COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 41 (2017) 800–813

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12347

Information-Theoretic Measures Predict the Human Judgment of Rhythm Complexity

Remi de Fleurian,[a] Tim Blackwell,[b] Oded Ben-Tal,[c] Daniel Müllensiefen[a]

[a]*Department of Psychology, Goldsmiths, University of London*
[b]*Department of Computing, Goldsmiths, University of London*
[c]*Department of Music, Kingston University*

Received 14 July 2014; received in revised form 16 November 2015; accepted 7 December 2015

Abstract

To formalize the human judgment of rhythm complexity, we used five measures from information theory and algorithmic complexity to measure the complexity of 48 artificially generated rhythmic sequences. We compared these measurements to human prediction accuracy and easiness judgments obtained from a listening experiment, in which 32 participants guessed the last beat of each sequence. We also investigated the modulating effects of musical expertise and general pattern identification ability. Entropy rate and Kolmogorov complexity were correlated with prediction accuracy, and highly correlated with easiness judgments. A logistic regression showed main effects of musical training, entropy rate, and Kolmogorov complexity, and an interaction between musical training and both entropy rate and Kolmogorov complexity. These results indicate that information-theoretic concepts capture some salient features of the human judgment of rhythm complexity, and they confirm the influence of musical expertise on complexity judgments.

Keywords: Rhythm perception; Rhythm complexity; Information theory measures; Entropy rate; Kolmogorov complexity; Musical expertise

1. Introduction

The notion of complexity in art has been of interest to research in psychology for more than a century. Following Wundt's idea that the enjoyment of a stimulus depends on its complexity (Wundt, 1896), a series of studies investigated the relationship between complexity and esthetic perception. In this endeavor, various ways of assessing complexity have been used. For instance, in his famous study leading to the finding of an inverted-U relationship between hedonic value and arousal potential, the visual stimuli used by

Correspondence should be sent to Remi de Fleurian Department of Psychology Goldsmiths, University of London Whitehead Building New Cross London, SE14 6NW United Kingdom. E-mail: r.defleurian@gmail.com

Berlyne (1970) were arbitrarily assigned to a category of complexity based on their background and on the number of human figures being featured. In the musical domain, stimulus complexity has been manipulated by varying the loudness, pitch, and duration of tones in tone sequences (Vitz, 1966) or by increasing the variety of chords and the amount of syncopations (Heyduk, 1975). Furthermore, complexity has sometimes been quantified by asking the participants to provide subjective ratings of complexity (Heyduk, 1975; North & Hargreaves, 1995). All these studies reported an inverted-U relationship between liking and complexity, and they highlighted the modulatory effects of repeated exposure and familiarity, musical training, and individual preference for a specific level of complexity.

The research presented in this study focuses on formalizing the human judgment of rhythm complexity, by using an information-theoretic or algorithmic measure of complexity as opposed to the subjective or arbitrary measures presented so far. There are several possible measures of complexity to consider. Some of them are derived directly from music theory, such as measures of rhythmic syncopation (Fitch & Rosenfeld, 2007; Gomez, Thul, & Toussaint, 2007), or from human performance, such as measures of rhythm reproduction ability (Essens, 1995; Povel & Essens, 1985). Other measures benefit from a solid information-theoretic grounding and have been shown to capture various features of the cognitive processing of music. For instance, Shannon entropy (Shannon, 1948) was found to be a good predictor of the amount of attention directed to a single voice in a piece of music containing multiple voices (Madsen & Widmer, 2006). Hansen and Pearce (2012) reported that Shannon entropy predicts the uncertainty of probe tones in melodies. Moreover, they identified an entropy-by-expertise interaction in the ratings of uncertainty, showing a difference in how musicians and non-musicians process complex music. Another candidate, the predictive information rate, is a measure of how much an event within a sequence reduces the uncertainty about the following events, while taking into account the information content of all the previous events (Abdallah & Plumbley, 2009, 2010, 2012; Bialek, Nemenman, & Tishby, 2001). Madsen and Widmer (2006) also mentioned the possible suitability of measures based on compression algorithms, such as LZ78 (Ziv & Lempel, 1978) or LZW (Welch, 1984). Indeed, LZ compressibility has been empirically tested for its ability to predict human judgments of rhythm complexity by Shmulevich and Povel (2000). The measure did not perform well, but the authors attributed this to the short length of the sequences used in their experiment, and suggested that LZ compressibility is likely to perform better with longer sequences. Moreover, LZ78 is able to provide an approximation of Kolmogorov complexity (Kolmogorov, 1965; Li & Sleep, 2004), a measure of randomness that has been successfully used to cluster melodies or music in the MIDI format in similar groups, based on their compressibility (Cilibrasi, Vitanyi, & de Wolf, 2004; Li & Sleep, 2004, 2005).

