Perceptual Validity of Information-Theoretic Measures 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*

Correspondence should be sent to Remi de Fleurian. E-mail: r.defleurian@gmail.com

Abstract

In order to identify a perceptually valid measure 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 implicit and explicit complexity 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 was correlated with implicit and explicit judgments, Kolmogorov complexity was highly correlated with explicit judgments, and scores on the implicit task were correlated with self-assessed musical perceptual abilities. A logistic regression showed main effects of entropy rate and musical training, and an interaction between entropy rate and musical training. These results indicate that information-theoretic concepts capture some salient features of human rhythm perception, and confirm the influence of musical expertise in the perception of rhythm complexity.

Keywords: Rhythm Perception; Rhythm Complexity; Information Theory Measures; 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 aesthetic 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 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 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 paper focused on identifying a perceptually valid measure of rhythm complexity, as opposed to the subjective or arbitrary measures presented so far. There were several possible measures of complexity to consider. Some of them have been 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 once again 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 perceptual validity as a measure of human rhythm complexity by Shmulevich and Povel (2000). The measure didn't perform well, but the authors attribute this to the short length of the sequences used in their experiment, and suggest 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 in order 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).

The relationship between data compression and aesthetics is supported by Schmidhuber's (2009) theoretical model. He compares the human mind to a self-improving, computationally limited observer, and approaches the question of complexity from an algorithmic point of view: he states that beauty comes from the challenge of discovering patterns and the ability to compress new data, i.e. a moderately complex stimulus will be perceived as beautiful because the strategies developed to understand that stimulus will 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. Moreover, he argues that the perceived complexity of a stimulus changes with exposure, in accordance with Berlyne's (1970) and Heyduk's (1975) findings.

For the purpose of this study, we selected five measures with a potential to serve as perceptually valid measures of rhythm complexity: Shannon entropy, entropy rate, excess entropy, transient information, and Kolmogorov complexity. These measures have firm theoretical foundations and are defined at a high level of abstraction and generality, i.e. 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 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 sub-sequences 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 $H(L) \to H_\infty = hL + E$ 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 proportional to the logarithm of the period of a repetitive sequence.

The final measure that we consider in this paper, 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 summary, the substring entropies $H(1)$, $H(2)$, … represent increasingly long range symbol correlations. The entire set of substring entropies can be characterized by three quantities h, E and T.

We used these five measures in an experiment aimed at capturing implicit (or objective) rhythm complexity, as well as explicit (or subjective) rhythm complexity, and we investigated two suspected modulatory effects of rhythm perception: musical expertise (Hansen & Pearce, 2012; North & Hargreaves, 1995; Vitz, 1966) and pattern identification ability (Schmidhuber, 2009). In the following, we used a broad meaning for 'rhythmic sequence', characterizing the general enfolding of binary events in time.

2. Method

2.1. Participants

32 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 Appendix A 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 randomly extracted 3 sub-sequences of 50 symbols from each sequence, with the following restriction: the sub-sequences 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 sub-sequences and kept in a separate file in order to be used as the answer key for the implicit task. 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 150bpm. Each sub-sequence was therefore almost 20 seconds long. Extracting three sub-sequences 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, in order to accustom the participants to the experimental procedure.

Due to the nature of the task, which required accurate following of the beat sub-sequences, we also provided the participants with a visual representation of the sub-sequences. We designed a beat visualization tool using PowerPoint and iMovie (Fig. 1), which consisted of a spiral made of 50 white dots. When a sub-sequence 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 in order to make sure the participants knew exactly which beat they needed to provide a judgment for.

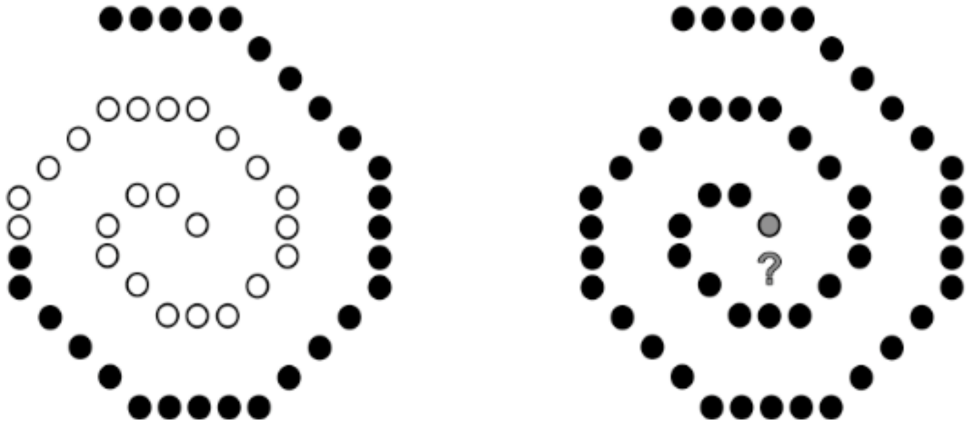

Fig. 1. Beat visualization tool at different stages of a sub-sequence. The first spiral shows a sub-sequence being played: the black dot moves along the spiral, synchronized with the beat. The second spiral shows the end of a sub-sequence: was the last beat supposed to be a hit or a rest?

