2), 478–487
DOI: 10.26650/CONS2024-1562373

Research Article / Araştırma Makalesi

A Comparative Study of Rhythmic Complexity Measures in Music Performance

Müzik Performansında Ritmik Karmaşıklık Ölçümleri Üzerine Karşılaştırmalı Bir Çalışma

Cihan YAYGIN[1], Ozan BAYSAL[2], Yavuz BURUK[3], Barış BOZKURT[4]

[1]İstanbul Technical University, Graduate School, Musicology and Music Theory Doctor of Philosophy Program, İstanbul, Türkiye
[2]İstanbul Technical University, Turkish Music State Conservatory, Department of Musicology, İstanbul, Türkiye
[3]İstanbul Technical University, Graduate School, Music Doctor of Philosophy Program, İstanbul, Türkiye
[4]College of Interdisciplinary Studies, Zayed University, Dubai

Corresponding author/ Sorumlu yazar : Cihan YAYGIN
E-mail / E-posta : chnyygn@gmail.com

ABSTRACT

The study of rhythmic complexity aims to determine the perceptual complexity of rhythms and how easily they can be remembered and performed. This research report evaluates two preliminary studies that approached rhythmic complexity from two complementary perspectives: conceptual understanding and performance. The research focuses on five rhythmic complexity measures: Lempel-Ziv, Keith, Tanguiane, Weighted Note-to-Beat Distance (WNBD), and Pressing. The rhythms used in the studies were created by the algorithmic software designed based on the conservatory entrance exam questions. In the first study, three expert raters evaluated the suitability and difficulty of 80 patterns. Their scores were compared with the values obtained from the complexity scales. The second study involved 11 conservatory students (5 female, 6 male) performing 16 selected rhythms from the initial pool.

The results demonstrated varying degrees of alignment between different scales and expert predictions under different conditions. In first study, Pressing, WNBD, Keith and Tanguiane complexity scales were found to be consistent with expert raters. In the second study, the expert raters provided the most accurate predictions for rhythm repetition difficulty, followed by the WNBD scale and the Tanguiane scale. In addition, there was a significant negative correlation between the average time spent and performance scores.

Keywords: Rhythm cognition, rhythm performance, rhythmic complexity scales, rhythmic skill assessment

ÖZ

Ritmik karmaşıklık çalışmaları, ritimlerin algısal karmaşıklığını, ne kadar kolay hatırlanıp icra edilebileceğini belirlemeyi amaçlayan bir araştırma alanıdır. Bu makalede, ritmik karmaşıklığa iki tamamlayıcı perspektiften yaklaşan (kavramsal algı ve ritmik kalıpların icrası.) iki ön çalışmanın sonuçlarını raporlanmıştır. Çalışmada beş ölçeğe odaklanılmıştır: Lempel-Ziv, Keith, Tanguiane, Weight Note-to-Beat Distance (WNBD) ve Pressing. Çalışmalarda kullanılan ritimler, Konservatuvar giriş sınavı soruları temel alınarak tasarlanan algoritmik yazılım tarafından oluşturulmuştur. İlk çalışmada, üç uzman değerlendirici, algoritmik olarak oluşturulan 80 ritim kalıbının uygunluğunu ve zorluğunu değerlendirmiştir. Verdikleri puanlar, karmaşıklık ölçeklerinden elde edilen değerlerle karşılaştırılmıştır. İkinci çalışmada 11 konservatuar öğrencisi (5 kadın, 6 erkek) ilk havuzdan seçilen 16 ritmi icra etmiştir.

Sonuçlar, farklı koşullar altında farklı ölçekler ve uzman tahminleri arasında değişen derecelerde uyum olduğunu göstermiştir. İlk çalışmada, Pressing, WNBD, Keith ve Tanguiane karmaşıklık ölçeklerinin uzman puanlayıcılarla tutarlı olduğu bulunmuştur. İkinci çalışmada, ritim tekrarı zorluğu için en doğru tahminleri uzman puanlayıcılar yapmış, bunu WNBD ölçeği ve Tanguiane ölçeği izlemiştir. Ayrıca, her bir soru için harcanan ortalama süre ile performans puanları arasında anlamlı bir negatif korelasyon bulunmuştur.

Anahtar Kelimeler: Ritmik algı, ritim performansı, ritmik karmaşıklık ölçekleri, ritmik beceri değerlendirmesi

Submitted / Başvuru : 06.10.2024
Revision Requested / Revizyon Talebi : 13.11.2024
Last Revision Received / Son Revizyon : 27.11.2024
Accepted / Kabul : 29.11.2024
Online Yayın / Pulished Online : 03.12.2024