Schmidhuber's (2009) theoretical model explored the relationship between data compression and esthetics. He compared the human mind to a self-improving, computationally limited observer, and approached the question of complexity from an algorithmic point of view: He stated that beauty comes from the challenge of discovering patterns and the ability to compress new data; that is, a moderately complex stimulus is perceived as beautiful because the strategies developed to understand that stimulus facilitate the subsequent understanding of similar stimuli. According to him, the limitations of the mind are a determining aspect of what an observer considers as enjoyable or not. It is therefore possible to consider that individual differences in pattern identification abilities might modulate the perception and enjoyment of complex stimuli. He also argued that the perceived complexity of a stimulus changes with exposure, in accordance with Berlyne's (1970) and Heyduk's (1975) findings. Moreover, the influence of complexity is not limited to the domain of art and esthetics. For instance, another study showed that children's attention shifted away from both visual and auditory stimuli that were either too simple or too complex, and focused on stimuli of intermediate complexity that were interesting but still understood, which suggests an effect of complexity on attention and learning (Kidd, Piantadosi, & Aslin, 2012, 2014).

For this study, we select five measures with a potential to formalize the human judgment of rhythm complexity: Shannon entropy (H), entropy rate (h), excess entropy (E), transient information (T), and Kolmogorov complexity (K). These measures have firm theoretical foundations and are defined at a high level of abstraction and generality; that is, they may be used to characterize the complexity of structures in any domain, by quantifying the complexity of a sequence of symbols, irrespective of what the symbols stand for.

The information-theoretic measures (H, h, E, and T) require a stationary probability distribution and, since they are based on probability, require an infinite symbol sequence. It could be argued that a non-stationary distribution (i.e., in which the symbol probability changes over "time," that is, as the sequence is read from left to right) might provide a better model of rhythm perception. However our aim here is not to provide a model per se but to investigate if abstract measures of complexity align with the perception of rhythm complexity. If they fail to quantify perceived complexity, then one explanation might be that subjects do indeed change their model of a rhythm sequence as the sequence progresses. Kolmogorov complexity, K, however, makes no assumption on stationarity and non-finiteness, and any discrepancy between K and any of H, h, E, and T might indicate that this assumption is too simplistic.

The most fundamental measure is the Kolmogorov complexity, usually denoted K (Li & Vitanyi, 2008). It is defined as the length of the shortest computer program that can generate a given symbol string. It is convenient, but not necessary, to use a binary alphabet in which symbol strings are sequences of 0's and 1's. K measures the complexity of a single object. A long and very predictable sequence (e.g., 1, 1, 1,...,1) could be produced by a very short program and therefore has a small K-complexity. Such a sequence is highly compressible. On the other hand, a random sequence has a large K-complexity because it can only be produced by a long program. A random sequence has no structural property that enables compression. K-complexity therefore measures randomness. K-complexity has the disadvantage of being incomputable, although upper bounds can be estimated by the degree of compressibility with respect to a particular compressor.

Shannon entropy, H, measures the information content of a typical symbol string from a particular source. It is given by the formula

$$H = -\sum p_i \log p_i \tag{1}$$

where p_i is the probability of the i'th symbol (Cover & Thomas, 2006). Although entropy is a function of a probability distribution and not of a single object, it can be computed for a single long message under the assumption that the distribution of symbols in the message matches the underlying probability distribution. Once more it ranges from small values for a very predictable sequence to high values for incompressible sequences.

The entropy rate, h, is the limit of the Shannon entropy per symbol of substrings of increasing length L. It captures the inherent randomness of a sequence when all correlations over longer and longer subsequences have been taken into account. It is zero for any repetitive (periodic) sequence and of value one for a sequence of 0's and 1's generated by the toss of a fair coin. Denoting the entropy of substrings of length L by $H(L)$, then

$$h = \lim_{L \to \infty} \frac{H(L)}{L} \tag{2}$$

The excess entropy, denoted E, has been described by numerous researchers and has been given various names (effective measure complexity, stored information, predictive information, Renyi entropy of order 1), although the mathematical definition is identical. It is defined as

$$E = \lim_{L \to \infty} H(L) - hL \tag{3}$$

If $H(L)$ acquires the asymptotic form $H(L) \longrightarrow H^{\infty} = hL + E$, then $E = H^{\infty}(0)$. Excess entropy has the advantage that it can distinguish repetitive patterns of different period. It has various interpretations. It may be related to the intrinsic memory of the source of the sequence or to the mutual information between two semi-infinite halves of the sequence. It is zero for a random sequence, and it is proportional to the logarithm of the period of a repetitive sequence.