We used the self-report questionnaire of the Goldsmiths Musical Sophistication Index, or Gold-MSI (Müllensiefen, Gingras, Musil, & Stewart, 2014), in order 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 the purpose of 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 (J. Raven, J. C. Raven, & Court, 1998).

[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/

2.3. Procedure

Each participant was tested individually in a quiet lab environment. The participants were first shown the 2 training sequences as many times as they wanted, and were given instructions on how to complete the answer sheet. They were requested, for each sub-sequence, to indicate whether the last beat was supposed to be a hit or a rest (implicit task), and were subsequently asked to provide a rating on a 7-point scale reflecting the easiness of indicating whether the last beat was supposed to be a hit or a rest (explicit 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 sub-sequence would only be played once. They were then shown the 48 sub-sequences in a randomized order, with a 5 seconds pause between each sub-sequence in order to leave enough time to complete the answer sheet.

The participants were then given instructions to complete the APM1, with a 10 minutes 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 sub-sequences of the implicit task, which we did not consider as a valid effort to guess the last beat of each sub-sequence.

For three of the generative algorithms, the three extracted sub-sequences differed significantly from each other in terms of the participants’ scores, as revealed by a Bonferroni-corrected Chi-Square test. Hence for these three algorithms, the extracted sub-sequences did not seem to represent a homogeneous sample in terms of their perceptual complexity. The sub-sequences from the following generative algorithms were therefore excluded from the data analysis: P12.88 ($\chi^2(2, N = 93) = 14.2, p = .001$, EVEN ($\chi^2(2, N = 93) = 19.2, p < .001$), and QP-random ($\chi^2(2, N = 93) = 15.7, p < .001$).

For the remaining 13 generative algorithms, we conducted Bonferroni-corrected correlation tests between the complexity values computed with each of the five selected measures. The results are shown in Table 1. We then averaged the participants’ scores across each sub-sequence for the implicit task ($M = 60.8\%$ correct, $SD = 16.3\%$) as well as the explicit task ($M = 3.76$ easiness rating on a 1-7 scale, $SD = 0.77$). Implicit and explicit scores were moderately but significantly correlated across participants ($r(37) = .457$, $p = .003$), demonstrating that the sub-sequences that were perceived as easier to solve were indeed solved more successfully.

We then averaged the participants’ scores across each generative algorithm for the implicit task as well the explicit task. Bonferroni-corrected one-tailed correlation tests were conducted between the information-theoretic complexity values of the algorithms and their implicit and explicit scores, to test the following hypothesis: Sequence complexity, as measured by the five complexity measures, correlates negatively with participants’ implicit scores as well as explicit scores. The correlation results are shown in Table 2. Both entropy rate and Kolmogorov complexity are significantly negatively correlated with explicit scores, and entropy rate is also negatively correlated with implicit scores. For both complexity measures, the magnitude of the correlation is greater for the explicit scores than for the implicit scores.

Table 1

Correlations between Complexity Values of Generative Algorithms

Complexity measure	Shannon entropy	Entropy rate	Excess entropy	Transient information	Kolmogorov complexity
Shannon entropy	1	.207	.321	.419	.272
Entropy rate	.207	1	-.347	-.388	.929**
Excess entropy	.321	-.347	1	.861**	-.088
Transient information	.419	-.388	.861**	1	-.096
Kolmogorov complexity	.272	.929**	-.088	-.096	1

* $p < .005$. ** $p < .001$.

Table 2

Correlations between Complexity Values of Generative Algorithms and their Implicit and Explicit Complexity

Complexity measure	Implicit scores	Explicit scores
Shannon entropy	-.271	-.198
Entropy rate	-.670*	-.909**
Excess entropy	.132	.030
Transient information	.097	.110
Kolmogorov complexity	-.592	-.973**

* $p < .01$. ** $p < .001$.

