

A Stimulus Set of 40 Popular Music Drum Patterns with Perceived Complexity Measures

Senn, Olivier; Hoesl, Florian; Jerjen, Rafael; Bechtold, Toni Amadeus; Kilchenmann, Lorenz; Rose, Dawn; Alessandri, Elena

DOI:

[10.1177/20592043231202576](https://doi.org/10.1177/20592043231202576)

License:

Creative Commons: Attribution (CC BY)

Document Version

Publisher's PDF, also known as Version of record

Citation for published version (Harvard):

Senn, O, Hoesl, F, Jerjen, R, Bechtold, TA, Kilchenmann, L, Rose, D & Alessandri, E 2023, 'A Stimulus Set of 40 Popular Music Drum Patterns with Perceived Complexity Measures', *Music & Science*, vol. 6, pp. 1-17.
<https://doi.org/10.1177/20592043231202576>

[Link to publication on Research at Birmingham portal](#)

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

A Stimulus Set of 40 Popular Music Drum Patterns with Perceived Complexity Measures

Olivier Senn¹ , Florian Hoesl¹ , Rafael Jerjen¹,
Toni Amadeus Bechtold^{1,2} , Lorenz Kilchenmann¹ ,
Dawn Rose¹  and Elena Alessandri¹ 

Abstract

This study presents an audio stimulus set of 40 drum patterns from Western popular music with empirical measurements of perceived complexity. The audio stimuli are meticulous reconstructions of drum patterns found in commercial recordings; they are based on careful transcriptions (carried out by professional musicians), drum stroke loudness information, and highly precise onset timing measurements. The 40 stimuli are a subset selected from a previously published larger corpus of reconstructed Western popular music drum patterns (Lucerne Groove Research Library). The patterns were selected according to two criteria: a) they only feature the bass drum, snare drum, and one or more cymbals, and b) they plausibly cover the complexity range of the corpus. Perceived stimulus complexity was measured in a listening experiment using a pairwise comparison design with 220 participants (4,400 trials). In each trial, participants were presented with two stimuli, and they stated which of the two sounded more complex to them. The comparison data then served to calculate complexity estimates using the Bradley–Terry probability model. The complexity estimates have an intuitive interpretation: they allow calculation of the probability that one pattern is considered more complex than another pattern in a pairwise comparison. To our knowledge, this is the first set of naturalistic music stimuli with meaningful perceived complexity estimates. The drum pattern stimuli and complexity measurements can be used for listening experiments in music psychology. The stimuli will further allow measures and models of drum pattern complexity to be assessed.

Keywords

Complexity, drumming, music perception, popular music, rhythm, stimuli

Submission date: 2 June 2023; Acceptance date: 5 September 2023

Introduction

The complexity of musical stimuli and its psychological effects on listeners (such as listeners' aesthetic appreciation or interest in the music, their groove experience, and many other kinds of music-evoked emotions) are a wide field of inquiry. Complexity has been shown to affect humans' appreciation of artifacts such as pictures (Nadal et al., 2010; Osborne & Farley, 1970; Vitz, 1966), narratives (Carney et al., 2014; Stokmans, 2003), music (Chmiel & Schubert, 2019), buildings (Imamoglu, 2000), advertisements (Cox & Cox, 1988), and websites (Mai et al., 2014; Nadkarni & Gupta, 2007; for a general overview,

see Mihelač and Povh, 2020; Van Geert & Wagemans, 2020). A popular theory posited by Berlyne (1963, 1971) claims that humans' preferences form an inverted-U function (also called a Wundt curve, see Wundt, 1874, p. 432)

¹ School of Music, Lucerne University of Applied Sciences and Arts, Lucerne, Switzerland

² Department of Music, University of Birmingham, Birmingham, UK

Corresponding author:

Olivier Senn, School of Music, Lucerne University of Applied Sciences and Arts, Lucerne, Switzerland.

Email: olivier.senn@hslu.ch



Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page

(<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

of stimulus complexity. According to this theory, humans prefer percepts of moderate complexity: the percepts need to be complex enough to be interesting, but not so complex as to be confusing. Studies in empirical music aesthetics have been equivocal in their findings: some provided evidence to support the inverted-U hypothesis with respect to musical stimuli (Beauvois, 2007; Heyduk, 1975; North & Hargreaves 1995, 1996; Steck & Machotka, 1975), whereas others failed to find evidence to support the theory or were inconclusive (Eisenberg & Thompson, 2003; Marin & Leder, 2013; Orr & Ohlsson, 2001, 2005; Russell, 1982; Vitz, 1964).

Besides music appreciation, the inverted-U hypothesis has also become relevant to the study of musical groove, whereby groove is defined as a pleasurable urge to move in response to music (Janata et al., 2012; Senn et al., 2020). Witek et al. (2014) found that popular music drum patterns with intermediate rhythmic complexity evoked a stronger groove experience than drum patterns with low or high rhythmic complexity. This finding was confirmed by several later studies (Cameron et al., 2023; Chmiel & Schubert, 2019; Matthews et al., 2019, 2020; Morillon & Zalta, 2021; Stupacher et al., 2020). A recent theory on the connection between complexity and groove hypothesizes that listeners use repetitive body movement to clarify metric uncertainties triggered by rhythmically complex stimuli. They “fill in the gaps” (Witek, 2017), replacing missing beats in the music by body movements such as foot or hand taps (Spiech et al., 2022; Stupacher et al., 2022). This theory suggests that the groove experience and associated body movement is essentially an embodied mechanism to clarify the meter underlying a complex musical rhythm.

To our knowledge, the concept of musical complexity has not yet been comprehensively defined in music psychology or in music scholarship. In the present study, we propose to approach musical complexity from a pragmatic and empirical point of view: The approach relies on the availability of one set of naturalistic musical stimuli that cover a wide complexity range and that are associated with reliable complexity measurements. This kind of stimulus set allows one to investigate which stimulus properties correlate with the complexity measures. Thus it contributes to theorizing musical complexity on an empirical basis and potentially to a definition of the concept.

Measuring the complexity of stimuli, however, is not a trivial problem. Past studies on music and complexity used either objective (i.e., stimulus-based) or subjective (i.e., perception-based) methods to assess the complexity of musical stimuli. Objective methods estimate complexity using structural properties of the stimuli themselves. Existing objective measures of rhythmic complexity are based on syncopation (Keith, 1991; Longuet-Higgins & Lee, 1984; Witek et al., 2014, 2015), offbeatness (Gómez et al., 2007), metric complexity (Toussaint, 2002), or weighted note-to-beat distance (Gómez et al., 2005). These approaches link rhythmic complexity to discrepancies between the *rhythm*

of the music (which is the organization of sound events in time) and the *meter* (which is the implied background of discrete regular time units against which the rhythm is interpreted; London, 2004). Another objective measure, rhythmic oddity (Chemillier, 2002, p. 176), relies on the unequal or equal division of overarching periodic time units (which may be governed by a meter or not).

In a number of studies, researchers have used approaches from information theory to develop objective measures of rhythmic complexity, such as Shannon entropy, entropy rate, excess entropy, transient information, or Kolmogorov complexity (De Fleurian et al., 2017; Thul & Toussaint, 2008; Witek et al., 2014). These methods represent rhythm as a discrete series of symbols and quantify the information content within this series. If the information content is high (or if redundancy is low), then the rhythm is understood to be complex.

Subjective measures of musical complexity are based on the notion that complexity is not a property of the stimulus itself, but “intrinsically related to how listeners perceive music” (Eerola, 2016, p. 2f.), which in itself is thought to be highly enculturated (Trainor et al., 2012). Pressing (1999) formulated the idea that subjectively perceived rhythmic complexity is linked to the “computational cost” (p. 2), which is the effort that the individual needs to make to relate a concrete rhythm to the underlying meter. With this idea, Pressing anticipated the *predictive coding* approach to rhythm perception, which centers on the listener’s effort to adapt an inner model to correctly predict how a rhythm continues at any moment (Lumaca et al., 2019; Senn, 2023; Vuust et al., 2009; Vuust & Witek, 2014).

From a subjective perspective, the perceiver is the ultimate judge of stimulus complexity. This is also the stance adopted in this study: We assume that stimulus complexity needs to be determined empirically in a listening experiment with human judges whereby a musical stimulus is complex, when listeners perceive it as such. Further, we assume that objective models of stimulus complexity based on the properties of the stimuli themselves are most useful if they reliably predict the subjectively measured complexity of the stimuli.

In several studies, researchers measured stimulus complexity subjectively using Likert-type rating scales: Heyduk (1975) let participants rate the complexity of four short piano pieces. Russell (1982) asked participants to judge the complexity of 40 audio excerpts from modern jazz. In North and Hargreaves’ (1995) study, participants rated the complexity of 60 short extracts from popular music tracks. Shmulevich and Povel (2000) let participants judge the complexity of 35 homophonous rhythms, each of them consisting of 9 rhythmic events. Orr and Ohlsson (2005) asked participants to rate the complexity of 40 jazz and 40 bluegrass improvisations. Each of these stimulus sets has properties that limit their usability: The Heyduk (1975) set is very small (four stimuli). The Russell (1982), North and Hargreaves (1995), and Orr and Ohlsson (2005) stimuli show differences of instrumentation across stimuli

and vary in many other uncontrolled ways. The Shmulevich and Povel (2000) stimuli consist of abstract rhythms that are not played on musical instruments and that are not idiomatic for any specific kind of music. In all of these studies, the stimulus sets were prepared *ad hoc* in order to answer specific research questions; the authors did not design them for sustained use in music psychology, and they seem unsuitable for this purpose.

Clemente et al. (2020) presented a large stimulus set of 200 monophonic melodies that were explicitly designed for repeated use in music psychology. The stimuli vary along several axes that have been found to be important to music perception: balance, contour, symmetry, and complexity. Many stimulus properties are kept invariant across the set (instrumental timbre, monophony, length); others are easily controllable (number of notes, melodic intervals, tessitura, tempo, etc.). The Clemente et al. (2020) stimuli thus have many valuable qualities. Yet, the single-note piano melodies without loudness, timing, or tempo variations are not particularly naturalistic. Consequently, they are not adequate for research contexts in which ecological validity is key.

Witek et al. (2014) presented a set of 50 simplified drum patterns drawn from Western popular music recordings with a wide range of degree of syncopation. These stimuli show a balance between ecological validity and experimental control. Nevertheless, the simplifications of the drum patterns are substantial: The hi-hat voice is reduced to a simple sequence of eighth notes, thus neutralizing the contribution of the hi-hat (and other cymbals) to the complexity of the originally recorded patterns. The stimuli present all patterns at a tempo of 120 bpm instead of the originally recorded tempi. The stimuli do not implement variations in loudness and/or microtiming/swing. Finally, 13 of the 50 stimuli are not derived from originally recorded popular music but were composed by the researchers to show extreme (high or low) syncopation. All these modifications affect the naturalness of the stimuli. Furthermore, the complexity measures associated with the 50 stimuli are purely objective (*Index of Syncopation*, based on the monophonic syncopation measure by Longuet-Higgins & Lee, 1984), and they have been found to correlate only moderately well with perceptual syncopation measures (Hoesl & Senn, 2018).

