Information-theoretic Modeling of Perceived Musical Complexity

Sarah A. Sauvé & Marcus T. Pearce
Queen Mary University of London, London, United Kingdom

What makes a piece of music appear complex to a listener? This research extends previous work by Eerola (2016), examining information content generated by a computational model of auditory expectation (IDyOM) based on statistical learning and probabilistic prediction as an empirical definition of perceived musical complexity. We systematically manipulated the melody, rhythm, and harmony of short polyphonic musical excerpts using the model to ensure that these manipulations systematically varied information content in the intended direction. Complexity ratings collected from 28 participants were found to positively correlate most strongly with melodic and harmonic information content, which corresponded to descriptive musical features such as the proportion of out-of-key notes and tonal ambiguity. When individual differences were considered, these explained more variance than the manipulated predictors. Musical background was not a significant predictor of complexity ratings. The results support information content, as implemented by IDyOM, as an information-theoretic measure of complexity as well as extending IDyOM's range of applications to perceived complexity.

MUSICAL COMPLEXITY HAS BEEN A phenomonen of interest in the study of music emotion (Balkwill & Thompson, 1999; Huron, 2006), musical preference (Berlyne, 1974; Burke & Gridley, 1990; North & Hargreaves, 1995), rhythm perception (Chen, Penhune, & Zatorre, 2008; Large, Fink, & Kelso, 2002; Shmulevich & Povel, 2000; Song, Simpson, Harte, Pearce, & Sandler, 2013; Zatorre, Chen, & Penhune, 2007), salience (Prince, Thompson, & Schmuckler, 2009), melody identification (Madsen & Widmer, 2007), and neural responses to music

(Birbaumer, Lutzenberger, Rau, Braun, & Mayer-Kress, 1996). It is important to distinguish between the complexity of music, which can be assessed using a range of acoustic and symbolic measures of musical structure, and the complexity perceived by a listener, which is likely to be related to some of those measures. A key challenge in music perception is to understand which measures of musical complexity influence perceived complexity.

Information-theoretic approaches provide an attractive way to address this challenge because they are domain-general and have a well-defined interpretation in terms of the storage and communication of information. Eerola (2016) investigated perceived melodic complexity by comparing feature-based models (called expectancy-violation models by Eerola) to information-theoretic models. Feature-based models were constructed using eight principles addressing melodic and rhythmic aspects of a given melody, such as pitch proximity, tonal ambiguity, note density, and rhythmic variation. Information-theoretic models computed statistical properties of pitch class, pitch interval, and note duration, using first to third-order distributions. Both types of models were tested on their ability to predict existing complexity ratings for seven datasets, encompassing a variety of musical styles. The best performing model was an expectancy-violation model with four predictors, namely tonal ambiguity, pitch proximity, entropy of rhythmic distribution, and entropy of pitch-class distribution. However, the use of entropy blurs the distinction between expectancy-violation models and information-theoretic models; therefore, the model was reduced to three predictors: tonal ambiguity, pitch proximity, and rhythmic variability. This model, named EV_3, explained 68% of the variance in complexity ratings for seven datasets, representing an optimal balance between predictive power and parsimony.

The present study will further investigate the link between information-theoretic measures of predictability and perceived musical complexity by extending Eerola's (2016) work in two ways: 1) adding causal evidence by explicitly manipulating the melodic, rhythmic, and harmonic information content of stimuli (Margulis, 2016); and 2) investigating the relative weight of melodic, rhythmic, and harmonic information in perceived complexity of polyphonic music. The use of

polyphonic stimuli and the inclusion of harmonic complexity go beyond the melodic and rhythmic measures used in previous research (Eerola, 2016).

We use the IDyOM model (Pearce, 2005, 2018) to assess the information-theoretic properties of musical stimuli. IDyOM is a variable-order Markov model (Begleiter, El-Yaniv, & Yona, 2004; Bunton, 1997) that uses a multiple-viewpoint framework (Conklin & Witten, 1995), allowing it to combine models of different representations of the musical surface. IDyOM uses statistical learning to acquire models of the structural regularities in music and then uses these models to generate probabilistic predictions for forthcoming musical events based on the preceding context. Given a musical context, IDyOM estimates the probability of different continuations of the context based on how often they have appeared in similar contexts in its previous experience of music. IDyOM's predictions combine probabilities derived from a long-term model trained on a large corpus, reflecting schematic learning of structure through long-term exposure to a musical style, and a short-term model, trained incrementally on the current piece of music, reflecting learning of local learning of motivic structure internal to a piece of music. IDyOM can generate probabilistic predictions for the pitch and timing of a musical note in a melodic context and the next chord in a harmonic sequence.