We also aggregated responses at participant level to obtain an overall performance score on the implicit task. All participants with an overall performance score below chance level were assigned the chance level score (50%). We ran Bonferroni-corrected correlation tests between the participants' performance scores on the implicit task and their scores on the APM1 and three subscales from the Gold-MSI self-report inventory (Perceptual Abilities, Musical Training and General Sophistication). The participants with a higher score on the Perceptual Abilities subscale performed better on the implicit task ($r(29) = .478$, $p = .007$). There was no significant correlation with Musical Training ($r(29) = .371$, $p = .040$), General Sophistication ($r(29) = .266$, $p = .149$), or with the scores on the APM1 ($r(29) = .135$, $p = .467$), 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.

In order to assess any potential interactions between sequence complexity (as assessed by entropy rate, the complexity measure that correlates the most strongly with implicit scores) and musical training, we computed a binomial mixed effect model with musical training and entropy rate as fixed effects, participant and generative algorithm as random effects, and sub-sequence as a nested random effect. The results are presented in Table 3. As expected, there was a significant main effect of entropy rate. Moreover, there was a significant main effect of musical training and a significant interaction between entropy rate and musical training. Similar results were obtained in a model using Kolmogorov complexity instead of entropy rate (Appendix B).

Table 3

Mixed Effects Model of the Influence of Musical Training and Entropy Rate on Implicit Task Performance, with Participant, Generative Algorithm and Sub-Sequence as Random Effects.

	Estimate	*SE*	*z*-value	*p*-value
(Intercept)	0.4840	0.1213	3.990	< .001
Entropy rate	-0.2639	0.1151	-2.293	.0219
Musical training	0.1493	0.0721	2.070	.0384
Entropy rate x Musical training	-0.1490	0.0609	-2.445	.0145

Note. Sub-sequences were coded as 1 for *correct answer* and 0 for *incorrect answer.*

3. Discussion

This study assessed the perceptual validity of five different general measures of complexity by comparing formal complexity measurements to explicit and implicit behavioral responses of human listeners on a novel rhythm perception task. Experimental results showed that implicit and explicit behavioral responses 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 explicit responses that reflect the subjectively perceived complexity of rhythms. In addition, only entropy rate was significantly correlated with the participants' implicit task responses. 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 perception 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 explicit responses 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 valid 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 that effects of musical expertise, as suggested by the correlation of the 'Perceptual Abilities' dimension of the Gold-MSI with participants' scores on the implicit task, and by the main effect of 'Musical Training' dimension of the Gold-MSI. 'Perceptual Abilities' is defined as the "self-assessment of a cognitive musical ability […] related to music listening skills", and 'Musical Training' is defined as the "extent of musical training and practice" and "degree of self-assessed musicianship" (Müllensiefen et al., 2014). The significant effects of individual perceptual ability and musical training are in line 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 entropy rate and musical training in the results of the implicit 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 (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).

However, 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 implicit measure of rhythm complexity. While this assumption is debatable, it seems to receive at least some support by the correlation between implicit and explicit scores. 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 sub-sequences (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. We tried to mitigate the effects of the instances where this assumption does not hold by discarding sub-sequences from algorithms that yielded inconsistent scores on the implicit task. Finally, we decided to use generative algorithms for the production of the rhythmic sequences in order obtain a large sample of experimental stimuli within a controlled parameter space. Of course, we are fully aware of the fact that the algorithmically generated sequences 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 aesthetic experience of music certainly involves more aspects than just the complexity of the stimulus, such as 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 are perceptually valid measures of rhythm complexity offers the possibility to study the aesthetic perception of rhythm in a rigorous and quantitative manner which can contribute to our understanding of the cognitive processes that underpin the judgment of the beauty in music.

Acknowledgements

The authors thank Dominik Chapman for his valuable input in some aspects of the design of this study.

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.

Abdallah, S., & Plumbley, M. (2010, December 9). *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

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.

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, NY: Academic Press.

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.

Berlyne, D. E. (1970). Novelty, complexity, and hedonic Value. *Perception & Psychophysics, 8*(5A), 279-286.

Bialek, W., Nemenman, I., & Tishby, N. (2001). Predictability, complexity, and learning. *Neural Computation, 13*, 2409-2463.

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, NY: Academic Press.

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

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.

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

Fitch, W. T., & Rosenfeld, A. J. (2007). Perception and production of syncopated rhythms. *Music Perception, 25*(1), 43-58.

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.

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.

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 *Proceedings of the Sound and Music Computing Conference*. Paris, France.

Li, M., & Sleep, R. (2005). Genre classification via an LZ78-based string kernel. In *Proceedings of the 6th International Conference on Music Information Retrieval* (pp. 252-259). London, UK.