The final measure that we consider in this study, the transient information, T, has been proposed as a means of distinguishing between sequences of the same period (and hence of identical h and E) and entropy. It measures the difficulty in synchronizing to a periodic process and captures a structural property that E fails to pick up (Crutchfield & Feldman, 2003). In terms of the asymptotic length-L entropy, it is calculated as

$$T = \sum_{L=0}^{\infty} H^{\infty}(L) - H(L) \tag{4}$$

In this study, we use these five measures in an experiment aimed at predicting the difficulty of correctly guessing the end of rhythmic patterns of various complexity, and we investigate two suspected modulatory effects of rhythm perception: musical expertise (Hansen & Pearce, 2012; North & Hargreaves, 1995; Vitz, 1966) and pattern identification ability (Schmidhuber, 2009). It is worth noting that most listeners associate rhythm with a regular pulse, but more broadly, rhythm can be defined as the general enfolding of binary events in time. In the sequences we present to our participants, a notional pulse is displayed visually but aurally; most sequences do not have a strong underlying pulse. Nevertheless, the task requires participants to make judgments about durational patterns, that is, rhythm (Parncutt, 1994).

2. Method

2.1. Participants

Thirty-two participants (15 women and 17 men), ranging in age from 21 to 57 years (*M* = 26.9 years, *SD* = 6.9 years), were recruited through various social networks and among graduate students at Goldsmiths, University of London. Most of the participants had received some musical training at some point in their lives (*M* = 4.67 years of musical training, *SD* = 3.46 years).

2.2. Materials

We selected 16 different generative algorithms with a diverse range of complexity values on the five selected complexity measures (see Supplementary Material for details of the generative algorithms). Some algorithms produced easily identifiable patterns, but we deliberately covered a wide range of complexity values in order for most of the stimuli to be hard to fully apprehend. We generated a sequence of 10^4 symbols (1's and 0's) from each algorithm, and we randomly extracted three subsequences of 50 symbols from each sequence, with the following restriction: The subsequences extracted from the same generative algorithm could not all end with the same symbol. The last symbol was removed from each of the 48 resulting subsequences and kept in a separate file to be used as the answer key for the prediction task (see Supplementary Material for the full set of subsequences and the answer key). SuperCollider (Wilson, Cottle, & Collins, 2011) was then used to replace the 1's by drum hits[1] and the 0's by rests, all representing quarter notes at 150 bpm. Each subsequence was therefore almost 20 s long. Extracting three subsequences for each generative algorithm was a compromise, aimed at balancing the requirements of our analysis with the overall length of the experiment. We generated two additional training sequences, of 20 symbols each, to accustom the participants to the experimental procedure.

Due to the nature of the task, which required accurate following of the beat subsequences, we also provided the participants with a visual representation of the subsequences. We designed a beat visualization tool using PowerPoint and iMovie (Fig. 1), which consisted of a spiral made of 50 white dots. When a subsequence was played, a black dot gradually filled the spiral, synchronized with the beat, until it reached the center, at which point a question mark appeared to make sure the participants knew exactly which beat they needed to provide a judgment for.

We used the self-report questionnaire of the Goldsmiths Musical Sophistication Index, or Gold-MSI (Müllensiefen, Gingras, Musil, & Stewart, 2014), to investigate the effects of musical expertise. The questionnaire allows the calculation of a general score of musical sophistication as well as individual scores for five subscales. For this study, we only used the subscales "Perceptual Abilities" and "Musical Training," as well as the General Musical Sophistication, because they are most closely related to the individual differences effects that previous studies have found (Hansen & Pearce, 2012; North & Hargreaves, 1995; Vitz, 1966).

Finally, we selected a shortened version of the Raven's Progressive Matrices (RPM), the Advanced Progressive Matrices: Set I (APM1), to assess the participants' ability to identify and reason with visual patterns. The RPM is a widely recognized test of pattern detection, and although it is often used as a predictor of "general intelligence," its initial purpose was to measure eductive ability, which is the "meaning-making ability" that allows one to make sense of more or less chaotic stimulus configurations (Raven, Raven, & Court, 1998).