IDyOM has been shown to accurately predict Western listeners' pitch expectations in behavioral, physiological and EEG studies (e.g., Egermann, Pearce, Wiggins, & McAdams, 2013; Hansen & Pearce, 2014; Omigie, Pearce, & Stewart, 2012; Omigie, Pearce, Williamson, & Stewart, 2013; Pearce, 2005; Pearce, Ruiz, Kapasi, Wiggins, & Bhattacharya, 2010). In many circumstances, IDyOM provides a more accurate model of listeners' pitch expectations than static rule-based models (e.g., Narmour, 1990; Schellenberg, 1997). It also been shown to account for expectations for the timing of melodic events (Sauvé, Sayed, Dean, & Pearce, 2018) and harmonic movement (Harrison & Pearce, 2018; Sears, Pearce, Caplin, & McAdams, 2018). Furthermore, IDyOM can simulate other psychological processes in music perception, including similarity perception (Pearce & Müllensiefen, 2017), recognition memory (Agres, Abdallah, & Pearce, 2017), phrase boundary perception (Pearce, Müllensiefen, & Wiggins, 2010), and aspects of emotional experience (Egermann et al., 2013; Gingras et al., 2016; Sauvé et al., 2018). The present research applies IDyOM to the question of musical complexity for the first time.

We propose the overall hypothesis that perceived complexity of polyphonic music is related to predictability given a learned model of the syntactic structure of a musical style. Therefore, we conceive of perceived complexity not in terms of the absolute properties or features of a piece of music but rather in terms of its congruity to the structural regularities of a musical style with which the listener is familiar. More specifically we hypothesize that information-theoretic measures of melodic, rhythmic, and harmonic predictability, as computed by IDyOM, accurately simulate perceived complexity of polyphonic music. To test this hypothesis, complexity ratings were collected for specially composed 3-voice stimuli for which we systematically manipulated the melody, rhythm, and harmony, using IDyOM to ensure that these manipulations changed information content for the respective musical parameter in the intended direction, while maintaining information content as constant as possible for the other two parameters. Participant ratings of complexity are expected to vary with information content across versions of the stimuli, such that higher information content predicts higher complexity ratings. For comparison with the information-theoretic measure of complexity as stylistic predictability, we apply Eerola's (2016) feature-based EV_3 model to the complexity ratings collected for this new set of stimuli.

In comparison to the correlational approach taken by Eerola (2016) using real-world musical examples, we attempt a more causal intervention by deliberately manipulating a stimulus in order to create systematic variations of information theoretic predictability, independently for rhythmic, melodic, and harmonic parameters. To the extent that the results corroborate the hypothesized relationship between information-theoretic predictability and perceived complexity, this allows us to be more confident that no other confounding factors can account for the observed effects. For the same reason, we compare feature-based models to increase our confidence that it is the predictability of the stimulus rather than features of the stimulus per se that influence perceived complexity.

The stimuli exhibit wide variations in predictability to ensure sufficient variance in information content across the three parameters (melody, rhythm, and harmony) to allow a robust statistical analysis of the relationship between information content and perceived complexity. This manipulation necessarily reduced the ecological validity of the stimuli in terms of stylistic congruence, sophistication, and musical quality. However, we are interested here in perception of complexity regardless of whether the stimuli adhere to any music-theoretic principles. Nonetheless, to accommodate the fact that our participants were encultured in Western tonal
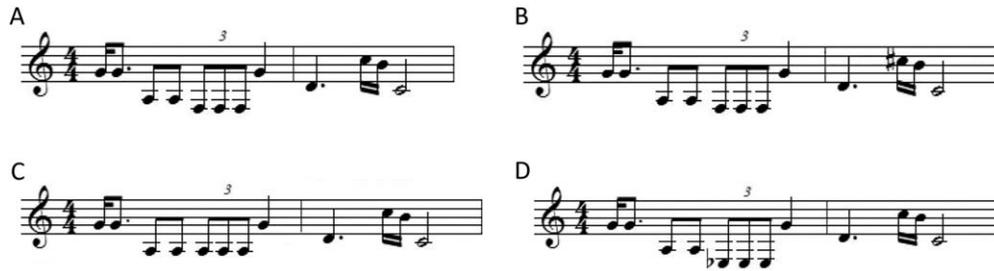
FIGURE 1. Middle voice (A), with each of the following modifications: A instead of F in the third beat of bar 1 (B), C♯ instead of C in the second half of the downbeat of bar 2 (C), and E♭ instead of F in the third beat of bar 1 (D).

music, the simulations used to inform the creation of the stimuli employ IDyOM models trained on Western tonal music. We predict that musical properties of the stimuli that are unpredictable for these models (based on learning the syntactic structure of Western tonal music) will also be perceived by listeners as unpredictable, and therefore complex. To ensure sufficient variance and the absence of ceiling effects in the complexity ratings, we encouraged participants to rate the complexity of the stimuli relative to other stimuli in the experiment, rather than on an absolute scale.

This approach provides a balance between experimental control and ecological validity. In comparison to studies using real-world stimuli (e.g., Eerola, 2016), it has greater experimental control but lower ecological validity, whereas in comparison to studies using artificial, stylistically unfamiliar stimuli (e.g., Loui, Wessel, & Kam, 2010), it has greater ecological validity but lower experimental control.