Li, M., & Vitanyi, P. (2008). *An introduction to Kolmogorov complexity and its applications* (3rd ed.). New York, NY: Springer.

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.

Povel, D. J., & Essens, P. (1985). Perception of temporal patterns. *Music Perception, 2*(4), 411-440.

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.

Shmulevich, I., & Povel, D. J. (2000). Measures of temporal pattern complexity. *Journal of New Music Research, 29*(1)*,* 61-69.

Vitz, P. C. (1966). Affect as a function of stimulus variation. *Journal of Experimental Psychology, 71*(1), 74-79.

Welch, T. A. (1984). A technique for high-performance data compression. *Computer, 17*(6), 8-19.

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.

Appendix A

Generative Algorithms and Associated Complexity Values

A very large number of generative sequences have been defined and examined in the mathematical and scientific literature (the Online Encyclopedia of Integer Sequences,[2] or OEIS, has over 220000 entries). We devised four broad groupings and collected, as far as possible, representative sequences which have known *H*, *h*, *E*, and *T*. We computed our own upper bound estimates to the Kolmogorov complexity by compressing sequences of length 10^7 with the Lempel-Ziv 1978 algorithm and counting the number of codewords per symbol (Li & Sleep, 2004). We computed *h*, *E* and *T* wherever values were not available by analyzing generative sequences of 10^7 symbols. The sequences were all chosen for ease of generation; the algorithms in each case amounted to only a few lines of Java code.

The groupings reflect the means of generation and overall statistical properties of the sequences. We have deliberately avoided any sequence of musical origin; our sequences are purely abstract entities.

1. Deterministic-Periodic

These sequences are repetitions of finite length substrings. They are generated by deterministic algorithms. We chose one sequence of period 5 (P5c: (10101)*), three sequences of period 12 (P12.165: (001010010101)*, P12.88: (000010111101)*, P12.116: (000100100101)*) and a single sequence of period 16 (P16: (1010111011101110)*). The sequences of period 12 were chosen to have entropy close to 1 (approximately equal numbers of 0's and 1's) and to have *T* values that spanned the possible range. The complexities of all inequivalent sequences of period 12 were computed; two sequences of period p are equivalent if they cannot be decomposed into sequences of smaller period, if they are identical under interchange of 0 and 1 or if one is a cyclic permutation of the other. The sequence complexities are given in Table A1. The entropy rates are zero as expected, and the *K*-complexities are low (0.01 - 0.0178) since these sequences are easily compressed.

2. Stochastic

This group contains sequences where each symbol is drawn independently from a probability distribution. Two sequences were produced using a pseudorandom number generator to produce 0's and 1's with probability 0.5 of either (Bernoulli-0.5 sequence), and with probability 0.7 of one symbol and 0.3 of the other (Bernoulli-0.7 sequences). Table A1 shows the complexity values. In this case, *E* and *T* are zero because the sequences are produced from sources without memory (symbols do not depend on the past) and because there is nothing to synchronize to. The *K*-complexities are high (0.0514 - 0.0574) because the sequences, ideally, are incompressible.

3. Deterministic-Stochastic

These processes mix a deterministic rule with a stochastic one. The 'golden mean' process (GM) produces strings with no consecutive 0's. A 0 or a 1 is generated with probability ½. The next symbol is certainly a 1 if a 0 was generated; otherwise a 0 or 1 is again generated with equal probabilities. The RRXOR process proceeds as follows: two Bernoulli-0.5 generations are immediately followed by their XOR i.e. sub-sequences 000,

[2] Accessible at: http://oeis.org/

011, 101, 110 occur with equal probability. The even process generates strings with even numbers of 1's bounded by any number of 0's. A 0 or a pair of 1's is generated with even probability followed by a Bernoulli-0.5 trial. A full account of these sequences can be found in Crutchfield and Feldman (2003). Table A1 tabulates the complexities. None of the measures is zero, indicating the mix of stochastic and deterministic elements. The sequences are fairly incompressible with high K-complexities (0.0415 - 0.0469).

4. Deterministic-Aperiodic

These sequences are deterministic but are not strictly periodic. The Thue-Morse process (Crutchfield & Feldman, 2003) is produced as follows. The sequence starts with 01. The complement of 01 is then added to produce 01 10. This is again followed by the complement to produce 0110 1001. The procedure is iterated.

The Fibonacci Word sequence (sequence A003849 of OEIS), or FW, is produced by adding the sequences at iteration n and at iteration $n + 1$ to produce the sequence at iteration $n + 2$. It begins S(0) = 0, S(1) = 01, S(2) = S(0)S(1) = 0 + 01 = 001 and continues S(3) = 01 + 001 = 01001.