2.3. Procedure

Each participant was tested individually in a quiet lab environment. The participants were first shown the two training sequences as many times as they wanted and were given instructions on how to complete the answer sheet. They were requested, for each subsequence, to indicate whether the last beat was supposed to be a hit or a rest

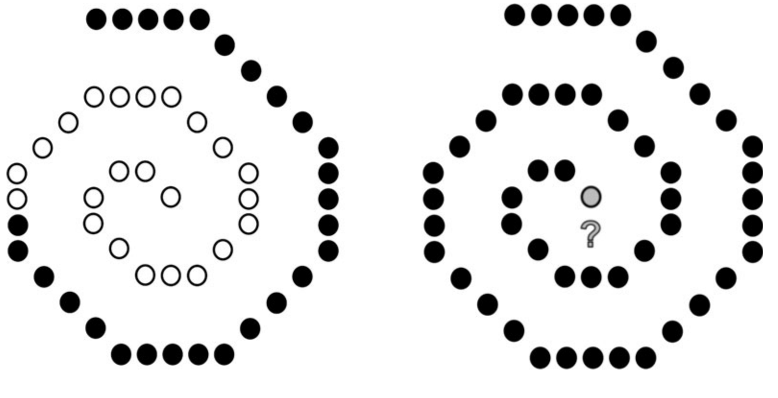

Fig. 1. Beat visualization tool at different stages of a subsequence. The first spiral shows a subsequence being played: The black dot moves along the spiral, synchronized with the beat. The second spiral shows the end of a subsequence: Was the last beat supposed to be a hit or a rest?

(prediction task), and they were subsequently asked to provide a rating on a seven-point scale reflecting the easiness of indicating whether the last beat was supposed to be a hit or a rest (judgment task).

Once the answer sheet was completed for the training sequences and once the experimental procedure was fully understood, the participants were told that they should try to guess whether the last beat was a hit or a rest if they were not sure about their answer. Participants were also told that each subsequence would only be played once. They were then shown the 48 subsequences in a randomized order, with a 5 s pause between each subsequence to leave enough time to complete the answer sheet.

The participants were then given instructions to complete the APM1, with a 10 min time limit, and the self-report questionnaire of the Gold-MSI, with no time limit.

2.4. Results

We excluded one participant from the data analysis because that participant gave strictly alternating responses of "hit" and "rest" for the last 43 subsequences of the prediction task, which we did not consider as a valid effort to guess the last beat of each subsequence.

We conducted Bonferroni-corrected significance tests of the Spearman's correlation coefficients between the complexity values computed with each of the five selected measures for each algorithm. The results are shown in Table 1.

We then averaged the participants' scores across each subsequence for the prediction task (M = 60.3% correct, SD = 17.8%) as well as the judgment task (M = 3.73 easiness rating on a 1–7 scale, SD = 0.83). A significance test of the Pearson's correlation coefficient showed that prediction accuracy and easiness judgments were moderately but significantly correlated across participants ($r(46) = .407$, $p = .004$), demonstrating that the subsequences that were perceived as easier to solve were indeed solved more successfully.

Bonferroni-corrected one-tailed significance tests of the Spearman's correlation coefficients were then conducted between the information-theoretic complexity values of the subsequences and their associated prediction accuracy and easiness judgments, to test the following hypothesis: Subsequence complexity, as measured by the five

Table 1
Correlations between complexity values of generative algorithms

Complexity Measure	Shannon Entropy	Entropy Rate	Excess Entropy	Transient Information	Kolmogorov Complexity
Shannon entropy	1	0.002	0.410	0.260	0.058
Entropy rate	0.002	1	−0.515	−0.400	0.947*
Excess entropy	0.410	−0.515	1	0.926*	−0.326
Transient information	0.260	−0.400	0.926*	1	−0.177
Kolmogorov complexity	0.058	0.947*	−0.326	−0.177	1

Note. $^*p < .001$.

complexity measures, correlates negatively with participants' prediction accuracy as well as easiness judgments. The correlation results are shown in Table 2. Both entropy rate and Kolmogorov complexity were significantly negatively correlated with prediction accuracy and easiness judgments. For both complexity measures, the magnitude of the correlation was greater for the easiness judgments than for the prediction accuracy.