In conclusion, presently there is no stimulus set available that enables the use of naturalistic music to study the effects of complexity on music listeners and that is associated with accurate subjective measures of perceived stimulus complexity. This study has the following goals: (1) it presents a set of 40 idiomatic, naturalistic drum pattern audio stimuli for which there is a spectrum of perceived complexity and that can be used in psychological research and in the modeling of drum pattern complexity; (2) it provides empirical subjective complexity measurements based on a listening experiment; (3) it uses these measurements to assess existing objective measures of drum pattern complexity and formulates a benchmark for future measures.

Methods and Materials

Forming the Stimulus Set

From the corpus of the Lucerne Groove Research Library, 40 popular music drum patterns were selected. This corpus currently consists of 251 audio reconstructions, transcriptions, and metadata relating to drum patterns from Western popular music (rock, funk, rhythm and blues, pop, disco, soul, heavy metal, and rock'n'roll) in common time (4/4 time signature). These patterns were originally recorded between the 1950s and the 2010s. The audio reconstructions of the drum patterns are based on expert transcriptions of eight-bar extracts from the original recordings and were prepared by two professional musicians (second and fourth author). Event timing relies on computer-assisted onset time measurements (precise to a few milliseconds). The loudness of each stroke is based on the perceptual judgment of the two transcribers (for information on the transcription, measurement, and reconstruction process, see Senn et al., 2018, pp. 5–7).

To select the 40 drum patterns for the current study, we applied two criteria: first, we wanted the drum patterns to be more or less homogenous in terms of their instrumentation. Patterns that included toms or additional percussion instruments were excluded from the selection. The remaining 184 drum patterns were eligible because they only used the bass drum, snare drum, hi-hat, and other cymbals.

Second, we wanted the stimulus set to cover the complexity range of the corpus and to fill this range more or less equidistantly. To achieve this, we calculated the *Index of Syncopation* (Witek et al., 2014, in the formalization of Hoesl & Senn, 2018) for the snare drum and bass drum voices of the 184 eligible patterns and ordered them by increasing degree of syncopation. The ordered list of drum patterns was divided into 20 shorter sublists of 9–10 drum patterns having a similar value on the *Index of Syncopation*. From each of these 20 syncopation-matched sublists, 2 patterns were randomly selected for this study's set of drum pattern stimuli. Even though the *Index of Syncopation* cannot be expected to be a perfect predictor of complexity (Hoesl & Senn, 2018, p. 10), it is nevertheless likely that the selected stimuli cover the complexity range of the original stimulus corpus to a great extent.

The 40 patterns were shortened from 8 measures duration to 4 measures, ending on the first beat of the fifth measure. Since most patterns repeat with periods of a half bar, one bar, or two bars, a duration of four bars is sufficient to present the period at least twice and thus reveal the periodic nature of the patterns. The shortness of the stimuli (between 8.8 and 16.5 s) facilitates their use in a listening experiment.

Stimuli were rendered from MIDI data in *Avid Pro Tools* (version 12.1) with drum samples from the *Toontrack Superior Drummer Custom & Vintage* audio samples library (version 2.4.4). The same set of drum samples was used for drumset instruments and playing techniques

throughout the 40 patterns, and light reverberation was added in *Pro Tools*. The reconstructions give the impression that they were recorded on the same drum kit under the same acoustic conditions. Information on each drum stroke (onset time, used audio sample, dynamics) can be found in the datafile *drumpatterns.Rda* (see *Data Availability*).

Experimental Design

The goal of the listening experiment was to measure the relative complexity of the 40 drum pattern stimuli as perceived by listeners. The experiment used an incomplete pairwise comparison design: In every trial, two stimuli were presented to the participants, who then selected which stimulus sounded more complex (“winner”) compared to the other (“loser”).

In music psychology, pairwise comparison designs are rarely used; Likert rating is much more frequent. Yet, human participants tend to carry out pairwise comparison tasks with more ease than Likert rating tasks (Clark et al., 2018; Laming, 1984; Phelps et al., 2015). In a comparison task, participants do not need to map their judgment onto an abstract ordinal scale, but they can rank two stimuli in a direct way. Additionally, each trial presents participants with all the information they need to carry out the task. They do not have to remember stimuli from earlier trials to provide a fair judgment. The approach also eliminates well-known biases connected to Likert-rating, such as response style bias (which range of a scale participants like to use, see Chen et al., 1995; Oishi et al., 2005) and the reference group effect (the same pattern might appear simple to a musician but complex to a non-musician, see Heine et al., 2002). Conversely, the comparison method is costly: Many trials are necessary to achieve a good differentiation. This is the case because every trial only offers ranking information on two stimuli; information pertinent to the size of the perceived complexity difference is lost. Participants only choose a winning pattern; they do not express whether the difference in terms of complexity is small or large.

The experiment was implemented as an incomplete pairwise comparison design because a complete design would have required that participants judged all 780 pairs that can be formed from the 40 drum patterns. The incomplete design was implemented as follows: The 40 stimuli were divided into two subsets. Each subset featured 20 drum patterns that covered the same range on the *Index of Syncopation*. The data collection was then carried out in two phases:

1. The first phase of the data collection (November 2020 – January 2021) established the relative complexity of the stimuli *within* these two subsets (see further details about the design in *Appendix B*). The design guaranteed that each participant judged all stimuli once in a comparison trial. It also

ensured that each combination of the 20 stimuli (190 combinations) appeared once in a block of 19 participants and that the presentation sequence of stimuli within trials and across the experiment was counterbalanced. The data from the first phase was used to create a ranking of the stimuli within the two subsets.

2. The second data collection phase (March 2021) established the relative complexity *between* the two subsets. In this second phase, each trial consisted of two stimuli, one from each of the two subsets, which were on the same complexity rank within their subset after phase 1. The presentation sequence of the trials was randomized and the presentation order of the two stimuli within each trial was counterbalanced.

The combined dataset collected in phases 1 and 2 allowed for the relative perceived complexity of all 40 stimuli to be estimated using one single probabilistic model (see *Statistical Analyses*).

Procedure

Participants filled the survey online on the SosciSurvey platform (www.soscisurvey.de). The survey was available in two languages, German and English. Participants gave informed consent and responded to a series of demographic questions (gender, age, country of residence, survey language skills, musical preferences). Participants' musical expertise was assessed using the *MSI Training* subscale of the Gold-MSI self-report inventory, which exists in English (Müllensiefen et al., 2014) and German (Schaal et al., 2014). In addition, participants assigned themselves to one of five expertise groups: *professional musicians*, *music students* (training to become professional musicians or music teachers), *amateur musicians*, *music listeners* (who listened to music frequently but did not actively play or sing music), and finally participants who were *not interested* in music.

Participants were asked to carry out the experiment in a quiet location using good quality headphones. They listened to one drum pattern (similar to the 40 experimental stimuli, but not part of the set) in order to test their audio equipment and set playback to a fairly loud, but agreeable volume. They were asked not to change the volume settings after this initial adjustment. Participants then carried out 20 experimental comparisons. In each trial, the two stimuli were presented as separate audio players within the browser window. Participants started the playback themselves and were allowed to listen to the stimuli as many times as they liked. They answered the question “Which of the two drum patterns sounds more complex to you?” using radio buttons. No definition of complexity was offered to the participants, because we wanted them to apply their own concept of complexity. Participants were advised to go with their gut feeling if they struggled to

make a decision. They could not advance in the survey without selecting a winning pattern in each trial (no ties were allowed).

After the last trial, participants were asked the question: “What makes a drum pattern complex?” They wrote their answers in a free text field but could opt out of answering this question. The data from this text field will be analyzed in a qualitative study in the future. Finally, participants could volunteer to share their e-mail addresses if they wanted to be informed about the outcome of the study and/or be invited to future surveys. This contact information was saved separately from the experimental data to preserve the anonymity of the participants. On average, it took participants 16 min to complete the survey.

Data Collection and Filtering

A total of 260 participants completed phase one of the survey between November 2020 and February 2021. They were recruited through two channels: 84 participants responded to an e-mail invitation that went out to the students and faculty of the Lucerne University of Applied Sciences and Arts (School of Music) and to personal connections of the authors. These participants filled either the German ($n = 63$) or the English language version ($n = 21$) of the survey. A total of 176 participants were recruited through Amazon’s MTurk platform; all of them took the survey in English. MTurk participants were remunerated with US\$6 for their participation. Other participants were remunerated with a food/drink voucher for the cafeteria on the university’s music campus. They could opt out of this remuneration. No partial study credits were offered to students.

In each trial, participants were asked to listen to both drum patterns in their entirety before choosing a winner. Observations ($n = 29$) were excluded if participants used less than 16 s on any of the trials, i.e., they did not listen all the way through the stimuli. Some participants ($n = 12$) reported having only basic skills (levels A1 or A2) in the respective survey language (German or English). For these participants, inclusion in the dataset was decided based on their answer to the open text question. Eight participants provided intelligible answers suggesting sufficient language skills and these were included in the final sample; the data of the remaining four participants were excluded. After this screening, the dataset consisted of 227 complete and valid observations. Due to dropouts and the screening procedure, the dataset did not show second-order balance: the 19 different combinations of trials (rows in Table 7 of Appendix B) had been filled between 11 and 14 times. To establish second-order balance, surplus cases were randomly dropped until all 19 combinations were represented exactly 11 times each.

Phase 2 of the experiment was offered in German only. Invitations to participate were circulated via e-mail among students and faculty of the Lucerne University of Applied Sciences and Arts (School of Music) and personal connections of the authors. Participants received a cafeteria voucher (with

the possibility to opt out). Phase 2 yielded complete and valid observations from 11 participants.

Participants

The data of $n = 220$ participants (209 responding in phase 1 and 11 in phase 2 of the experiment) were used for complexity modeling. These participants (74 female, 142 male, 3 other, 1 no answer) had a mean age of 38 years, ranging from 19 to 81 ($SD = 12.3$). Participants lived predominantly in the United States (126) and in Switzerland (65). The remaining participants were from India (8), Brazil (5), Germany (4), Australia (3), Canada (3), Italy (2), or Sri Lanka (1); three did not provide information about their country of residence.