INTRODUCTION

Rhythmic complexity, a central theme in the domains of musicology, psychology, and information technologies, represents a focal point at their intersection. Known variously as *complexity of musical rhythm, rhythm complexity, temporal pattern complexity, and complexity of rhythm*, the field has witnessed the development of several tools aimed at enhancing the understanding of rhythmic complexity (de Fleurian et al., 2017; Shmulevich & Povel, 2000; Thul, 2008; Thul & Toussaint, 2008). In the psychology literature, researchers have not only delved into rhythmic structures but have also probed temporal perception and its cognitive underpinnings (Bolton, 1894; Buffardi, 1971; Clarke, 1987; Fraisse, 1956, 1982; Povel, 1981, 1984; Povel & Essens, 1985; Sternberg et al., 1982; Thomas & Brown, 1974). Researchers such as Povel have delineated the fundamental tenets and limitations of rhythmic complexity, while contributions from music theory and mathematics have illuminated the computational aspects of rhythm perception. Various scales have been proposed to measure the perceptual complexity of rhythm, allowing for the quantification of the inherent complexity within rhythmic structures. To date, the literature has presented 10 such scales (Arom & Ligeti, 1991; Essens, 1995; Gómez et al., 2005; Keith, 1991; Lempel & Ziv, 1976; Longuet-Higgins & Lee, 1984; Povel & Essens, 1985; Pressing, 1999; Shmulevich & Povel, 2000; Smith & Honing, 2006; Tanguiane, 1993, 1994; Toussaint, 2002, 2003, 2004, 2005).[1]

This article reports two complementary studies comparing the predictive accuracy of different rhythmic complexity scales. An array of algorithmically generated rhythm patterns suitable for assessing rhythm skills was assessed according to five such scales (outlined below): Lempel-Ziv, Tanguaine, Keith, Weight Note-to-Beat Distance (WNBD), and Pressing. Utilizing Thul's classification, we primarily focused on the pattern-matching scales of Pressing (1999), Keith (1991), and Tanguiane (1993, 1994).[2] Each scale takes into account different factors underlying the perception of rhythm, emphasizing their relative importance. The effectiveness of these scales was assessed by asking experts to evaluate rhythm questions and student performers to reproduce the rhythms they were played.

Rhythmic complexity scales

The Lempel-Ziv scale

The Lempel-Ziv scale (1976) uses a compression algorithm that aims to create a system capable of evaluating the randomness and complexity of finite sequences using digital data. The application of the Lempel-Ziv scale, which is also used in areas such as heart arrhythmia, in musical rhythmic structures was introduced in Povel and Schumulevich's article (Shmulevich & Povel, 2000).

The method is illustrated in Figure 1. The smallest rhythmic/temporal unit in each pattern, such as a 16th note, is represented by 1 (beat) or 0 (rest). Each new sequence of digits is called a sub-sequence for example: '1' '00' '101' etc. Each new sub-sequence encountered in a rhythmic sequence, notated from left to right, is counted, and the rhythmic pattern is scored by evaluating consecutive connections. Povel and Schumelevich suggest that it would be more efficient to write the rhythmic line twice, as the musical rhythmic patterns would be too short for the Lempel-Ziv scale. This process continues until either a new sub-sequence cannot be found or the end of the loop, where the pattern occurs twice, is reached (Shmulevich & Povel, 2000).[3] The total number of sub-sequences obtained reflects the complexity of the rhythmic sequence. In Figure 1-d, first sub-sequence (which is a unit) is "1" then second is "11" third is "10" . . . last sub-sequence will be "1101" and the total number of sub-sequences is 16 so the complexity score is 16. It should be noted that the triplet is considered the smallest unit for this rhythmic phrase.

Keith scale

In the Keith scale (1991) syncopated rhythms are classified into three categories: Hesitation, in which a note starts on the quarter note boundary but ends on the eighth note boundary, as in starting on the beat and ending off the beat (scored 1); *Anticipation*, in which the second event begins before the beat and ends on the beat (scored 2); and *Syncopation* (scored 3), in which hesitation and anticipation are combined, as shown in Figure 2. The scores are added together and the higher the score the more complex the rhythmic pattern. The Keith scale can be used only to classify rhythms consisting of units divisible by two.[4]

[1] A comprehensive comparison of these scales is presented in Thul's thesis (Thul, 2008), where basic evaluations are employed through a phylogenetic tree method to elucidate the interrelationships among these scales. As observed by Thul, the most successful scales are those that incorporate human-based measures of metrical complexity testing.

[2] Other scales in the field were excluded from the study as some of their assumptions were not clearly stated. For instance, the method proposed by Schumulevich and Povel (2000), extensively discussed in Thul's thesis (Thul, 2008), lacks clarity regarding the parameter "d," which they claim to impact rhythm repetition. Despite Thul's use of rhythms from Schumulevich and Povel's work, this ambiguity remained unaddressed, leading us to exclude this scale.

[3] Uncertainties exist regarding the implementation of the method discussed in Shumulevich and Povel's article. The method described as follows: if a subsequence ends at the same

a) 

b) **111 100 100 101 100 101 101 101**

c) **111 100 100 101 100 101 101 101 111 100 100 101 100 101 101 101**

d) **1-11-10-0-100-101-1001-01-1011-011-111-00-10010-1100-10110-1101**

Figure 1. a) Rhythmic sequence in staff notation; b) Rhythmic sequence notated digitally in the binary system (The smallest unit into which 1 beat is divided. ,here is 1/3 for triplet); c) Final sequence in which the same pattern occurs twice; d) Each new subsection is divided by "-".