We also aggregated responses at participant level to obtain an overall performance score on the prediction task (M = 60.3% correct, SD = 8.2%). We ran Bonferroni-corrected correlation tests between the participants' performance scores on the prediction task and their scores on the APM1 and three dimensions from the Gold-MSI self-report inventory (Perceptual Abilities, Musical Training and General Sophistication). We found correlations of moderate size between performance scores and the three dimensions of the Gold-MSI, that is, Perceptual Abilities ($r(29) = .439$, $p = .014$), Musical Training ($r(29) = .386$, $p = .032$), and General Sophistication ($r(29) = .314$, $p = .085$), as revealed by significance tests of the Pearson's correlation coefficients. However, possibly due to the small sample size, none of the correlations were significant at the Bonferroni-corrected significance level of $p = .0125$. The correlation between the performance scores and the APM1 was almost zero ($r(29) = .082$, $p = .662$), as revealed by a significance test of the Spearman's correlation coefficient, although it is worth mentioning that a large proportion of the participants obtained the maximal score on the APM1, which could have led to a lack of correlation due to a ceiling effect.

To assess any potential interactions between sequence complexity (as assessed by entropy rate and Kolmogorov complexity, the complexity measures that correlated the most strongly with prediction accuracy) and musical training, we computed binomial mixed effect models with musical training and sequence complexity as fixed effects, participant and generative algorithm as random effects, and subsequence as a nested random effect. The results are presented in Table 3 for entropy rate and in Table 4 for Kolmogorov complexity. As expected, there were significant main effects of both complexity measures. Moreover, there were significant main effects of musical training and significant interactions between musical training and both entropy rate and Kolmogorov complexity.

Table 2
Correlations between complexity values of subsequences and their associated prediction accuracy and easiness judgments

Complexity Measure	Prediction Accuracy	Easiness Judgments
Shannon entropy	0.022	0.166
Entropy rate	−0.407*	−0.834**
Excess entropy	0.208	0.361
Transient information	0.086	0.158
Kolmogorov complexity	−0.402*	−0.866**

Note. *p < .01, **p < .001.

Table 3
Mixed effects model of the influence of musical training and entropy rate on prediction accuracy, with participant, generative algorithm, and subsequence as random effects

	Estimate	*SE*	*z*-value	*p*-value
(Intercept)	0.4541	0.2219	2.047	0.0407
Entropy rate	−0.2842	0.0991	−2.868	0.0041
Musical training	0.1475	0.0614	2.404	0.0162
Entropy rate x Musical training	−0.1432	0.0550	−2.601	0.0093

Note. Subsequences were coded as 1 for *correct answer* and 0 for *incorrect answer*.

Table 4
Mixed effects model of the influence of musical training and Kolmogorov complexity on prediction accuracy, with participant, generative algorithm, and subsequence as random effects

	Estimate	*SE*	*z*-value	*p*-value
(Intercept)	0.4539	0.2221	2.044	0.0409
Kolmogorov complexity	−0.2803	0.0998	−2.808	0.0050
Musical training	0.1470	0.0613	2.399	0.0165
Kolmogorov complexity × Musical training	−0.1267	0.0551	−2.301	0.0214

Note. Subsequences were coded as 1 for *correct answer* and 0 for *incorrect answer*.

3. Discussion

This study assessed the ability of five different general measures of complexity to capture the human judgment of rhythm complexity by comparing formal complexity measurements to prediction accuracy and easiness judgments of human listeners on a novel rhythm perception task. Experimental results showed that prediction accuracy and easiness judgments correlated moderately, indicating that sequences that are perceived as more complex are indeed less predictable. Out of the five assessed complexity measures, only entropy rate and Kolmogorov complexity were significantly correlated with prediction task responses, and with judgment task responses that reflect the subjectively perceived complexity of rhythms. The entropy rate of a sequence can be interpreted as the departure from periodicity. For instance, a periodic sequence that has each symbol randomized with a small probability p has an entropy rate that grows with p. The results therefore suggest that the judgment of rhythm complexity scales with departure from periodicity. It is worth mentioning that entropy rate and Kolmogorov complexity per symbol of the generative algorithms were highly correlated in our study. We observe that these measures are fundamentally measures of the randomness of infinite length sequences; entropy rate and the Kolmogorov complexity per symbol scale from small (ordered and almost periodic sequences) to large (incompressible and random sequences). Entropy rate has not been investigated in a psychological context so far, but the strong correlation of Kolmogorov complexity with easiness judgments confirms suggestions by Shmulevich and Povel (2000). More generally, our results fit within the growing body of research that provides evidence that salient features of music can be captured by formal measures of complexity (Cilibrasi et al., 2004; Essens, 1995; Fitch & Rosenfeld, 2007; Gomez et al., 2007; Hansen & Pearce, 2012; Li & Sleep, 2004, 2005; Madsen & Widmer, 2006; Povel & Essens, 1985; Shmulevich & Povel, 2000).