The quasi-periodic linear sequence (Tao, 2006), or QP-linear, is formed by the linearly structured set

$$L = \{n\colon [\alpha n] < \delta\} \tag{5}$$

where [x] denotes the fractional part of x and, for this experiment, $\alpha = \sqrt{7}$ and $\delta = 0.5$. The n'th term of the sequence is 1 if n is in L and 0 otherwise. The resulting sequence is almost periodic since symbols at n and $n + L$ are correlated by virtue of the identity

$$[\alpha(n + L)] - [\alpha n] = [\alpha L] \tag{6}$$

The quasi-periodic quadratic sequence (QP-quad) is formed in a similar way to QP-linear. The quadratically structured set is

$$Q = \{n\colon [\alpha n\char`^2] < \delta\} \tag{7}$$

and the n'th term is 1 if n is in Q. The sequence is more random than QP-linear but has some vestige of periodicity.

The quasi-periodic random sequence (QP-rand) has even more stochasticity; it is formed in a similar way to QP-quad except values of n in Q are further subject to a test. They are included in R with probability δ'. The definition is

$$Q = \{n\colon [\alpha n^2] < \delta \ AND \ U(0,1)\} < \delta' \tag{8}$$

where $U(0, 1)$ is the uniform distribution on [0, 1] and, for this experiment, $\delta' = 0.5$.

The increasing randomness of the quasi-periodic sequences is reflected by the increasing h and K values. The very high T complexities indicate the unpredictable nature of these sequences.

Finally, sequence W is designed to be particularly challenging for the LZ78 algorithm (Shor, 2005). It contains all possible codewords at each length. The sequence starts 0|1, and continues 0|1|00|01|10|11. Its empirical K-complexity as computed by LZ78 compressibility (Table A1) is even higher than the incompressible Bernoulli-generated sequences. This result is presumably due to the finite length of the sequences and the imperfections of the pseudorandom number generator.

Table A1

Complexity Values of Generative Algorithms used in the Trials

Group	Algorithm	*H*	*h*	*E*	*T*	*K*
1	P5c	.971	0[a]	0.232[a]	4.87[a]	.00100
	P12.165	.980	0	3.59	14.8	.00154
	P12.88	1	0	3.59	8.42	.00155
	P12.116	.918	0	3.59	11.7	.00154
	P16	.896	0	4	16.6	.00178
2	B-0.5	1	1	0	0	.0574
	B-0.7	.881	.881	0	0	.0514
3	GM	.918	.667[a]	0.252[a]	0.252[a]	.0415
	RRXOR	1	.667[a]	2[a]	9.43[a]	.0469
	EVEN	.918	.667[a]	0.902[a]	3.03[a]	.0432
4	TM	1	.083[a]	4.168	16.4	.00674
	QP-linear	1	.019	5.14	33.2	.00488
	QP-quadratic	.970	.023	3.7	26.2	.0121
	QP-random	.877	.650	1.52	10.2	.0417
	W	1	.693	5.59	53.84	.0582
	FW	.960	.087	2.63	9.04	.00474

[a]These values of *h*, *E* and *T* were taken from Crutchfield and Feldman (2003).

References

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.

Li, M., & Sleep, R. (2004). Melody classification using a similarity metric based on Kolmogorov complexity. In *Proceedings of the Sound and Music Computing Conference*. Paris, France.

Shor, P. (2005). *Lempel-Ziv* [PDF document]. Retrieved from lecture notes online website: http://www-math.mit.edu/~shor/PAM/lempel_ziv_notes.pdf

Tao T. (2006). The dichotomy between structure and randomness, arithmetic progression, and the primes. In M. Sanz-Solé, J. Soria, J. L. Varona, & J. Verdera (Eds.), *Proceedings of the International Congress of Mathematicians* (pp. 581-608). Madrid, Spain: EMS Publishing House.

Appendix B

Table B1

Mixed Effects Model of the Influence of Musical Training and Kolmogorov Complexity on Implicit Task Performance, with Participant, Generative Algorithm and Sub-Sequence as Random Effects.

	Estimate	*SE*	*z*-value	*p*-value
(Intercept)	0.4839	0.1226	3.949	< .001
Kolmogorov complexity	-0.2592	0.1164	-2.228	.0259
Musical training	0.1490	0.0726	2.053	.0401
Kolmogorov complexity x Musical training	-0.1289	0.0617	-2.089	.0368

Note. Sub-sequences were coded as 1 for *correct answer* and 0 for *incorrect answer.*