Figure 2. (a), (b): No syncopation; (c) hesitation; (d) anticipation; (e) syncopation. The dots represent the underlying metrical "beats" (Keith, 1991, p. 134).

Keith's work has a background that expresses the divisions of a beat in fractions. This approach was also the source of the Tanguaine and WNBD scales. However, Keith points out that the rhythmic expression he uses in his work can be ambiguous: "*Deciding on the relative rhythmical "strength" of a hesitation, anticipation, and syncopation is an interesting philosophical problem. Most listeners would probably agree that an anticipation is stronger than a hesitation, and a syncopation is strongest of all. For simplicity, therefore, we define the syncopation value (s) of these three types of events as follows: hesitation = 1; anticipation = 2; syncopation =3 (= anticipation + hesitation)*" (Keith, 1991, p. 134). Despite its limitations, the Keith scale is significant as it provides a foundational approach for other rhythmic scales. It was also used by Povel and Schumulevich (2000). In our study, it was tested on human subjects for the first time in comparison with the WNBD and Tanguiane scales derived from it.

Tanguiane scale

Tanguiane's method divides rhythmic patterns into single notes and sub-units (Tanguiane, 1993), which can then be elaborated by subdividing the durations (Shmulevich & Povel, 2000). For example, a rhythmic pattern might be broken down into sub-units like 1000, 1111, 1010, etc., where each number represents a specific rhythmic duration. These sub-units are then analyzed to determine the complexity of the rhythm, with more sub-units indicating higher complexity. The complexity score of a rhythm is based on the number of distinct sub-units it contains, thus, a rhythm with sub-units such as 111, 100, and 101 would have a complexity score of 3. This method aligns with Gestalt psychology and builds on Lerdahl & Jackendoff's (1996) conceptual framework of rhythmic structures.

The Weighted Note-to-Beat Distance (WNBD)

The Weighted Note-to-Beat Distance (WNBD) scale measures the distance between the first beat (onset) and subsequent beats, establishing a criterion based on the tempo ratio and the distance from strong beats (Gómez et al.,

position as a predetermined subsequence during the second repetition of the rhythm, it is assumed that no new subsequences are generated in the remaining part of the rhythm. For the present paper, we shall abstain from addressing these matters and proceed by presenting the data from the source without attempting to rectify these errors

4 Although it may seem possible to adapt scales to measure rhythmic structures that do not follow divisions by two, this approach is susceptible to perceptual and functional misconceptions. It is crucial to consider the perceptual differences of structures like triplets (see Clarke, 1987). Therefore, the adaptation of scales to subdivisions, such as triplets, emerges as a significant area of research deserving detailed examination. Considering the focus of our paper, we will not delve into the intricacies of this topic. In our research, we chose not to include questions involving triplets when evaluating the Keith and Pressing scales to ensure methodological rigor and avoid potential misinterpretations.

2005; Thul & Toussaint, 2008), with greater distances indicating higher complexity. The measure is defined as the sum of the inverse distances of all notes from their nearest strong beats, normalized by the total number of notes. Unlike the Keith scale, this method integrates the tempo as a significant component. It assesses perceptual complexity by assuming that the rhythm is played according to a predetermined tempo. In the original article (Gómez et al., 2005), the method is described as follows:

> "Firstly, notes are supposed to end where the next note starts. Let ei, ei+1 be two consecutive strong beats in the meter. By strong beats we just mean pulses. Also, let x denote a note that starts after or on the strong beat ei but before the strong beat ei+1; we first define T(x) = mind(x, ei), d(x; ei+1), where d denotes the distance between notes in terms of duration. Here the distance between two adjacent strong beats is taken as the unit and, therefore, the distance d is always a fraction. . .
> The WNBD measure D(x) of a note x is then defined as follows: 0, if x = ei; 1/T(x) if note x ≠ ei ends before or at ei+1; 2/T(x) , if note x 6 ≠ ei ends after ei+1 but before or at ei+2; and 1/T(x) , if note x ≠ ei ends after ei+2. Let n denote the number of notes of a rhythm. Then, the WNBD measure of a rhythm is the sum of all D(x), for all notes x in the rhythm, divided by n." (Gómez et al., 2005).

In a simplified summary: Each beat is assigned a value, T(x), and evaluated based on three conditions: 1) Note start on a strong beat do not receive any value. 2) If the sycopated note does not exceed the second strong beat, the value is 1/T(x). 3) If the beat intersects with the strong beats, the value is 1/T(x) but if the beat surpasses the second strong beat, the value is 2/T(x). The total value is then divided by the number of beats in the rhythm. To demonstrate this method with the example in Figure 1; the rhythm with 14 beats, there are 8 strong beats that will not take a value and 6 beats between these beats, each written as a triplet. Each of these beats is T(x)= 1/3 away from the strong beats. This means that each of them will score 3 points when processed. As a result, 3+3+3+3+3+3+3+3+3=18 is divided by 14. WNBD score is 1.28 for same rhythm in Figure 1.