However, we only found low correlations between participants' responses and excess entropy, transient information, and Shannon entropy. Shannon entropy has previously been found to capture the attention given to a specific melodic line in a piece of music (Madsen & Widmer, 2006) or the uncertainty of probe tones in melodies (Hansen & Pearce, 2012). In comparison with our findings, this might indicate that Shannon entropy may be better suited for capturing the complexity of pitch sequences rather than rhythmic sequences. It is worth remarking that Shannon entropy, $H(1)$, is a function of a probability distribution and, unlike the other measures considered here, is not sensitive to symbol order. We would expect that a suitable measure of rhythm complexity should take into account the relative positions of beats, and not just their probability distribution. The low correlation between participants' responses and excess entropy (which distinguishes sequences of different periodicity) and transient information (which differentiates between sequences of the same period) further indicates that rhythm complexity is perceived as a departure from periodicity, no matter what the periodicity actually is (although sequences of very long periodicity were not included in our study).

Our results also confirm the effect of musical expertise, as suggested by the main effect of the "Musical Training" dimension of the Gold-MSI. "Musical Training" is defined as the "extent of musical training and practice" and "degree of self-assessed musicianship" (Müllensiefen et al., 2014). The significant effect of musical training is consistent with Schmidhuber's (2009) theory, which states that the understanding of a complex stimulus depends on the previous acquisition of strategies to understand similar stimuli. This is also in agreement with the findings reported by Hansen and Pearce (2012), North and Hargreaves (1995), and Vitz (1966) about the effects of musical abilities on the perception of music complexity. Hansen and Pearce (2012) also found an entropy-by-expertise interaction in their results. We identified a similar interaction between musical training and both entropy rate and Kolmogorov complexity in the results of the prediction task, which suggests that domain-specific expertise provides an advantage when dealing with low-randomness stimuli, and becomes detrimental as randomness increases.

As stated above, Kolmogorov complexity and entropy rate both essentially measure the randomness of a sequence. Therefore, a possible interpretation of our results is that sequences that are less random (e.g., rhythmic patterns with a short period length) are easier to process because they can be processed within the limits of human working memory capacity. The working memory model, as proposed by Baddeley and Hitch in 1974 (for a detailed description of auditory working memory, see also Baddeley & Logie, 1992), includes components that are relevant for musical processing. Lee (2004) found evidence for the existence of a specific rhythmic component in working memory, and Jerde, Childs, Handy, Nagode, and Pardo (2011) showed that working memory for rhythm activated different brain areas compared to passive listening of rhythms. Based on these results, it is reasonable to assume that our experimental task of completing rhythmic sequences could indeed be recruiting cognitive processes associated with working memory. This working memory interpretation of our results can also accommodate for the modulatory effect of musical expertise, as there is evidence for increased working memory capacity due to domain-specific expertise (Chase & Ericsson, 1982), and for a relationship between auditory working memory abilities and the extent of musical training (Bailey & Penhune, 2010).

This is one of the several possible interpretations of the connection between the modeling approach we introduced in this study and the participants' mental representations related to the judgment of rhythm complexity. The high correlation coefficients for both Kolmogorov complexity and entropy rate on the prediction task suggest a linear relationship between the randomness of a rhythmic sequence and the human ability to fully process it. While the data from this experiment alone do not allow the identification of the specific cognitive resources or mechanisms involved in such a relationship, the significant interaction between musical training and the formal complexity measures in the mixed effects models allows us to infer that the involved cognitive mechanisms can be modeled more closely with formal complexity measures in the participants with higher musical training. Possible explanations could therefore be related to increased working memory capacity as discussed above, to more structured representations of distributions of transition probabilities for the elements of the auditory sequences, or to differences in attentional patterns or in memorization of longer rhythmic patterns for instance. The precise differences in cognitive processing could be investigated in subsequent research.

It is important to remember that the results of this study rely on certain modeling assumptions. We assumed that prediction accuracy for the last beat of a sequence can serve as a cognitively adequate measure of rhythm complexity. While this assumption is debatable, it seems to receive at least some support by the correlation between prediction accuracy and easiness judgments. Moreover, we assumed that complexity values as computed by the five complexity measures are comparable when computed for large sequences (infinite length complexity or over 10^7 symbols) and for shorter subsequences (50 symbols). This is due to our choice of using some non-periodic sequences and to the difficulty of defining complexity for short, non-periodic, but structured sequences. Finally, we decided to use generative algorithms for the production of the rhythmic sequences in order to obtain a large sample of experimental stimuli within a controlled parameter space. Of course, even though these algorithmic sequences fit the definition of rhythm by being "patterned configuration[s] of attacks that may or may not be constrained overall by a meter or associated with a particular tempo" (Randel, 1986, p. 700), we are fully aware of the fact that they probably lack ecological validity and are only remotely related to rhythmic patterns from real music. However, having established the similar behavior of entropy rate and Kolmogorov complexity compared to human judgments on this set of artificially generated stimuli, an extension of this study could use rhythmic sequences taken from existing music pieces and apply formal complexity measurements in a similar way. A follow-up experiment should also revisit the effect of individual differences in general pattern recognition ability but should aim to avoid the ceiling effect on the APM1 task, for example, by using the second and longer set of the Matrices.