A total of 37 participants self-identified as professional musicians, 34 as music students, and 43 as amateur musicians. The remaining participants stated they were either music listeners (104) or not interested in music (2). Participants had a mean *MSI Training* score of 25.25, which is not significantly different from the UK population norm of 26.52 ($t_{219} = -1.306$, $p = .193$) reported by Müllensiefen et al. (2014, p. 10). This result is surprising, since more than half of the participant sample consisted of professional or amateur musicians and music students ($n = 114$). This musician subsample did indeed have a high mean *MSI Training* score of 36.87. The remaining participants (music listeners, not interested) counterbalanced this with a very low mean *MSI Training* score of 12.75. A total of 50 participants (all from the MTurk subsample) had an *MSI Training* score of 7, which is the minimum value on this scale and indicates no music training at all.

Statistical Analyses

The experimental data were analyzed using the Bradley–Terry probability model, which was developed to analyze the outcomes of pairwise comparison experiments. The model was first studied by Zermelo (1929), then formalized by Bradley and Terry (1952), and later subsumed under the family of generalized linear models, due to its close relationship to the logistic probability model (Agresti, 2007, 2012; Bradley & El-Helbawy, 1976; Perez-Ortiz & Mantiuk, 2017). All statistical analyses were carried out with *R* (version 3.6.3) in the RStudio (version 1.1.463) environment. The BradleyTerry2 library (version 1.1–2) was used for complexity modeling (Turner & Firth, 2012). The ggplot2 (version 3.3.3) library was used for preparing plots and figures.

Results

Complexity Estimates and Their Interpretation

The experiment yielded 4,400 valid trials. Every stimulus participated in 220 trials: it was compared 11 times against each of the 19 other stimuli in the same subset

(phase 1 of the experiment) and another 11 times against one complexity-matched stimulus from the other set (phase 2 of the experiment). A Bradley–Terry model was fitted to the data. The fit was good ($\chi^2 = 4,517.1$, $df = 4,361$, $p = .049$) with only a very mild tendency of overdispersion (reduced chi-squared statistic: $\chi^2 / df = 1.036$, see Taylor, 1997, pp. 268–271). The Bradley–Terry coefficients $\hat{\beta}_i$ were estimated for each stimulus. These coefficients quantify the *Perceived Complexity* of each drum pattern. The stimuli are listed in Table 1 with increasing *Perceived Complexity* estimates (see also Figure 1).

The Bradley–Terry coefficients allow the estimation of the probability that one drum pattern wins a pairwise

comparison trial against another drum pattern when judged by a random member of the listener population, as represented by the participant sample. The estimated success probability $\hat{\Pi}_{ij}$ that stimulus i wins over stimulus j is given by

$$\hat{\Pi}_{ij} = \frac{e^{\hat{\beta}_i - \hat{\beta}_j}}{1 + e^{\hat{\beta}_i - \hat{\beta}_j}}.$$

Coefficients $\hat{\beta}_i$ and $\hat{\beta}_j$ are the Bradley–Terry coefficients corresponding to stimuli i and j , respectively (see Agresti, 2007, p. 266). The expression for the estimated success probability $\hat{\Pi}_{ij}$ is the logistic function with argument

Table 1. *Perceived Complexity* (Bradley–Terry Estimates $\hat{\beta}$) of the 40 stimuli using data from 4,400 trials. SE is the standard error of the difference between $\hat{\beta}$ s of neighboring stimuli, where Down refers to the next less complex and Up to the next more complex neighbor. The label (letter) was used to reference the stimulus during the experiment (note that stimulus 4 is lowercase “l” and stimulus 10 is uppercase “i”).

Stimulus (Label)				Song Title	Audiofile	Perceived Complexity ($\hat{\beta}_i$)	SE	
							Down	Up
1	(L)	A Kind of Magic	01_TayR_2.mp3	0.400	—	0.291		
2	(f)	(Sittin' On) The Dock of the Bay	02_JacA_2.mp3	0.408	0.291	0.258		
3	(a)	Smells Like Teen Spirit	03_GroD_3.mp3	0.476	0.258	0.254		
4	(l)	Boogie Wonderland	04_JohR_2.mp3	0.573	0.254	0.289		
5	(C)	Vultures	05_JorS_2.mp3	0.784	0.289	0.248		
6	(Q)	Kashmir	06_BonJ_5.mp3	1.182	0.248	0.269		
7	(s)	Street of Dreams	07_FreJ_3.mp3	1.210	0.269	0.260		
8	(O)	Change the World	08_Robj_3.mp3	1.230	0.260	0.237		
9	(B)	Let's Dance	09_HakO_3.mp3	1.263	0.237	0.234		
10	(l)	Space Cowboy	10_McKD_2.mp3	1.415	0.234	0.260		
11	(c)	I Feel for You	11_NelP_5.mp3	1.632	0.260	0.210		
12	(q)	Virtual Insanity	12_McKD_3.mp3	1.859	0.210	0.253		
13	(M)	Bravado	13_PeaN_4.mp3	1.902	0.253	0.246		
14	(i)	Let's Go Dancing	14_BroG_5.mp3	1.951	0.246	0.206		
15	(h)	Discipline	15_FreJ_5.mp3	2.091	0.206	0.244		
16	(j)	Pass the Peas	16_Staj_4.mp3	2.151	0.244	0.250		
17	(b)	The Pump	17_PhiS_5.mp3	2.189	0.250	0.204		
18	(o)	Roxanne	18_CopS_1.mp3	2.216	0.204	0.249		
19	(H)	Dreamin'	19_YouE_5.mp3	2.413	0.249	0.249		
20	(m)	Soon I'll Be Loving You Again	20_GadJ_1.mp3	2.511	0.249	0.249		
21	(G)	Summer Madness	21_BroG_4.mp3	2.530	0.249	0.249		
22	(j)	Listen Up!	22_HakO_2.mp3	2.586	0.249	0.205		
23	(k)	Jungle Man	23_ModJ_5.mp3	2.752	0.205	0.208		
24	(d)	Shake Everything You Got	24_ChaD_5.mp3	3.051	0.208	0.252		
25	(P)	Chicken	25_SteB_4.mp3	3.052	0.252	0.252		
26	(n)	Cissy Strut	26_ModJ_3.mp3	3.080	0.252	0.211		
27	(g)	Far Cry	27_PeaN_2.mp3	3.120	0.211	0.212		
28	(p)	Alone + Easy Target	28_GroD_5.mp3	3.130	0.212	0.214		
29	(t)	Soul Man	29_JorS_3.mp3	3.263	0.214	0.256		
30	(N)	Ain't Nobody	30_Robj_1.mp3	3.300	0.256	0.218		
31	(A)	Diggin' On James Brown	31_GarD_5.mp3	3.314	0.218	0.219		
32	(F)	In the Stone	32_JohR_3.mp3	3.342	0.219	0.258		
33	(r)	Southwick	33_SteB_1.mp3	3.360	0.258	0.221		
34	(e)	You Can Make It if You Try	34_ErrG_1.mp3	3.447	0.221	0.261		
35	(T)	The Dump	35_DeiA_3.mp3	3.464	0.261	0.221		
36	(D)	Killing In the Name of	36_WilB_3.mp3	3.564	0.221	0.225		
37	(K)	Cold Sweat	37_StuC_4.mp3	3.763	0.225	0.229		
38	(S)	Hyperpower	38_FreJ_4.mp3	3.902	0.229	0.242		
39	(E)	Rock Steady	39_PurB_5.mp3	4.394	0.242	0.261		
40	(R)	Jelly Belly	40_MarB_4.mp3	4.701	0.261	—		

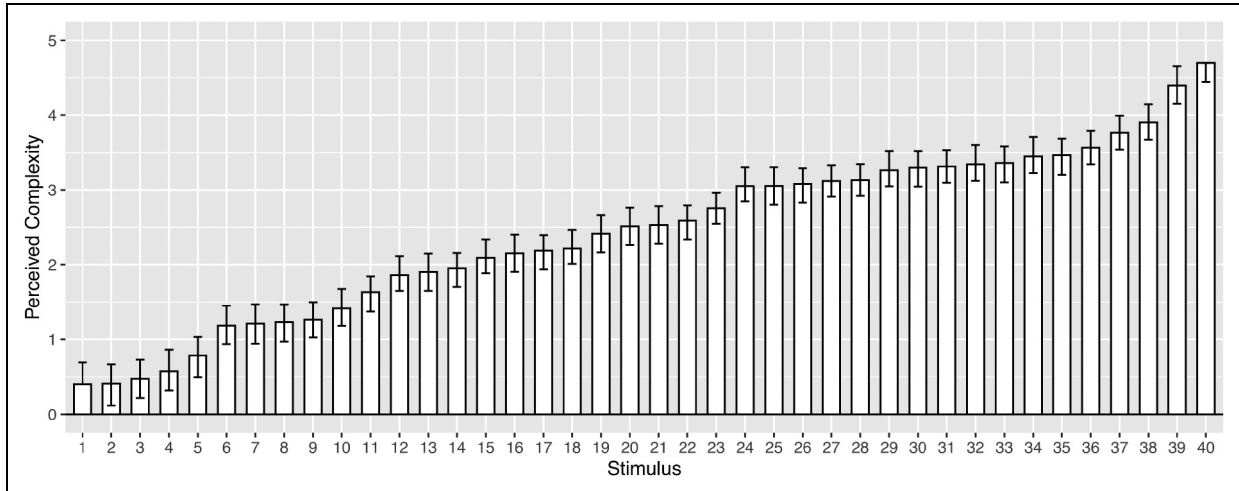


Figure 1. Perceived complexity (Bradley-Terry estimates $\hat{\beta}$) of the 40 stimuli. Error bars indicate the standard error for the difference between the $\hat{\beta}$ s of neighboring stimuli.

$\hat{\beta}_i - \hat{\beta}_j$. An example: The estimated probability that stimulus 6 (“Kashmir”) is considered to be more complex than stimulus 1 (“A Kind of Magic”) in a trial is (using the $\hat{\beta}_6$ and $\hat{\beta}_1$ coefficients from Table 1)

$$\hat{\Pi}_{6,1} = \frac{e^{\hat{\beta}_6 - \hat{\beta}_1}}{1 + e^{\hat{\beta}_6 - \hat{\beta}_1}} = \frac{e^{1.182 - 0.400}}{1 + e^{1.182 - 0.400}} \cong 0.686.$$

We can expect the “Kashmir” drum pattern to be considered more complex than the “A Kind of Magic” drum pattern in 68.6% of the trials. Conversely, the probability that “A Kind of Magic” is considered more complex than “Kashmir” is $\hat{\Pi}_{1,6} \cong 0.314$. Since no ties were allowed, it is true for all i and j that $\hat{\Pi}_{ij} + \hat{\Pi}_{ji} = 1$.