Pressing scale

The Pressing scale scores the complexity of rhythmic structures using values prepared in accordance with cognitive complexity perception (Pressing, 1999). The fundamental idea involves grouping rhythmic patterns, comprising any combination of 16th-notes in one beat structures and assigning a value to each group. Implementation-wise, it bears a strong resemblance to Keith scale. In his introductory work on the scale, Pressing defined the "pattern complexity(pc)" scores for specific rhythmic patterns (rhythmic cells), as depicted in Figure 3. To measure complexity in this method, the pc values in Figure 3 are summed as the associated rhythmic cell appears in the beats. The sum of all pc values gives the complexity of the rhythm.

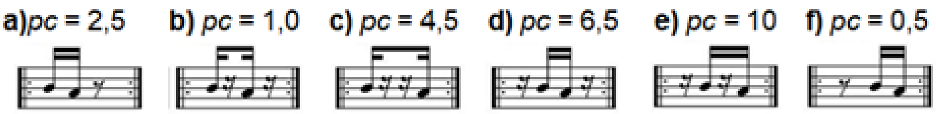

Figure 3. Rhythms with various pattern complexities (pc) in the Pressing Scale (Pressing, 1999)

METHOD

Our research involved two studies, drawing on the results obtained from two groups of participants, expert raters and student performers. Both studies were conducted using the Google Colab platform, enabling online data collection.

The survey tool was designed to anonymously store the information and voice recordings received from the participants. Participants involved in the research were provided with comprehensive information regarding the nature, purpose, and procedures of the study. Prior to the studies, ethical clearance was obtained from Izmir Democracy University, Publication Ethics Committee for Science and Engineering Sciences. Participants were assured of confidentiality and their voluntary participation. They were informed about their rights to withdraw from the study at any stage without penalty. Consent forms indicating their understanding and willingness to participate in accordance with ethical rules were obtained by all participants through the google colab tool with a checkbox.

Rhythm pattern generation

The rhythmic patterns utilized in this study were automatically generated through a custom rule-based generative tool coded in Python.[5] Data recorded from a real exam was also taken into account to prepare the rules.[6] These patterns were

[5] For this tool: https://github.com/yavuzburuk/ritim (last access date: 24.11.2024)
[6] MAST rhythm dataset. Zenodo. https://doi.org/10.5281/zenodo.7243752 (last access date: 24.11.2024)

designed to model rhythm questions commonly encountered in undergraduate conservatory entrance exams, serving as a reference point. Additionally, these generated patterns were converted into sound files, simulating the auditory environment encountered during exam scenarios (e.g., the sound of a pen tapping on a table in a reverberant room). The generative tool was employed to create random rhythmic patterns, focusing on two distinct levels, each featuring two different rhythm types (options). Level 1 consisted of patterns without syncopations, whereas Level 2 included syncopations within the second measures of each rhythmic phrase. Option 1 centered on triplet divisions, while Option 2 incorporated patterns utilizing eighth and sixteenth-note divisions. Each rhythm question spanned two measures and featured 12-14 beats, with nonrecurring patterns.

Study 1

In the first study, seven participants who were all expert raters evaluated a sample of 80 rhythms, generated as described above. All participants served as faculty members in the musicology, music theory, and composition departments of the conservatories affiliated to various universities. In addition to rating the questions, participants were also asked about their expertise in preparing questions for conservatory entrance exams and their experiences as jury members. Subsequently, they were tasked with assessing 20 questions from two different levels, each having two options, resulting in a total of 80 questions. The 5-point Likert-type scale they used ranged from 1 (very easy) to 5 (very difficult) to determine the suitability of the questions for inclusion in conservatory entrance exams. The results of the study clarified that questions rated 1 or 5 were considered unsuitable for exams (too easy or complex). The participants had the autonomy to view the questions randomly within their designated question set. The rhythms were presented in the form of notation, but the participants could listen to the MIDI-realized audio of each rhythm if they desired (Figure 4).

Figure 4. Study 1 – screenshot a question from the online survey

Study 2

The second study involved 11 participants; 5 female, 6 male and ages ranging between 18-28 (mean age 25). All participants were conservatory students, who were instructed to replicate 16 rhythms selected from the initial pool of 80 (Figure 5) by tapping them on a table using a pen. The 16 questions encompassed various levels of difficulty and were selected according to the evaluations of the expert raters in Study 1 and the scores obtained from different complexity scales. Particular attention was given to the ratings provided by the three expert participants with prior experience of preparing exam questions.