Finally, we acknowledge that the esthetic experience of music certainly involves more aspects than just the complexity of the stimulus, such as the ability to trigger emotions as well as the effects of enculturation, semantic context, stylistic preferences, and many more. However, the main findings of this study that Kolmogorov complexity and entropy rate can suitably measure the perceived complexity of rhythmic patterns offers the possibility to study the esthetic perception of rhythm in a rigorous and quantitative manner which can contribute to our understanding of the cognitive processes that underpin the judgment of beauty in music.

Acknowledgments

The authors thank Dominik Chapman for his valuable input in some aspects of the design of this study, as well as the editor and two anonymous reviewers for their insightful comments on previous versions of this manuscript.

Note

1. We used a sampled tom-tom recording uploaded by the user ‘quartertone’ on freesound, retrieved from http://www.freesound.org/people/quartertone/sounds/129946/

References

Abdallah, S., & Plumbley, M. (2009). Information dynamics: Patterns of expectation and surprise in the perception of music. *Connection Science*, *21*(2–3), 89–117. doi:10.1080/09540090902733756.

Abdallah, S., & Plumbley, M. (2010). Predictive information, multi-information, and binding information. Technical report C4DM-TR10-10. Version 1.1. Retrieved from http://www.eecs.qmul.ac.uk/~markp/2010/AbdallaPlumbley10-tr1010-1.1.pdf [2015, November 16].

Abdallah, S., & Plumbley, M. (2012). A measure of statistical complexity based on predictive information with application to finite spin systems. *Physical Letters A*, *376*(4), 275–281. doi:10.1016/j.physleta.2011.10.066.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (vol. 8, pp. 47–89). New York: Academic Press. doi:10.1016/s0079-7421(08)60452-1.

Baddeley, A. D., & Logie, R. H. (1992). Auditory imagery and working memory. In D. Reisberg (Ed.), *Auditory imagery* (pp. 179–197). Hillsdale, NJ: Erlbaum.

Bailey, J. A., & Penhune, V. B. (2010). Rhythm synchronization performance and auditory working memory in early- and late-trained musicians. *Experimental Brain Research*, *204*(1), 91–101. doi:10.1007/s00221-010-2299-y.

Berlyne, D. E. (1970). Novelty, complexity, and hedonic value. *Perception & Psychophysics*, *8*(5A), 279–286. doi:10.3758/bf03212593.

Bialek, W., Nemenman, I., & Tishby, N. (2001). Predictability, complexity, and learning. *Neural Computation*, *13*, 2409–2463. doi:10.1162/089976601753195969.

Chase, W. G., & Ericsson, K. A. (1982). Skill and working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (vol. 16, pp. 1–58). New York: Academic Press. doi:10.1016/s0079-7421(08)60546-0.

Cilibrasi, R., Vitanyi, P., & de Wolf, R. (2004). Algorithmic clustering of music based on string compression. *Computer Music Journal*, *28*(4), 49–67. doi:10.1162/0148926042728449.

Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

Crutchfield, J. P., & Feldman, D. P. (2003). Regularities unseen, randomness observed: Levels of entropy convergence. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *13*(1), 25–54. doi:10.1063/1.1530990.

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

Fitch, W. T., & Rosenfeld, A. J. (2007). Perception and production of syncopated rhythms. *Music Perception*, *25*(1), 43–58. doi:10.1525/mp.2007.25.1.43.

Gomez, F., Thul, E., & Toussaint, G. (2007). An experimental comparison of formal measures of rhythmic syncopation. In *Proceedings of the International Computer Music Conference* (pp. 101–104). Ann Arbor, MI: MPublishing.

Hansen, N. C., & Pearce, M. T. (2012). Shannon entropy predicts perceptual uncertainty in the generation of melodic pitch expectations. In E. Cambouropoulos, C. Tsougras, P. Mavromatis, & K. Pastiadis (Eds.), *Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music* (pp. 406–407). Thessaloniki, Greece: Aristotle University of Thessaloniki.