Inference on the difference between the $\hat{\beta}$ estimates is based on the Wald approximation (see Kutner et al., 2004, p. 578). The Bradley-Terry model provides an estimate of the standard error for the difference $\hat{\beta}_i - \hat{\beta}_j$ with respect to all $C_2^{40} = 780$ pairs of stimuli i and j (see datafile `se_matrix.Rda` in the supplemental material). The datafile `p_matrix.Rda` presents results of formal significance tests against the null hypothesis that $\Pi_{ij} = 0.5$ (which is equivalent to $\beta_i = \beta_j$) for each pair of stimuli. Table 1 and Figure 1 (error bars) only show the standard errors between each stimulus and its lower (Down) or upper (Up) neighbor on the list. The measurement error for the *Perceived Complexity* estimates $\hat{\beta}_i$ themselves is approximately $\hat{\sigma}_{\beta} = 0.175$.

The most complex stimulus 40 (“Jelly Belly”) has a very high probability of $\hat{\Pi}_{40,1} = .987$ to win a trial against the most simple stimulus 1 (“A Kind of Magic”), which indicates that the stimuli cover a complexity range that is perceivable by the vast majority of listeners. None of the stimuli that are neighbors in Table 1 show significant complexity differences: The increase of complexity from one stimulus to the next more complex stimulus seems to be smooth. There is one exception: Stimulus 39 is significantly more complex than stimulus 38 ($p = .042$).

Effect of Musical Training on Complexity Judgments

The sample of participants consisted of people with extreme differences in terms of musical training, as measured by the *MSI Training* scale. This raised the concern whether the sample represented a homogenous population or not. To address this concern, we investigated whether the experimental data from participants who were musically trained differed significantly from the data provided by the less musically trained participants. In general, people with substantial musical training show superior discrimination abilities across many domains of music perception (Besson et al., 2007; Chartrand et al., 2008; Neuhaus et al., 2006; Penhune, 2019). Accordingly, we would expect that people with more musical training have better abilities to detect complexity differences, compared to the less trained.

If so, then music experts’ complexity judgments should agree with each other often, resulting in lopsided distributions of wins between high- and low-complexity stimuli. Conversely, the musical non-experts can be expected to perceive complexity differences less clearly. In this case, we expect the trials to be decided by chance more often, which would result in a more uniform distribution of wins across stimuli.

Table 2 lists the frequencies of wins for each stimulus, grouped by musical training. The high-training group consists of participants who had an *MSI Training* score above 26.52 ($n = 105$), which is a score that is higher than the UK population average for this scale (Müllensiefen et al., 2014, p. 10). The low-training group had *MSI Training* scores below 26.52 ($n = 115$).

A χ^2 -test of independence between participants’ musical training and the wins of the 40 stimuli provided no evidence for an effect of expertise on the outcomes of the trials ($\chi^2_{(39)} = 42.099, p = .338$). The low- and high-training subsamples do not differ significantly in their responses to the comparison task. We conclude that the participants were

Table 2. Number of wins for each stimulus among participants with high ($MSI\ Training > 26.52$) or low ($MSI\ Training < 26.52$) musical expertise.

Stimulus	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
High Training	14	22	16	12	23	32	30	31	25	24	33	53	34	48	48	49	48	45	43	50
Low Training	19	16	23	29	21	25	34	26	32	38	49	39	47	47	53	42	57	61	59	69
Stimulus	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
High Training	58	59	66	73	66	81	60	62	79	64	67	66	72	74	72	65	81	74	86	95
Low Training	49	63	64	71	63	64	86	85	73	75	73	74	83	84	73	84	76	88	93	93

sampled from a population that is homogenous with respect to the ability to perceive and judge drum pattern complexity.

Training Set and Test Set

Future users of the stimulus set may want to divide the stimuli into two subsets of 20 stimuli to define one set as a training set for complexity model development and the other as a test set for model assessment. The variable *Set* (data file *stimuli.Rda*) provides such a bipartition, which makes sure that the two subsets are similarly composed with respect to the *Perceived Complexity* estimates. The two sets are closely matched with respect to the mean, the standard deviation, skewness, and excess kurtosis of *Perceived Complexity* (Table 3). An Anderson–Darling test provided no evidence that the sets were drawn from different populations with respect to the *Perceived Complexity* of the stimuli ($A^2 = 0.210$, $p = .996$).

Control Variables

The 40 drum pattern stimuli closely reproduce the rhythm, timing, and dynamics of the drum patterns played in the original recordings. Thanks to the detailed reconstruction, the stimuli are highly idiomatic and allow for the creation of experiments with good ecological validity. However, this implies that the stimuli differ in many ways besides the differences between the rhythmic patterns as represented in the transcriptions (*transcriptions.pdf*). These additional dimensions may need to be controlled when the stimuli are used in an experimental setting.

The drum pattern data file (*drumpatterns.Rda*) lists all information that was used to create the audio reconstructions. For every note onset ($n = 2,758$), it specifies the instrument and playing technique (*Instrument*), the physical onset time in seconds (*Seconds*), the metric time in beats (*Beats*), the audio sample from the *Toontrack Superior Drummer Custom & Vintage* library (version 2.4.4) used for reconstruction (*AudioSample*), and the associated loudness (*LoudnessDB*). Quadratic regression models were fitted to the notes' physical onset time (*Seconds*) and metrical time (*Beats*) data to calculate local tempi in beats per minute (*Local Tempo BPM*). This method allows slight tempo variations to be detected (many of the original

Table 3. Mean, standard deviation, skewness, and excess kurtosis of the *Perceived Complexity* estimates in the two subsets A and B. Set A consists of stimuli 2, 4, 5, 9, 10, 12, 14, 15, 16, 18, 20, 23, 25, 29, 30, 31, 33, 37, 38, 39; and set B of the remaining 20 stimuli.

Descriptive statistic	Set A	Set B
Mean	2.416	2.433
Standard deviation	1.143	1.138
Skewness	−0.166	−0.166
Excess kurtosis	−1.124	−0.887

recordings were not synchronized with a metronome click track and therefore show some tempo drift). Microtiming was calculated as onsets' time deviations from the fit values in seconds (*Microtiming Seconds*) and as a proportion of the local beat duration (*Microtiming Beats*). The information in the drum pattern data file can be used to form new control variables or to create new audio reconstructions in which one or more of these properties is manipulated.

The stimuli data file (*stimuli.Rda*) contains discographic and complexity-related data but also a series of descriptive statistics associated with each of the 40 stimuli. These statistics were derived from the drum pattern file and the audio files: They provide the duration of each stimulus in seconds (*Duration*), the loudness in decibels (*Loudness RMS*), the local tempi at the earliest (*Initial Tempo BPM*) and latest (*Final Tempo BPM*) onsets, the change of tempo happening in between (*Tempo Drift BPM*), and the absolute value of this change (*Absolute Tempo Drift BPM*). Finally, they contain the standard deviation of the microtiming values in each pattern, measured in seconds (*Microtiming Seconds SD*) or as proportions of the beat (*Microtiming Beats SD*).

Table 4 presents the results of tests that investigate whether *Perceived Complexity* is associated with any of these control variables. The tests provided no evidence for a dependence between the complexity measure and any of the control variables: Complexity does not seem to vary systematically in response to the duration of the stimulus (*Duration*), its loudness (*Loudness RMS*), its tempo (*Initial Tempo BPM*, *Final Tempo BPM*), tempo change (*Tempo Drift BPM*), tempo instability (*Absolute Tempo Drift BPM*), or microtiming (*Microtiming Seconds SD*, *Microtiming Beats SD*).

Table 4. Pearson correlation coefficients (r) of control variables with *Perceived Complexity*.

Control variables	r	t	df	p
Duration	0.274	1.759	38	.087
Loudness RMS	0.222	1.402	38	.169
Initial Tempo BPM	−0.122	−0.756	38	.454
Final Tempo BPM	−0.105	−0.649	38	.520
Tempo Drift BPM	0.193	1.214	38	.232
Absolute Tempo Drift BPM	0.175	1.093	38	.281
Microtiming Seconds SD	0.252	1.609	38	.116
Microtiming Beats SD	0.288	1.856	38	.071

Perceived Complexity and Objective Measures of Drum Pattern Complexity

How do the subjective *Perceived Complexity* estimates relate to existing objective measures of drum pattern complexity? The following analysis has two purposes: first, it assesses the fit of four objective measures of complexity and sets an initial benchmark for future complexity modeling. Second, it permits the assessment of how the size of the measurement error of *Perceived Complexity* relates to the error variances of the objective measures: The empirical subjective *Perceived Complexity* measures are only useful for the improvement of the objective measures if their measurement error is distinctly smaller than the error variance of the objective measures. The following four objective measures are investigated:

- *Number of Onsets*: The count of all note onsets played in a pattern is a simple baseline measure of complexity, assuming that more complex patterns tend to consist of more notes.
- *Syncopation Index*: Syncopation measure originally proposed by Witek et al. (2014) in the formalization by Hoesl and Senn (2018, p. 5–9).
- *Revised Syncopation Index*: Syncopation index provided by Hoesl and Senn (2018, p. 10–11), which is based on Witek et al. (2014) but re-estimates the weights of the metric positions within the bar.
- *Kolmogorov Complexity (BDM)*: This measure is an estimated value of the Kolmogorov complexity calculated using the block decomposition method (BDM) introduced by Dakos and Soler-Toscano (2017). The *accs* (version 0.2–5) and *spatialwarnings* (version 3.0.3) libraries were used to estimate Kolmogorov complexity in R.

The *Syncopation Index* and *Revised Syncopation Index* have been designed to register syncopation in the bass drum and the snare drum voices of a drum pattern; they do not consider syncopation in the cymbal voices. Prediction inaccuracies might be caused by syncopation in the cymbal voices that contribute to stimulus complexity but are not measured by either of the two methods.

Table 5 shows that all four measures were strongly and positively correlated with *Perceived Complexity*. The observed Pearson correlation coefficients ranged between $r = 0.600$ (*Number of Onsets*) and $r = 0.748$ (*Revised Syncopation Index*). The best methods performed near the benchmark that Honing and Smith (2006) established for the monophonic syncopation measures of Longuet-Higgins and Lee (1984, $r = 0.75$) and Palmer and Krumhansl (1990, $r = 0.73$).

Figure 2 shows scatterplots of *Perceived Complexity* (vertical axis) against the four objective measures (horizontal axis). It reveals that the syncopation-based measures have poor discrimination in the lower syncopation range. The best-fitting model, *Revised Syncopation Index*, predicts *Perceived Complexity* with a coefficient of determination of $R^2 = 0.560$. When regressed on *Revised Syncopation Index*, the *Perceived Complexity* measurements have a standard deviation of $s = 0.844$ about the estimated mean. This variability is greater than the measurement error of the *Perceived Complexity* estimates of $\hat{\sigma}_\beta = 0.175$. This implies that even the best of the four methods can be improved without overfitting the data.