Figure 5. Algorithmically generated rhythm questions that were used in our study

The rhythm questions were presented randomly. Participants listened to the audio recording and were required to replicate the rhythms as described above. Each participant had three opportunities to listen, with the rhythm played immediately upon pressing the play button, and a repeat after a 4-second pause. The tool recorded the number of listening repetitions and the time taken for each question. Participants had the option to save their responses or request to play the question again after reviewing their recording. This process yielded a total of 176 audio recordings, which were subsequently evaluated by two experts and assigned one of the following labels: 1 (*unsuccessful*); 2 (*major mistake*); 3 (*minor mistake*) and 4 (*successful*).

RESULTS

Study 1

We assessed the reliability of expert evaluations using Cronbach's alpha, yielding a high internal consistency of α = .866. To measure inter-rater agreement, Krippendorff's alpha coefficient was calculated, resulting in α = .400 for the seven experts. To enhance reliability, we re-calculated Krippendorff's alpha using a subset of three expert raters who were experienced in preparing conservatory entrance exam questions (P1, P3, P7). This produced the higher value of α = .512, signifying moderate agreement. Subsequent analyses were based on the data from these three participants. Pearson correlations were used to assess the agreement between expert opinions and complexity scales. All units exhibited skewness and kurtosis values within the range of -1.5 to +1.5, indicating a normal data distribution.

As Keith and Pressing scales do not measure triplet values, the comparisons for these two scales were based on 40 questions without triplets. For the other scales, evaluations were made based on 80 questions. The Pearson correlation relationships between the scales, the means of the experts, the scales and the experts are as shown in Table 1. Accordingly, Pressing and Keith for 40 questions and WNBD, Tanguiane and Lempel-Ziv for 80 questions including triplets, respectively, are in line with the experts' averages.

Table 1. Pearson correlations of expert ratings with five different rhythmic complexity measures

	M(*SD*)	N	P1	P3	P7	Mean P1, P3, P7	LZ	Pressing	Keith	Tanguiane	WNBD
P1	3.88(0.85)	80	1								
P3	3.57(1.30)	80	.741**	1							
P7	4.40(0.73)	80	.591**	.535**	1						
Mean P1, P3, P7	4.12(0.79)	80	.892**	.921**	.767**	1					
LZ	18,02(1.81)	80	.222*	.337**	.162	.296**	1				
Pressing	19.78(8.09)	40	.783**	.692**	.715**	.815**	.085	1			
Keith	7.07(3.04)	40	.672**	.687**	.737**	.780**	.326*	.866**	1		
Tanguiane	3.48(1.16)	80	.637**	.571**	.533**	.665**	-.054	.709**	.749**	1	
WNBD	1.52 (0.36)	80	.661**	.671**	.501**	.715**	.526**	.885**	.811**	.537**	1

LZ: Lempel-Ziv
*p < .05; **p <.01.

Excluding triplet questions, the internal consistency for the remaining 40 questions, as determined by Cronbach's alpha test, was α = .787. Using the Krippendorff method, the scores of the three expert raters with expertise in question preparation (P1, P3, and P7) yielded α = .589. Comparing the scales with the mean scores of these three participants (M = 4.125; SD = 0.79), Pearson's correlations were r = .815 for Pressing (M = 19.7875; SD = 8.09787), r = .800 for WNBD (M = 1.688; SD = ,36319), r = .780 for Keith (M = 7.0750; SD = 3.0499), r = .724 for Tanguiane (M = 3.2750; SD = 1.2807), and r = .408 for Lempel-Ziv (M = 19.7250; SD = .6300) (all p <.001). Consequently, the most successful scales among the 40 questions were Pressing, WNBD, Keith, Tanguiane, and Lempel-Ziv.

It is evident that the Pressing, WNBD, Keith, and Tanguiane complexity scales closely align with expert judgments regarding rhythm difficulty. These scales demonstrate high internal consistency and provide accurate approximations of expert opinions in predicting rhythm difficulty. However, their performance is less optimal when assessing triplet patterns. Additionally, the two difficulty levels in our question preparation algorithm consistently aligns with expert opinions.

Study 2

Krippendorff's alpha coefficient, which did not yield a high result in the initial study, increased to α =.873 for the 16 selected questions. Furthermore, the internal consistency test utilizing ratings from three participants demonstrated a notably high Cronbach's alpha value of .948. In the case of these 16 questions, calculations indicated strong consistency (Cronbach's α = .949) and agreement (Krippendorff's α = .6782) among all seven participants in the first study. Thus, the reliability of both internal consistency and inter-rater agreement for these 16 questions can be confidently affirmed. Detailed expert ratings and complexity scores for the 16 questions utilized in Study 2 are presented in Table 2.