Heyduk, R. G. (1975). Rated preference for musical compositions as it relates to complexity and exposure frequency. *Perception & Psychophysics*, *17*(1), 84–90. doi:10.3758/bf03204003.

Jerde, T. A., Childs, S. K., Handy, S. T., Nagode, J. C., & Pardo, J. V. (2011). Dissociable systems of working memory for rhythm and melody. *NeuroImage*, *57*(4), 1572–1579. doi:10.1016/j.neuroimage.2011.05.061.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS ONE*, *7*(5), e36399. doi:10.1371/journal.pone.0036399.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2014). The Goldilocks effect in infant auditory attention. *Child Development*, *85*(5), 1795–1804. doi:doi: 10.1111/cdev.12263.

Kolmogorov, A. N. (1965). Three approaches to the quantitative definition of information. *Problems of Information Transmission*, *1*(1), 1–7.

Lee, K. M. (2004). Rhythmic working memory and imagery ability. In S. D. Lipscomb, R. Ashley, R. O. Gjerdingen & P. Webster (Eds.), *Proceedings of the 8th International Conference on Music Perception & Cognition* (pp. 228–230). Adelaide, Australia: Causal Productions.

Li, M., & Sleep, R. (2004). Melody classification using a similarity metric based on Kolmogorov complexity. In E. Saint-James, G. Assayag, C. Agon & M. Chemillier (Eds.) *Proceedings of the Sound and Music Computing Conference* (pp. 126–129). Barcelona, Spain: SMCNetwork.

Li, M., & Sleep, R. (2005). Genre classification via an LZ78-based string kernel. In T. Crawford, & M. Sandler (Eds.) *Proceedings of the 6th International Conference on Music Information Retrieval* (pp. 252–259). London: Queen Mary, University of London.

Li, M., & Vitanyi, P. (2008). *An introduction to Kolmogorov complexity and its applications* (3rd ed.). New York, NY: Springer. doi:10.1007/978-0-387-49820-1.

Madsen, S. T., & Widmer, G. (2006). Music complexity measures predicting the listening experience. In M. Baroni, A. R. Addessi, R. Caterina, & M. Costa (Eds.), *Proceedings of the 9th International Conference on Music Perception & Cognition* (pp. 1812–1816). Bologna, Italy: University of Bologna.

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS ONE*, *9*(2), e89642. doi:10.1371/journal.pone.0089642.

North, A. C., & Hargreaves, D. J. (1995). Subjective complexity, familiarity, and liking for popular music. *Psychomusicology*, *14*, 77–93. doi:10.1037/h0094090.

Parncutt, R. (1994). A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Perception*, *11*(4), 409–464. doi:10.2307/40285633.

Povel, D. J., & Essens, P. (1985). Perception of temporal patterns. *Music Perception*, *2*(4), 411–440. doi:10.2307/40285311.

Randel, D. M. (Ed.) (1986). *The New Harvard Dictionary of Music. Cambridge, MA, & London*. United Kingdom: The Belknap Press of Harvard University Press.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's progressive matrices and vocabulary scales. Section 4: The advanced progressive matrices*. San Antonio, TX: Harcourt Assessment.

Schmidhuber, J. (2009). Simple algorithmic theory of subjective beauty, novelty, surprise, interestingness, attention, curiosity, creativity, art, science, music, jokes. *Journal of SICE*, *48*(1), 21–32.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x.

Shmulevich, I., & Povel, D. J. (2000). Measures of temporal pattern complexity. *Journal of New Music Research*, *29*(1), 61–69. doi:10.1076/0929-8215(200003)29:01;1-p;ft061.

Vitz, P. C. (1966). Affect as a function of stimulus variation. *Journal of Experimental Psychology*, *71*(1), 74–79. doi:10.1037/h0022619.

Welch, T. A. (1984). A technique for high-performance data compression. *Computer*, *17*(6), 8–19. doi:10.1109/mc.1984.1659158.

Wilson, S., Cottle, D., & Collins, N. (2011). *The SuperCollider book*. Cambridge, MA: MIT Press.

Wundt, W. M. (1896). *Grundriss der Psychologie*. Leipzig, Germany: W. Engelmann.

Ziv, J., & Lempel, A. (1978). Compression of individual sequences via variable-rate coding. *IEEE Transactions on Information Theory*, *24*(5), 530–536. doi:10.1109/tit.1978.1055934.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Generative algorithms and associated complexity values.

Appendix S2. Subsequences and answer key.

Appendix S3. Example of calculations of complexity values.