Stimuli and Data Availability

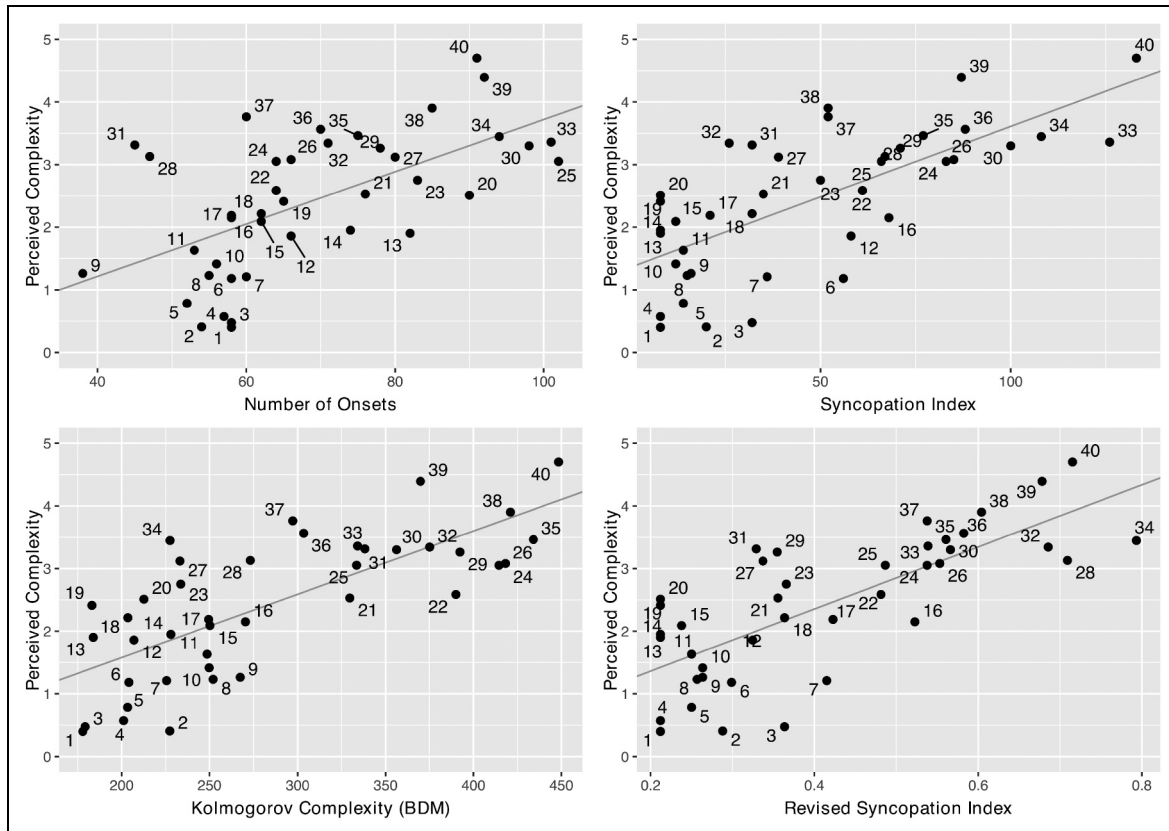
The audio reconstructions of the stimuli (mp3), transcriptions in drumset notation (transcriptions.pdf), and all information used to create the stimuli (drumpatterns.Rda) can be found online at [Authors' online resources](#). These resources also provide anonymized demographic information about the 220 participants (participants.Rda), the outcomes of the 4,400 pairwise comparison trials (contests.Rda), discographic information, perceived complexity measures, control variables, and further information about the 40 stimuli (stimuli.Rda, p_matrix.Rda, se_matrix.Rda). Discographic information about the 40 drum patterns is presented in *Appendix A* of this study (Table 6, below).

Discussion

This study presents a set of stimuli consisting of 40 drum patterns drawn from Western popular music. These stimuli replicate the original drum patterns faithfully by using information from detailed transcriptions in drumset notation, measurements of highly precise note onset timing, and subjective assessments of each stroke's loudness, provided by two experienced professional musicians. Drum patterns can be heard alone quite frequently in this kind of repertoire: Many songs start with a solo drum pattern (examples are Michael Jackson's "Billie Jean," Stevie Wonder's "Superstition," or Led Zeppelin's "When The Levee Breaks") or there may be a drum break later in the song (e.g., James Brown's "Funky Drummer" or The Winstons' "Amen, Brother"). Consequently, drum patterns are likely to evoke a naturalistic listening experience and may be used for experiments that aim at ecological validity. Drum patterns are considered essential elements of groove

Table 5. Pearson correlations of four measures of rhythm complexity with *Perceived Complexity*.

Method	Source(s)	<i>r</i>	<i>t</i>	<i>df</i>	<i>p</i>	<i>R</i> ²
Number of Onsets	Ad-hoc measure	0.600	4.626	38	<.001	0.360
Syncopation Index	Witek et al. (2014); Hoesl and Senn (2018)	0.701	6.057	38	<.001	0.491
Kolmogorov Complexity (BDM)	Dakos and Soler-Toscano (2017)	0.736	6.693	38	<.001	0.541
Revised Syncopation Index	Hoesl and Senn (2018)	0.748	6.953	38	<.001	0.560

**Figure 2.** Scatterplots of *Perceived Complexity* against four objective measures of rhythmic complexity: *Number of Onsets*, *Kolmogorov complexity (BDM)*, *Syncopation Index*, *Revised Syncopation Index*; see also Table 5.

in Western popular music; hence this set of stimuli will be particularly useful for groove research.

The high similarity of the audio reconstructions with the originally recorded drum patterns comes at the cost of experimental control: The stimuli do not just vary with respect to complexity but show differences in many other dimensions. To account for the effect of these types of variation, a series of control variables has been operationalized that allow experimenters to monitor the effects of nuisance variability. All data that informed stimuli creation have been provided as a dataset (drumpatterns.Rda). This information can be used to develop further control variables or to create new stimuli sets in which specific types of nuisance variability are suppressed.

The listening experiment provided a subjective measure of *Perceived Complexity* for each of the 40 popular music drum pattern stimuli. Since the *Perceived Complexity* estimates are equivalent to the Bradley–Terry coefficients $\hat{\beta}$, they have a clear probabilistic interpretation: They allow

the estimation of the probability that any of the 40 stimuli wins a pairwise comparison trial against any of the other stimuli in the surveyed listener population. To our knowledge, this is the first stimulus set in the music domain that offers this intuitive interpretation of measurements. The analysis further suggests that the stimuli cover a complexity range that is noticeable for the vast majority of listeners and that they fill this range in a more or less equidistant way.

We did not find an effect of musical training on the complexity judgments of the participants. Consequently, it is plausible that the participant sample was drawn from a homogenous population with respect to their ability to discern differences of complexity. This result contributes a nuance to the general findings in music psychology that music training enables superior discrimination across many domains of music perception (Besson et al., 2007; Chartrand et al., 2008; Neuhaus et al., 2006; Penhune, 2019): Musical training does not appear to alter perception of

complexity in popular music drum patterns. This counterintuitive result might be due to the ubiquitous presence of popular music in the media of the globalized West: An adult person with intact hearing is likely to have heard songs from this repertoire on many thousand occasions during their lifetime. So, people both with and without formal music training might implicitly be familiar with the properties of popular music drum patterns by sheer exposure.

It would be interesting to repeat this study's experiment with children, adolescents, and adults to investigate whether these groups show different abilities to distinguish between complex and simple stimuli. Carrying out this experiment with professional drummers/percussionists could potentially also lead to a different result, as this population may not only consider the perceived complexity of the patterns but also the difficulty of learning and playing them. We did not explicitly collect data from the highly proficient drummer population, so this hypothesis cannot be tested on the basis of this study's data.

We did not find significant relationships between eight loudness-, tempo-, or microtiming-related control variables and the *Perceived Complexity* estimates. This result does not imply that these variables are irrelevant as predictors of perceived complexity in general. It simply indicates that, for these 40 stimuli, in which these variables take idiomatic values, no systematic relationship of these variables with *Perceived Complexity* was observed. However, more extreme manipulations of the control variables (such as strongly increasing or decreasing the tempi), are likely to affect the perceived complexities of the stimuli.

Four objective measures of stimulus complexity were positively correlated with the *Perceived Complexity* estimates. The two best fitting measures (*Kolmogorov Complexity*, *Revised Syncopation Index*) approached the benchmark of $r = .75$ established by Honing and Smith (2006). However, with a fit of $R^2 = .560$, not even the *Revised Syncopation Index* can be recommended without reservation as a reliable measure of drum pattern complexity; there is considerable room for improvement. Recently, Senn (2023) presented an objective method for the estimation of drum pattern complexity based on ideas from predictive coding, which had an excellent fit with the empirical data ($R^2 = .852$).

Experimenters can use the complexity measurements, the control variables (*stimuli.Rda*), and the pairwise significance test data (*p_matrix.Rda*) to choose drum pattern subsets according to their specific requirements. The pairwise comparison data (*contests.Rda*) is available if researchers want to add new stimuli to the set; in this case it will be necessary to measure the perceived complexity of the new stimuli against each other and in relation to the existing stimuli in the course of a listening experiment and fit a Bradley–Terry model to the combined data.

The stimulus set and measurements have a series of limitations. First, the drum patterns are drawn from commercial recordings of Western popular music; they only cover the complexity range that is idiomatic within this repertoire. Consequently, these stimuli will not be adequate for

studies that investigate the effects of extremely high or low complexity. Second, rhythm is the primary source of complexity in these drum pattern stimuli. Other types of complexity (harmonic, melodic, formal, or others) are not varied in this stimulus set. This means that experimental results based on these stimuli will not permit statements on musical complexity in general. Third, we can expect that the perceived complexity of a stimulus decreases with familiarity (Hannon et al., 2012). We did not measure participants' familiarity with the stimuli, and therefore we cannot assess the relevance of this effect for this study. The complexity measurements rely on the perception of participants that predominantly live in Western, educated, industrialized, rich and democratic (WEIRD) countries (Muthukrishna et al., 2020). Researchers should be aware that experimental results based on these stimuli are primarily valid for the sampled repertoire and population. It may not be warranted to generalize results to different musical repertoires and to other human populations (Apicella et al., 2020; Henrich et al., 2010; Jacoby et al., 2021).

Conclusion

This study (1) presented a set of forty naturalistic stimuli that vary along a complexity dimension. (2) It provided a perceived complexity measurement that has an intuitive probabilistic interpretation for each of the stimuli. (3) These subjective measurements were then used to test existing objective measures of drum pattern complexity and to formulate a benchmark for future objective measures. The stimuli are more naturalistic than those presented in Clemente et al. (2020) and are associated with a ground truth measurement of complexity that is more accurate than the *Indices of Syncopation* found in Hoesl and Senn (2018) and Witek et al. (2014). The stimulus set and perceptual complexity measurements will be a useful resource for listening experiments in which stimulus complexity is a relevant independent variable. They also allow one to assess how well objective measures of stimulus complexity predict subjectively perceived complexity. They will support the development of reliable objective complexity measures that are closely aligned with empirical subjective measurements, and they will allow theories and definitions of musical complexity to be investigated. The set will enable the testing of hypotheses on the relationship between stimulus complexity and the experience of groove, building on work by Witek et al. (2014), Matthews et al. (2019), Sioros et al. (2022) and Stupacher et al. (2022). The stimuli will also be useful for the development of complexity models that are based on cognitive theories (e.g., predictive coding or dynamic attending theory, see Senn, 2023) potentially contributing to a better understanding of subjectively perceived complexity.