Table 2. Expert participants' ratings and rhythmic complexity scores for the 16 questions

Question	P1	P3	P7	Mean P1,3,7	LZ	Pressing	Keith	Tanguiane	WNBD
1	3	3	3	3	16			3	1,29
2	3	3	4	3,33	17			3	1,29
3	2	1	3	2	16			3	1,15
4	4	3	4	3,67	15			3	1,29
5	3	3	3	3	19	17	6	2	1,5
6	5	4	5	4,67	20	23	8	2	1,69
7	2	1	3	2	18	5	0	1	1
8	4	4	4	4	20	17	6	2	1,85
9	5	5	5	5	17			5	1,93
10	5	5	5	5	16			6	1,71
11	4	4	4	4	17			5	1,75
12	4	4	4	4	15			4	1,25
13	5	5	5	5	20	27	9	3	1,84
14	5	5	5	5	20	33	12	4	2,33
15	4	4	4	4	20	11,5	4	5	1,28
16	4	4	5	4,33	20	25,5	10	5	1,71

The scores of the student participants were then compared to these complexity ratings. Pearson correlations between the experts' ratings and students' performances for the 16 questions revealed a significant and strong relationship, indicating that as complexity, as assessed by the experts, increased, student performances significantly declined (Table 3). The mean scores of the 16 questions provided by the seven initial study participants also significantly correlated with student performance scores (r = -.853, p <.001), affirming the accuracy of expert predictions in assessing rhythm repetition difficulty.

Table 3 presents a summary of the interplay between complexity scores derived from three raters, five complexity scales, and student performance. Notably, the Lempel-Ziv scale exhibits no significant correlation with students' performances (r = -.395, p = .13), a result possibly influenced by the relatively small sample size and the scale's focus on longer sequences. Similarly, the Pressing scale, with only eight questions, does not significantly correlate with student performances (r = -.533, p = .174). The Keith scale, tested on eight questions, also lacks a significant correlation with student performance (r = -.589, p = .124). In contrast, the Tanguiane Scale displays a significant correlation with students' success in rhythm repetition (r = -.570, p = .021), indicating moderate alignment. Notably, the WNBD scale shows the highest performance, with a significant correlation with student achievements (r = -.661, p = .005), making it the most successful in predicting rhythm difficulty.

In summary, the expert participants provided the most accurate predictions for rhythm repetition difficulty, followed by the WNBD scale and the Tanguiane scale. The Pressing and Keith scales, tested with a limited question pool, may not have yielded significant results due to the small sample size. The Lempel-Ziv scale, despite its focus on longer sequences, also exhibited low efficiency, emerging as the least successful scale in both studies. Additionally, it is noteworthy that there is a significant negative correlation (r = -.606, p = .013) between the mean time spent on each question and performance scores, implying that for highly complex rhythms, repeated listening does not contribute significantly to performance success.

Table 3. Pearson correlations of the experienced graders and the complexity mesures with student performance scores (SPS)

	SPS	Mean Time	P1	P3	P7	Mean P1,3,7	LZ	Pressing	Keith	Tanguiane	WNBD
SPS	1	-.606*	-.830**	-.843**	-.738**	-.839**	-.395	-.533	-.589	-.570*	-.661**
Mean Time	-.606*	1	.571*	.612*	.583*	.611*	.157	.555	.596	.418	.534*
N	16	16	16	16	16	16	16	8	8	16	16

SPS: Student Performance Score
LZ: Lempel-Ziv
*p < .05; **p <.01.

DISCUSSION AND CONCLUDING REMARKS

In this study we investigated two fundamental aspects of rhythmic complexity: the cognitive processes related to understanding rhythmic structures and the complexity linked to their execution. The first study focused on experts, allowing us to observe the scales' success in predicting difficulty within a relatively stable group. In the second study, despite the expectation of consistent results from conservatory students, the desired clarity still needs to be achieved. Nevertheless, the findings suggest potential success with an increased sample size We found a notable consistency between students with early rhythmic training and the assessments made by expert raters with expertise in evaluating rhythmic proficiency. This alignment suggests a potential for agreement in human judgments, even in the face of the inherent subjectivity associated with complexity. However, each complexity scale offers a distinctive perspective on the conceptualisation of rhythm. Consequently, the measurability of rhythm is also related to the manner in which it is expressed.

The scales under investigation utilize diverse methods to represent rhythmic structures, which they define as stimuli over time. For instance, the Tanguine scale focuses on the divisibility hierarchy between rhythmic structures, the WNBD scale emphasizes the proportions to the strong beat, whereas the Lempel-Ziv scale assigns binary codes to the rhythmic units.The WNBD scale offers a comprehensive approach to measuring rhythmic structures by defining a distance proportion within the time framework. However, this method requires an understanding of the rhythm as a whole, necessitating the identification of strong beats. However, it necessitates knowledge of the rhythm as a whole, requiring the identification of strong beats. On the other hand, the Lempel-Ziv scale quantifies each value (either 1 or 0) according to a predetermined unit (as for example one millisecond), enabling the expression and measurement of the rhythm without prior knowledge. Nevertheless, it disregards the musical attributes of these units and lacks a perceptual basis. but completely ignores human pattern recognition information, thereby permitting distinct values to be assigned at the basic unit level.