Action Editor

Jessica Grahn, Western University, Brain and Mind Institute & Department of Psychology.

Peer Review

Deniz Duman, University of Jyväskylä, Department of Music, Art and Culture Studies.

Maria Witek, University of Birmingham, Department of Music.

Contributorship

Olivier Senn: Conceptualization, Methodology, Investigation, Formal Analysis, Writing – Original Draft, Writing – Review & Editing, Visualization, Project Administration.

Florian Hoesl: Resources, Investigation, Data Curation, Writing – Review & Editing

Rafael Jerjen: Resources, Methodology, Investigation, Formal Analysis, Writing – Review & Editing

Toni Bechtold: Resources, Writing – Review & Editing

Lorenz Kilchenmann: Resources, Writing – Review & Editing

Dawn Rose: Writing – Review & Editing

Elena Alessandri: Funding Acquisition, Writing – Review & Editing, Supervision

Data Availability

The stimuli and datasets generated and/or analyzed during the current study are available at *Authors' online resources* (<https://zenodo.org/record/6343826#>).

Consent to Participate and Publish

Informed consent was obtained from all individual participants included in the study.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.


Ethics Statement


This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committee of the Lucerne University of Applied Sciences and Arts (October 30, 2020/No. EK-HSLU 005 M20).


Funding


The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung (grant number 100016 192398/1).


ORCID iDs


Olivier Senn  <https://orcid.org/0000-0002-4944-3115>

Florian Hoesl  <https://orcid.org/0000-0002-6700-5354>

Toni Amadeus Bechtold  <https://orcid.org/0000-0002-4768-5997>

Lorenz Kilchenmann  <https://orcid.org/0000-0003-3723-6298>

Dawn Rose  <https://orcid.org/0000-0003-2945-9491>

Elena Alessandri  <https://orcid.org/0000-0003-3750-3851>

References

- Agresti, A. (2007). *An Introduction to categorical data analysis* (2 edition). Wiley-Interscience.
- Agresti, A. (2012). *Categorical data analysis* (3 edition). Wiley.

- Apicella, C., Norenzayan, A., & Henrich, J. (2020). Beyond WEIRD: A review of the last decade and a look ahead to the global laboratory of the future. *Evolution and Human Behavior*, 41(5), 319–329. <https://doi.org/10.1016/j.evolhumbehav.2020.07.015>
- Beauvois, M. W. (2007). Quantifying aesthetic preference and perceived complexity for fractal melodies. *Music Perception: An Interdisciplinary Journal*, 24(3), 247–264. <https://doi.org/10.1525/mp.2007.24.3.247>
- Berlyne, D. E. (1963). Complexity and incongruity variables as determinants of exploratory choice and evaluative ratings. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, 17(3), 274–290. <https://doi.org/10.1037/h0092883>
- Berlyne, D. E. (1971). *Aesthetics and psychobiology*. Appleton-Century-Crofts.
- Besson, M., Schön, D., Moreno, S., Santos, A., & Magne, C. (2007). Influence of musical expertise and musical training on pitch processing in music and language. *Restorative Neurology and Neuroscience*, 25(3–4), 399–410.
- Bradley, R. A., & El-Helbawy, A. T. (1976). Treatment contrasts in paired comparisons: Basic procedures with application to factorials. *Biometrika*, 63(2), 255–262. <https://doi.org/10.1093/biomet/63.2.255>
- Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika*, 39(3/4), 324–345. <https://doi.org/10.2307/2334029>
- Cameron, D. J., Caldarone, N., Psaris, M., Carrillo, C., & Trainor, L. J. (2023). The complexity-aesthetics relationship for musical rhythm is more fixed than flexible: Evidence from children and expert dancers. *Developmental Science*, 26(5). <https://doi.org/10.1111/desc.v26.5>
- Carney, J., Wlodarski, R., & Dunbar, R. (2014). Inference or enaction? The impact of genre on the narrative processing of other minds. *PLOS ONE*, 9(12), e114172. <https://doi.org/10.1371/journal.pone.0114172>
- Chartrand, J.-P., Peretz, I., & Belin, P. (2008). Auditory recognition expertise and domain specificity. *Brain Research*, 1220, 191–198. <https://doi.org/10.1016/j.brainres.2008.01.014>
- Chemillier, M. (2002). Ethnomusicology, ethnomathematics. The logic underlying orally transmitted artistic practices. In G. Assayag, H. G. Feichtinger, & J. F. Rodrigues (Eds.), *Mathematics and music: A diderot mathematical forum* (pp. 161–183). Springer. https://doi.org/10.1007/978-3-662-04927-3_10
- Chen, C., Lee, S., & Stevenson, H. W. (1995). Response style and cross-cultural comparisons of rating scales among East Asian and North American students. *Psychological Science*, 6(3), 170–175. <https://doi.org/10.1111/j.1467-9280.1995.tb00327.x>
- Chmiel, A., & Schubert, E. (2019). Unusualness as a predictor of music preference. *Musicae Scientiae*, 23(4), 426–441. <https://doi.org/10.1177/1029864917752545>
- Clark, A. P., Howard, K. L., Woods, A. T., Penton-Voak, I. S., & Neumann, C. (2018). Why rate when you could compare? Using the “EloChoice” package to assess pairwise comparisons of perceived physical strength. *PLOS ONE*, 13(1), e0190393. <https://doi.org/10.1371/journal.pone.0190393>
- Clemente, A., Vila-Vidal, M., Pearce, M. T., Aguiló, G., Corradi, G., & Nadal, M. (2020). A Set of 200 Musical stimuli varying

- in balance, contour, symmetry, and complexity: Behavioral and computational assessments. *Behavior Research Methods*, 52, 1491–1509. <https://doi.org/10.3758/s13428-019-01329-8>
- Cox, D. S., & Cox, A. D. (1988). What does familiarity breed? Complexity as a moderator of repetition effects in advertisement evaluation. *Journal of Consumer Research*, 15(1), 111–116. <https://doi.org/10.1086/209149>
- Dakos, V., & Soler-Toscano, F. (2017). Measuring complexity to infer changes in the dynamics of ecological systems under stress. *Ecological Complexity*, 32, 144–155. <https://doi.org/10.1016/j.ecocom.2016.08.005>
- De Fleurian, R., Blackwell, T., Ben-Tal, O., & Müllensiefen, D. (2017). Information-theoretic measures predict the human judgment of rhythm complexity. *Cognitive Science*, 41(3), 800–813. <https://doi.org/10.1111/cogs.12347>
- Eerola, T. (2016). Expectancy-violation and information-theoretic models of melodic complexity. *Empirical Musicology Review*, 11(1), 2–17. <https://doi.org/10.18061/emr.v11i1.4836>
- Eisenberg, J., & Thompson, W. F. (2003). A matter of taste: Evaluating improvised music. *Creativity Research Journal*, 15(2–3), 287–296. <https://doi.org/10.1080/10400419.2003.9651421>
- Gómez, F., Melvin, A., Rappaport, D., Toussaint, G. T., Service, M. M., & Toussaint, G. T. (2005). Mathematical measures of syncopation. In *Proc. BRIDGES: Mathematical Connections in Art, Music and Science*, 73–84.
- Gómez, F., Thul, E., & Toussaint, G. T. (2007). An experimental comparison of formal measures of rhythmic syncopation. Proceedings of the International Computer Music Conference (ICMC) 2007, 101–104
- Hannon, E. E., Soley, G., & Ullal, S. (2012). Familiarity overrides complexity in rhythm perception: A cross-cultural comparison of American and Turkish listeners. *Journal of Experimental Psychology: Human Perception and Performance*, 38(3), 543–548. <https://doi.org/10.1037/a0027225>
- Heine, S. J., Lehman, D. R., Peng, K., & Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective Likert scales?: The reference-group effect. *Journal of Personality and Social Psychology*, 82(6), 903–918. <https://doi.org/10.1037/0022-3514.82.6.903>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2–3), 61–83. <https://doi.org/10.1017/S0140525X0999152X>
- Heyduk, R. G. (1975). Rated preference for musical compositions as it relates to complexity and exposure frequency. *Perception & Psychophysics*, 17(1), 84–90. <https://doi.org/10.3758/BF03204003>
- Hoesl, F., & Senn, O. (2018). Modelling perceived syncopation in popular music drum patterns: A preliminary study. *Music & Science*, 1, 2059204318791464. <https://doi.org/10.1177/2059204318791464>
- Honing, H., & Smith, L. M. (2006). Evaluating and extending computational models of rhythmic syncopation in music. Proceedings of the International Computer Music Conference (ICMC) 2006, 1–4.
- Imamoglu, Ç (2000). Complexity, liking and familiarity: Architecture and non-architecture Turkish students' assessments of traditional and modern house facades. *Journal of Environmental Psychology*, 20(1), 5–16. <https://doi.org/10.1006/jevp.1999.0155>
- Jacoby, N., Polak, R., Grahn, J., Cameron, D., Lee, K. M., Godoy, R., Undurraga, E. A., Huanca, T., Thalwitzer, T., Dombia, N., Goldberg, D., Margulis, E., Wong, P. C. M., Jure, L., Rocamora, M., Fujii, S., Savage, P. E., Ajimi, J., & Konno, R., ..., J. H. McDermott (2021). Universality and cross-cultural variation in mental representations of music revealed by global comparison of rhythm priors. *PsyArXiv*. <https://doi.org/10.31234/osf.io/b879v>
- Janata, P., Tomic, S. T., & Haberman, J. M. (2012). Sensorimotor coupling in music and the psychology of the groove. *Journal of Experimental Psychology: General*, 141(1), 54–75. <https://doi.org/10.1037/a0024208>
- Keith, M. (1991). *From polychords to polya: Adventures in musical combinatorics*. Vinculum Pr.
- Kutner, M., Nachtsheim, C., Neter, J., & Li, W. (2004). *Applied linear statistical models* (5th edition). McGraw-Hill/Irwin.
- Laming, D. (1984). The relativity of 'absolute' judgements. *British Journal of Mathematical and Statistical Psychology*, 37(2), 152–183. <https://doi.org/10.1111/j.2044-8317.1984.tb00798.x>
- London, J. (2004). *Hearing in time: Psychological aspects of musical meter*. Oxford University 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>
- Lumaca, M., Haumann, N. T., Brattico, E., Grube, M., & Vuust, P. (2019). Weighting of neural prediction error by rhythmic complexity: A predictive coding account using mismatch negativity. *European Journal of Neuroscience*, 49(12), 1597–1609. <https://doi.org/10.1111/ejn.14329>
- Mai, R., Hoffmann, S., Schwarz, U., Niemand, T., & Seidel, J. (2014). The shifting range of optimal web site complexity. *Journal of Interactive Marketing*, 28(2), 101–116. <https://doi.org/10.1016/j.intmar.2013.10.001>
- Marin, M. M., & Leder, H. (2013). Examining complexity across domains: Relating subjective and objective measures of affective environmental scenes, paintings and music. *PloS One*, 8(8), 1–35. <https://doi.org/10.1371/journal.pone.0072412>
- Matthews, T. E., Witek, M. A. G., Heggli, O. A., Penhune, V. B., & Vuust, P. (2019). The sensation of groove is affected by the interaction of rhythmic and harmonic complexity. *PLOS ONE*, 14(1), e0204539. <https://doi.org/10.1371/journal.pone.0204539>
- Matthews, T. E., Witek, M. A. G., Lund, T., Vuust, P., & Penhune, V. B. (2020). The sensation of groove engages motor and reward networks. *NeuroImage*, 214, 116768. <https://doi.org/10.1016/j.neuroimage.2020.116768>
- Mihelač, L., & Povh, J. (2020). The impact of the complexity of harmony on the acceptability of music. *ACM Transactions on Applied Perception*, 17(1), 3:1-3:27. <https://doi.org/10.1145/3375014>
- Morillon, B., & Zalta, A. (2021). Prominence of delta oscillatory rhythms in the motor cortex and their relevance for auditory perception. Rhythm Perception and Performance Workshop