The results demonstrated that the complexity scales did not consistently yield results across different attempts. The inadequacies of the scales in this regard have been previously identified in the field (Thul, 2008); nevertheless, through these two conducted studies, one can discuss the reasons behind these inadequacies more effectively in further research. The differences between the two studies can be attributed to their distinct focuses - judgmental difficulty in Study1 and performative challenges in Study2. While achieving an average level of success for both studies is possible, the significant discrepancy observed with the Pressing scale requires further examination. The Pressing scale, based on human-based cognitive methods, showed substantial alignment with expert opinions in Study1 but did not demonstrate significant agreement with students' performances in Study2. This raises questions about the suitability of cognitive-based scales in assessing rhythmic complexity in a performative context.

Despite their differences, the scales, utilizing unit-based scoring approaches to represent rhythm, yield similar results when used to assess perceptual complexity. However, it is important to note that pattern-matching scales may sometimes be misleading. Our selected sample revealed significant variations for rhythms predicted to be of similar difficulty. This discrepancy highlights the limitations of these scales in capturing the nuanced perceptual differences in rhythm. The expression of rhythm greatly influences the structure of these complexity scales, and these assumptions may not consistently align with perceptual aspects. Therefore, further testing of complexity scales with grouped rhythms or matching according to selected criteria is essential to ensure their consistency and necessitates reevaluation.

Ethics Committee Approval: Ethics committee approval was received for this study from the ethics committee of Izmir Democracy University, Publication Ethics Committee for Science and Engineering Sciences (Date: September 1, 2022, Number: 2022/04).

Informed Consent: Written informed consent was obtained from participants who participated in this study.

Peer Review: Externally peer-reviewed.

Author Contributions: Conception/Design of Study- C.Y., O.B.; Data Acquisition- C.Y., O.B.; Data Analysis/Interpretation- C.Y., O.B., B.B., Y.B.; Drafting Manuscript- C.Y., O.B., B.B., Y.B.; Critical Revision of Manuscript- C.Y., O.B., B.B., Y.B.; Final Approval and Accountability- C.Y., O.B., B.B., Y.B.; Material and Technical Support- B.B., Y.B.; Supervision- O.B.

Conflict of Interest: The authors have no conflict of interest to declare.

Grant Support: This study was supported by Scientific and Technological Research Council of Turkiye. Grant Number: 121E198. The authors thank TUBITAK for their support.

Etik Komite Onayı: Bu çalışma için etik komite onayı İzmir Demokrasi Üniversitesi, Fen ve Mühendislik Bilimleri Yayın Etiği Kurulu'ndan (Tarih: 1 Eylül 2022, Sayı: 2022/04) alınmıştır.

Katılımcı Onamı: Yazılı onam bu çalışmaya katılan katılımcılardan alınmıştır.

Hakem Değerlendirmesi: Dış bağımsız.

Yazar Katkıları: Çalışma Konsepti/Tasarım- C.Y., O.B.; Veri Toplama- C.Y., O.B.; Veri Analizi/Yorumlama- C.Y., O.B., B.B., Y.B.; Yazı Taslağı- C.Y., O.B., B.B., Y.B.; İçeriğin Eleştirel İncelemesi- C.Y., O.B., B.B., Y.B.; Son Onay ve Sorumluluk- C.Y., O.B., B.B., Y.B.; Malzeme ve Teknik Destek- B.B., Y.B.; Süpervizyon – O.B.

Çıkar Çatışması: Yazarlar çıkar çatışması bildirmemiştir.

Finansal Destek: Bu çalışma Türkiye Bilimsel ve Teknolojik Araştırma Kurumu tarafından desteklenmiştir. Hibe Numarası: 121E198. Yazarlar destekleri için TÜBİTAK'a teşekkürlerini sunarlar.

ORCIDs ID of the author(s) / Yazar(lar)ın ORCID ID'leri

Cihan YAYGIN	0000-0002-3329-3650
Ozan BAYSAL	0000-0002-7271-9095
Yavuz BURUK	0009-0001-2223-6087
Barış BOZKURT	0000-0002-0177-0758

REFERENCES / KAYNAKLAR

Arom, S., & Ligeti, G. (1991). African Polyphony and Polyrhythm. (*No Title*). https://doi.org/10.1017/cbo9780511518317

Baysal, O., & Bozkurt, B. (2022). MAST rhythm dataset [Data set]. Zenodo. https://doi.org/10.5281/zenodo.7243752

Bolton, T. L. (1894). Rhythm. *The American Journal of Psychology, 6*(2), 145. https://doi.org/10.2307/1410948

Buffardi, L. (1971). Factors affecting the filled-duration illusion in the auditory, tactual, and visual modalities. *Perception & Psychophysics, 10*(4), 292–294. https://doi.org/10.3758/bf03212828

Clarke, E. F. (1987). Levels of structure in the organization of musical time. *Contemporary Music Review, 2*(1), 211–238. https://doi.org/10.1080/07494468708567059

de Fleurian, R., Blackwell, T., Ben-Tal, O., & Müllensiefen, D. (2016). Information-Theoretic Measures Predict the Human Judgment of Rhythm Complexity. *Cognitive Science, 41*(3), 800–813. https://doi.org/10.1111/cogs.12347

Essens, P. (1995). Structuring temporal sequences: Comparison of models and factors of complexity. *Perception & Psychophysics, 57*(4), 519–532. https://doi.org/10.3758/bf03213077

Fraisse, P. (1956). Fraisse (Paul). — Les Structures Rythmiques. Etude psychologique, Erasme, 1956, Bruxelles-Paris. *Bulletin de Psychologie, 13*(177), 655–657. https://www.persee.fr/doc/bupsy_0007-4403_1960_num_13_177_8359_t1_0655_0000_1

Fraisse, P. (1982). Rhythm and Tempo. *Psychology of Music, 1*(1), 149–180. https://doi.org/10.1016/b978-0-12-213562-0.50010-3

Gómez, F., Melvin, A., Rappaport, D., & Toussaint, G. (n.d.). *Mathematical Measures of Syncopation*. Retrieved January 31, 2024, from https://research.cs.queensu.ca/home/daver/Pubs/MyPDF/MeasureSycopa.pdf

Keith, M. (1991). From Polychords to Pólya: Adventures in Musical Combinatorics. Vinculum Press.

Lempel, A., & Ziv, J. (1976). On the Complexity of Finite Sequences. *IEEE Transactions on Information Theory, 22*(1), 75–81. https://doi.org/10.1109/tit.1976.1055501

Lerdahl, F., & Jackendoff, R. S. (1996). A Generative Theory of Tonal Music, reissue, with a new preface. 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. https://doi.org/10.2307/40285271

Povel, D.-J. (1981). Internal representation of simple temporal patterns. *Journal of Experimental Psychology: Human Perception and Performance, 7*(1), 3–18. https://doi.org/10.1037/0096-1523.7.1.3

Povel, D.-J. (1984). A theoretical framework for rhythm perception. *Psychological Research, 45*(4), 315–337. https://doi.org/10.1007/bf00309709

Povel, D.-J., & Essens, P. (1985). Perception of Temporal Patterns. Music Perception: *An Interdisciplinary Journal, 2*(4), 411–440. https://doi.org/10.2307/40285311

Pressing, J. (1999). *Cognitive complexity and the structure of musical patterns.* http://dub.ucsd.edu/Mu206/CogComplex-music.pdf

Shmulevich, I., & Povel, D. J. (2000). Measures of Temporal Pattern Complexity. *Journal of New Music Research, 29*(1), 61–69. https://doi.org/10.1076/0929-8215(200003)29:01;1-P;FT061

Smith, M. L., & Honing, H. (2006). Evaluating and Extending Computational Models of Rhythmic Syncopation in Music. *International Computer Music Conference Proceedings, 2006*. http://hdl.handle.net/2027/spo.bbp2372.2006.139

Sternberg, S., Knoll, R., & Zukojsky, P. (1982). *The Psychology of Music* (pp. 181–239). Academic Press. http://www.musicalobservations.com/publications/timing.pdf

Tanguiane, A. (1993). Artificial Perception and Music Recognition. In *Springer eBooks*. Springer Berlin, Heidelberg. https://doi.org/10.1007/bfb0019384

Tanguiane, A. (1994). A Principle of Correlativity of Perception and Its Application to Music Recognition. *Music Perception, 11*(4), 465–502. https://doi.org/10.2307/40285634

Thomas, E. C., & Brown, I. (1974). Time perception and the filled-duration illusion. *Perception & Psychophysics, 16*(3), 449–458. https://doi.org/10.3758/bf03198571

Thul, E. (2008). Measuring the complexity of musical rhythm. *Escholarship.mcgill.ca.* C3S2E '08: Proceedings of the 2008 C3S2E Conference. https://escholarship.mcgill.ca/concern/theses/r494vp484

Toussaint, G. (2002). *Mathematical Connections in Art, Music, and Science A Mathematical Analysis of African, Brazilian and Cuban Clave Rhythms.* https://archive.bridgesmathart.org/2002/bridges2002-157.pdf

Toussaint, G. (2004). *A Comparison Of Rhythmic Similarity Measures*. International Society for Music Information Retrieval Conference. https://www.ee.columbia.edu/~dpwe/ismir2004/CRFILES/paper134.pdf

Toussaint, G. (2005a). *Classification and Phylogenetic Analysis of African Ternary Rhythm Timelines.* https://cgm.cs.mcgill.ca/~godfried/teaching/mir-reading-assignments/Classification-and-Phylogenetic-Analysis-of-African-Ternary-Rhythm-Timelines.pdf

Toussaint, G. (2005b). *The Euclidean Algorithm Generates Traditional Musical Rhythms.* https://cgm.cs.mcgill.ca/~godfried/publications/banff-extended.pdf

How cite this article / Atıf Biçimi

Yaygın, C., Baysal, O., Buruk, Y., & Bozkurt, B. (2024). A comparative study of rhythmic complexity measures in music performance. *Konservatoryum – Conservatorium, 11*(2), 478–487. https://doi.org/10.26650/CONS2024-1562373