- (RPPW) 2021. <https://www.uio.no/ritmo/english/news-and-events/events/conferences/2021/RPPW/videos/talks/61%20morillon.mp4?vrtx=view-as-webpage>
- 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. <https://doi.org/10.1371/journal.pone.0089642>
- Muthukrishna, M., Bell, A. V., Henrich, J., Curtin, C. M., Gedranovich, A., McInerney, J., & Thue, B. (2020). Beyond Western, Educated, Industrial, Rich, and Democratic (WEIRD) psychology: Measuring and mapping scales of cultural and psychological distance. *Psychological Science*, 31(6), 678–701. <https://doi.org/10.1177/0956797620916782>
- Nadal, M., Munar, E., Marty, G., & Cela-Conde, C. J. (2010). Visual complexity and beauty appreciation: Explaining the divergence of results. *Empirical Studies of the Arts*, 28(2), 173–191. <https://doi.org/10.2190/EM.28.2.d>
- Nadkarni, S., & Gupta, R. (2007). A task-based model of perceived website complexity. *MIS Quarterly*, 31(3), 501–524. <https://doi.org/10.2307/25148805>
- Neuhaus, C., Knösche, T. R., & Friederici, A. D. (2006). Effects of musical expertise and boundary markers on phrase perception in music. *Journal of Cognitive Neuroscience*, 18(3), 472–493. <https://doi.org/10.1162/jocn.2006.18.3.472>
- North, A. C., & Hargreaves, D. J. (1995). Subjective complexity, familiarity, and liking for popular music. *Psychomusicology: A Journal of Research in Music Cognition*, 14(1–2), 77–93. <https://doi.org/10.1037/h0094090>
- North, A. C., & Hargreaves, D. J. (1996). Responses to music in aerobic exercise and yogic relaxation classes. *British Journal of Psychology*, 87(4), 535. <https://doi.org/10.1111/j.2044-8295.1996.tb02607.x>
- Oishi, S., Hahn, J., Schimmack, U., Radhakrishnan, P., Dzokoto, V., & Ahadi, S. (2005). The measurement of values across cultures: A pairwise comparison approach. *Journal of Research in Personality*, 39(2), 299–305. <https://doi.org/10.1016/j.jrp.2004.08.001>
- Orr, M. G., & Ohlsson, S. (2001). The relationship between musical complexity and liking in jazz and bluegrass. *Psychology of Music*, 29(2), 108–127. <https://doi.org/10.1177/0305735601292002>
- Orr, M. G., & Ohlsson, S. (2005). Relationship between complexity and liking as a function of expertise. *Music Perception*, 22(4), 583–611. <https://doi.org/10.1525/mp.2005.22.4.583>
- Osborne, J. W., & Farley, F. H. (1970). The relationship between aesthetic preference and visual complexity in abstract art. *Psychonomic Science*, 19(2), 69–70. <https://doi.org/10.3758/BF03337424>
- Palmer, C., & Krumhansl, C. L. (1990). Mental representations for musical meter. *Journal of Experimental Psychology: Human Perception and Performance*, 16(4), 728–741. <https://doi.org/10.1037/0096-1523.16.4.728>
- Penhune, V. B. (2019). Musical expertise and brain structure: The causes and consequences of training. In M. H. Thaut & D. A. Hodges (Eds.), *The Oxford handbook of music and the brain* (pp. 419–438). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198804123.013.17>
- Perez-Ortiz, M., & Mantiuk, R. K. (2017). A practical guide and software for analysing pairwise comparison experiments. *ArXiv:1712.03686 [Cs, Stat]*. <http://arxiv.org/abs/1712.03686>
- Phelps, A. S., Naeger, D. M., Courtier, J. L., Lambert, J. W., Marcovici, P. A., Villanueva-Meyer, J. E., & MacKenzie, J. D. (2015). Pairwise comparison versus Likert scale for biomedical image assessment. *American Journal of Roentgenology*, 204(1), 8–14. <https://doi.org/10.2214/AJR.14.13022>
- Pressing, J. (1999). Cognitive complexity and the structure of musical patterns. *Proceedings of the 4th Conference of the Australasian Cognitive Science Society*, 4, 1–8. <http://dub.ucs-d.edu/Mu206/CogComplex-music.pdf>
- Russell, P. A. (1982). Relationships between judgements of the complexity, pleasingness and interestingness of music. *Current Psychology*, 2(1), 195–201. <https://doi.org/10.1007/BF03186760>
- Schaal, N. K., Bauer, A.-K. R., & Müllensiefen, D. (2014). Der Gold-MSI: Replikation und Validierung eines Fragebogeninstrumentes zur Messung Musikalischer Erfahrung anhand einer deutschen Stichprobe. *Musicae Scientiae*, 18(4), 423–447. <https://doi.org/10.1177/1029864914541851>
- Senn, O. (2023). A predictive coding approach to modelling the perceived complexity of popular music drum patterns. *Heliyon*, 9(4), e15199. <https://doi.org/10.1016/j.heliyon.2023.e15199>
- Senn, O., Bechtold, T., Rose, D., Câmara, G. S., Düvel, N., Jerjen, R., Kilchenmann, L., Hoesl, F., Baldassarre, A., & Alessandri, E. (2020). Experience of Groove Questionnaire: Instrument development and initial validation. *Music Perception*, 38(1), 46–65. <https://doi.org/10.1525/mp.2020.38.1.46>
- Senn, O., Kilchenmann, L., Bechtold, T., & Hoesl, F. (2018). Groove in drum patterns as a function of both rhythmic properties and listeners' attitudes. *PLOS ONE*, 13(6), e0199604. <https://doi.org/10.1371/journal.pone.0199604>
- 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](https://doi.org/10.1076/0929-8215(200003)29:01;1-P;FT061)
- Sioros, G., Madison, G., Cocharro, D., Danielsen, A., & Gouyon, F. (2022). Syncopation and groove in polyphonic music: Patterns matter. *Music Perception*, 39(5), 503–531. <https://doi.org/10.1525/mp.2022.39.5.503>
- Spiech, C., Sioros, G., Endestad, T., Danielsen, A., & Laeng, B. (2022). Pupil drift rate indexes groove ratings. *Scientific Reports*, 12(1), 11620. <https://doi.org/10.1038/s41598-022-15763-w>
- Steck, L., & Machotka, P. (1975). Preference for musical complexity: Effects of context. *Journal of Experimental Psychology: Human Perception and Performance*, 1(2), 170–174. <https://doi.org/10.1037/0096-1523.1.2.170>
- Stokmans, M. J. W. (2003). How heterogeneity in cultural tastes is captured by psychological factors: A study of Reading fiction. *Poetics*, 31(5), 423–439. <https://doi.org/10.1016/j.poetic.2003.09.003>
- Stupacher, J., Matthews, T. E., Pando-Naude, V., Foster Vander Elst, O., & Vuust, P. (2022). The sweet spot between predictability and surprise: Musical groove in brain, body, and social interactions. *Frontiers in Psychology*, 13(906190), 1–9. <https://www.frontiersin.org/articles/10.3389/fpsyg.2022.906190> <https://doi.org/10.3389/fpsyg.2022.906190>

- Stupacher, J., Witek, M. A. G., Vuoskoski, J. K., & Vuust, P. (2020). Cultural familiarity and individual musical taste differently affect social bonding when moving to music. *Scientific Reports*, 10(1), 10015. <https://doi.org/10.1038/s41598-020-66529-1>
- Taylor, J. R. (1997). *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements* (2nd edition). University Science Books.
- Thul, E., & Toussaint, G. T. (2008). Rhythm complexity measures: A comparison of mathematical models of human perception and performance. *International Conference on Music Information Retrieval (ISMIR)*, 9, 663–668.
- Toussaint, G. (2002). A mathematical analysis of African, Brazilian, and Cuban clave rhythms. *Bridges: Mathematical Connections in Art, Music, and Science*, 157–168.
- Trainor, L. J., Marie, C., Gerry, D., Whiskin, E., & Unrau, A. (2012). Becoming musically enculturated: Effects of music classes for infants on brain and behavior. *Annals of the New York Academy of Sciences*, 1252(1), 129–138. <https://doi.org/10.1111/j.1749-6632.2012.06462.x>
- Turner, H., & Firth, D. (2012). Bradley-Terry models in R: The BradleyTerry2 package. *Journal of Statistical Software*, 48(1), 1–21. <https://doi.org/10.18637/jss.v048.i09>
- Van Geert, E., & Wagemans, J. (2020). Order, complexity, and aesthetic appreciation. *Psychology of Aesthetics, Creativity, and the Arts*, 14(2), 135–154. <https://doi.org/10.1037/aca0000224>
- Vitz, P. C. (1964). Preferences for rates of information presented by sequences of tones. *Journal of Experimental Psychology*, 68(2), 176–183. <https://doi.org/10.1037/h0043402>
- Vitz, P. C. (1966). Preference for different amounts of visual complexity. *Behavioral Science*, 11(2), 105–114. <https://doi.org/10.1002/bs.3830110204>
- Vuust, P., Ostergaard, L., Pallesen, K. J., Bailey, C., & Roepstorff, A. (2009). Predictive coding of music – Brain responses to rhythmic incongruity. *Cortex*, 45(1), 80–92. <https://doi.org/10.1016/j.cortex.2008.05.014>
- Vuust, P., & Witek, M. A. G. (2014). Rhythmic complexity and predictive coding: A novel approach to modeling rhythm and meter perception in music. *Frontiers in Psychology*, 5(01111), 1–14. <https://doi.org/10.3389/fpsyg.2014.01111>
- Witek, M. A. G., Clarke, E. F., Wallentin, M., Kringelbach, M. L., & Vuust, P. (2014). Syncopation, body-movement and pleasure in groove music. *PLoS ONE*, 9(4), 1–12. <https://doi.org/10.1371/journal.pone.0094446>
- Witek, M. A. G. (2017). Filling in: Syncopation, pleasure and distributed embodiment in groove. *Music Analysis*, 36(1), 138–160. <https://doi.org/10.1111/musa.12082>
- Witek, M. A. G., Clarke, E. F., Wallentin, M., Kringelbach, M. L., & Vuust, P. (2015). Correction: Syncopation, body-movement and pleasure in groove music. *PLOS ONE*, 10(9), e0139409. <https://doi.org/10.1371/journal.pone.0139409>
- Wundt, W. M. (1874). *Grundzüge der physiologischen Psychologie*. W. Engelmann.
- Zermelo, E. (1929). Die Berechnung der Turnier-Ergebnisse als ein Maximumproblem der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 29(1), 436–460. <https://doi.org/10.1007/BF01180541>

Appendix

Appendix A: Discographic Information

Table 6. Discographic source information on the 40 drum patterns.

Stimulus	Original recording (song title, act, album title, year)	Original drummer	Minutage	Audiofile
1	A Kind of Magic, Queen, A Kind of Magic, 1986	Roger Taylor	0:26–0:35	01_TayR_2.mp3
2	(Sittin' on) The Dock of the Bay, Otis Redding, The Dock of the Bay, 1967	Al Jackson Jr.	0:28–0:39	02_JacA_2.mp3
3	Smells Like Teen Spirit, Nirvana, Nevermind, 1991	Dave Grohl	0:34–0:44	03_GroD_3.mp3
4	Boogie Wonderland, Earth, Wind & Fire, I Am, 1978	Ralph Johnson	0:32–0:41	04_JohR_2.mp3
5	Vultures, John Mayer, Continuum, 2005	Steve Jordan	0:01–0:12	05_JorS_2.mp3
6	Kashmir, Led Zeppelin, Physical Graffiti, 1974	John Bonham	0:18–0:32	06_BonJ_5.mp3
7	Street of Dreams, Guns N' Roses, Chinese Democracy	Josh Freese	1:13–1:26	07_FreJ_3.mp3
8	Change the World, Eric Clapton, Phenomenon, 1996	John Robinson	2:41–2:53	08_RobJ_3.mp3
9	Let's Dance, David Bowie, Let's Dance, 1982	Omar Hakim	0:08–0:18	09_HakO_3.mp3
10	Space Cowboy, Jamiroquai, The Return of the Space Cowboy, 1994	Derrick McKenzie	0:07–0:18	10_McKD_2.mp3
11	I Feel for You, Prince, Prince, 1979	Prince	2:44–2:57	11_NelP_5.mp3
12	Virtual Insanity, Jamiroquai, Travelling Without Moving, 1996	Derrick McKenzie	0:09–0:21	12_McKD_3.mp3
13	Bravado, Rush, Roll The Bones, 1991	Neil Peart	0:08–0:19	13_PeaN_4.mp3
14	Let's Go Dancin', Kool & The Gang, As One, 1982	George Brown	1:20–1:35	14_BroG_5.mp3
15	Discipline, Nine Inch Nails, The Slip, 2008	Josh Freese	0:55–1:04	15_FreJ_5.mp3
16	Pass the Peas, The J.B.'s, Food For Thought, 1971	Jabo Starks	0:17–0:28	16_Staj_4.mp3
17	The Pump, Jeff Beck, There & Back, 1980	Simon Phillips	4:02–4:18	17_PhiS_5.mp3
18	Roxanne, The Police, Outlandos d'Amour, 1978	Stewart Copeland	1:02–1:14	18_CopS_1.mp3
19	Dreamin', Loleatta Holloway, Loleatta, 1976	Earl Young	0:20–0:33	19_YouE_5.mp3
20	Soon I'll Be Loving You Again, Marvin Gaye, I Want You, 1975	James Gadson	1:05–1:17	20_GadJ_1.mp3
21	Summer Madness, Kool & The Gang, Light Of The Worlds, 1974	George Brown	1:52–2:05	21_BroG_4.mp3
22	Listen Up!, The Omar Hakim Experience, We Are One, 2014	Omar Hakim	1:54–2:10	22_HakO_2.mp3
23	Jungle Man, The Meters, Rejuvenation, 1974	Joseph Modeliste	0:00–0:13	23_ModJ_5.mp3
24	Shake Everything You Got, Maceo Parker, Roots & Grooves, 2007	Dennis Chambers	1:52–2:08	24_ChaD_5.mp3
25	Chicken, Maceo Parker, Mo' Roots, 1991	Bill Stewart	0:11–0:24	25_SteB_4.mp3
26	Cissy Strut, The Meters, The Meters, 1968	Joseph Modeliste	0:03–0:16	26_ModJ_3.mp3
27	Far Cry, Rush, Snakes & Arrows, 2006	Neil Peart	0:50–1:03	27_PeaN_2.mp3
28	Alone + Easy Target, Foo Fighters, Foo Fighters, 1994	Dave Grohl	0:22–0:34	28_GroD_5.mp3
29	Soul Man, The Blues Brothers, Briefcase Full Of Blues, 1978	Steve Jordan	0:21–0:31	29_JorS_3.mp3
30	Ain't Nobody, Rufus & Chaka Khan, Stompin' At The Savoy Live, 1982	John Robinson	2:22–2:33	30_RobJ_1.mp3
31	Diggin' On James Brown, Tower Of Power, Soul Vaccination Live, 1998	David Garibaldi	0:08–0:17	31_GarD_5.mp3
32	In the Stone, Earth, Wind & Fire, I Am, 1978	Ralph Johnson	0:53–1:03	32_JohR_3.mp3
33	Southwick, Maceo Parker, Mo' Roots, 1991	Bill Stewart	0:13–0:26	33_SteB_1.mp3
34	You Can Make It if You Try, Sly And The Family Stone, Stand!, 1968	Greg Errico	2:47–2:58	34_ErrG_1.mp3
35	The Dump, Lettuce, Live in Tokyo, 2003	Adam Deitch	5:58–6:10	35_DeiA_3.mp3
36	Killing In the Name of, Rage Against The Machine, Rage Against The Machine, 1992	Brad Wilk	0:17–0:30	36_WilB_3.mp3
37	Cold Sweat, James Brown, Cold Sweat, 1967	Clyde Stubblefield	0:08–0:18	37_StuC_4.mp3
38	Hyperpower!, Nine Inch Nails, Year Zero, 2006	Josh Freese	0:00–0:16	38_FreJ_4.mp3
39	Rock Steady, Aretha Franklin, Young, Gifted And Black, 1971	Bernard Purdie	2:30–2:44	39_PurB_5.mp3
40	Jelly Belly, Medeski Martin & Wood, Shack-Man, 1996	Billy Martin	0:20–0:31	40_MarB_4.mp3

Appendix B: Design of the Listening Experiment

The 40 stimuli were divided into two sets of 20 (uppercase and lowercase sets). Each stimulus in one set was matched with one stimulus in the other set that had the same (or very similar) *Index of Syncopation*. The goal was to create two sets that uniformly covered the entire complexity range. Each stimulus was then randomly assigned an uppercase or lowercase letter label. The labels of the stimuli are listed in Table 1.

Phase 1. Each of the 19 rows in Table 7 shows 10 pairs of letters from A to T. Every sequence contains each letter exactly once and, across the 19 rows, all 190 pairwise combinations of letters appear exactly once. The sequence of letters within each pair and across the 10 pairs were counterbalanced.

Each participant was randomly assigned to 1 of the 19 rows and to a starting set (upper- or lowercase). The partic-

ipant first carried out the 10 trials with the stimuli from the starting set, as specified by the row, and then the 10 trials with the stimuli from the other set, specified by the same row. So, if a participant was assigned row 4 and started with the uppercase set, the trials first presented JS, BI, KP, ..., DQ in trials 1–10, and then stimuli js, bi, kp, ..., dq from the lowercase set were presented in trials 11–20.

Phase 2. In phase 2, participants carried out 20 trials. In each trial, they compared one stimulus from the uppercase subset and one stimulus from the lowercase subset. The stimuli were selected to have the same complexity rank in their respective set after phase 1. These pairs were: fL, Ca, lQ, sO, cB, lq, Mi, Jh, Hb, oG, mP, Aj, kN, dF, Tn, Dp, gK, St, Er, eR. The sequence of trials was randomized for each participant.

Table 7. Design of the listening experiment, phase 1. Each participant was randomly assigned 1 row and carried out all pairwise comparisons (trials) of this sequence for the 20 uppercase and the 20 lowercase stimuli.

Row	Trial									
	1	2	3	4	5	6	7	8	9	10
1	KS	IL	JP	EH	CQ	RT	GA	OF	DM	NB
2	HO	DL	KB	GF	IT	EQ	MR	PS	AC	NJ
3	NG	AT	CM	KR	EJ	DF	BH	PL	IQ	OS
4	JS	BI	KP	LC	AM	TF	ON	RE	GH	DQ
5	OD	TE	LG	QK	IC	HF	SA	MJ	PN	BR
6	JF	MS	DE	RA	TH	QG	PB	NL	CK	OI
7	IH	QN	KO	EB	PA	FL	GM	RS	DC	TJ
8	RO	ND	KI	SE	AJ	GT	ML	PH	FC	BQ
9	EA	OT	IG	PF	HD	MK	QR	JL	NS	BC
10	GJ	AH	CS	FN	OE	DK	TM	IR	LB	PQ
11	AD	SH	NC	LT	OG	RJ	KE	BM	PI	FQ
12	HL	OC	GK	ST	JD	NI	RF	MQ	AB	EP
13	TD	IJ	KH	OM	AF	LQ	EN	PR	GC	BS
14	NA	CJ	KT	EI	BF	LO	QS	HR	GD	PM
15	PO	AQ	GS	ID	CT	EL	NR	FK	HM	JB
16	GB	KJ	NH	DS	OA	CP	MI	RL	FE	TQ
17	GE	JH	DP	IA	FS	OQ	NM	CR	LK	BT
18	NT	EM	FI	QJ	AK	CH	BO	PG	SL	RD
19	GR	EC	OJ	NK	LA	FM	HQ	BD	IS	PT