Method

PARTICIPANTS

Data were collected from 28 participants (12 female), mean age 43.03 ($SD = 16.34$) and mean Gold-MSI music training subscale (Müllensiefen, Gingras, Musil, & Stewart, 2014) score 36.60 ($SD = 9.31$) out of a maximum possible score of 49, indicating a relatively high level of music training overall. Participants were recruited through musicology and psychology mailing lists and social media. Ethical approval was obtained from the Queen Mary Research Ethics Committee, QMREC1536a.

Stimuli

The stimuli were composed by the first author, using IDyOM to ensure that information content varied systematically in the desired direction between levels of

complexity for each of the three musical parameters while keeping information content for the remaining two parameters as constant as possible. A total of 24 basic stimuli were designed, eight for each of the three musical parameters corresponding to eight levels of information-theoretic predictability with increasing information content, as measured by IDyOM (described in further detail below). Each stimulus consists of two bars written for three voices, and only the two outer voices were manipulated according to information content. Each of the 24 basic stimuli were created in four versions, with the outer voices combined with one of four different middle voices for a total of 96 different two-bar musical excerpts. As shown in Figure 1, two of the four possible middle voices contain all in-key notes and two contain out-of-key notes according to the implied harmony of the target melody alone, but in-key notes and out-of-key notes could become out-of-key or in-key depending on the harmonic context in which they are set. This relationship between the voices will be accounted for as features of all stimulus voices are incorporated into the analysis. A complete set of 24 stimuli for one target melody is shown in the Appendix.

Each stimulus was created as a MIDI file, then rendered to an audio file, with a violin sound applied to the upper voice, a clarinet sound applied to the middle voice, and a bassoon sound applied to the lower voice. These instruments were chosen so that each outer/inner voice timbral pair would be roughly equally perceptually dissimilar, using timbre dissimilarity ratings collected in a previous study (Sauvé, Stewart, & Pearce, 2014).

COMPLEXITY MEASURES

IDyOM generates probability distributions for each event in a piece of music that are conditioned upon the preceding musical context and the prior musical experience of the model. The probability of each event can be log-transformed to yield its *information content* (IC) according to the model (MacKay, 2003), which reflects

how unpredictable the model finds a note in a particular context. Averaging information content for all the events in a stimulus produces a measure of the overall predictability of the musical sequence. We use this as a measure of complexity under the hypothesis that unpredictable sequences (those with high information content) will be perceived as more complex than predictable sequences (those with low information content).

Middle voice. IDyOM was configured to predict chromatic pitch using a representation that combines scale degree with pitch interval, such that each event is represented as a pair of values: the scale degree (0–11) and the melodic pitch interval in semitones (with sign distinguishing ascending and descending intervals). IDyOM parameters were the same as used in previous research (Hansen & Pearce, 2014; Omigie et al., 2012, 2013; Pearce, 2005, 2018; Pearce et al., 2010): Both the short-term and long-term models were used, the latter being trained on 903 folk songs and chorales (datasets 1, 2, and 9 from Table 4.1 in Pearce, 2005, comprising 50,867 notes). Rhythmic complexity was simulated by configuring IDyOM to predict note onset using a representation of interonset interval (the difference in onset time between successive events) with other parameters the same as above.

The mean melodic information content of middle voices A, B, C, and D (Figure 1) was 5.61, 6.55, 8.66, and 8.13 respectively and the mean rhythmic information content was 3.70 for all versions.

Melodic complexity levels. The outer voices were designed to vary in mean information content between eight levels, such that level 1 had the lowest mean IC and level 8 the highest. Mean melodic information content for each complexity level ranged from 2.79 to 7.62 ($SD = 1.57$). Figure 2 illustrates the mean information content for all parameters for all 24 basic stimuli. Harmonic complexity varied across levels (range = 3.47–6.50, $SD = 1.05$) while rhythmic complexity did not (IC = 1.52).

Rhythmic complexity levels. IDyOM was configured to predict note onset in the outer voices of each stimulus, where note onset sequences in these voices were manipulated to have higher (level 8) or lower (level 1) IC. As shown in Figure 2, average rhythmic information content for each complexity level ranged from 1.49 to 3.04 ($SD = 0.53$). Melodic information content varied in a narrow range from 3.55 to 4.32 ($SD = 0.28$) while harmonic information content remained constant at 3.47 across all levels of rhythmic complexity. The variation in melodic information content is caused

by the variation in interval content in each rhythmic complexity level as the pattern of repeated and non-repeated notes changes with the increasingly complex rhythm.

Harmonic complexity levels. Chord progressions were four chords long, two beats per chord and harmonic complexity was measured using IDyOM trained on chord progressions from the Montreal Billboard Corpus (Burgoyne, Wild, & Fujinaga, 2011), in which the alphabet of possible chords includes major, minor, seventh, and various extension chords in almost any inversion. Each stimulus was encoded as a chord progression as a series of four integers, one for each chord (e.g., the progression I–IV–V–I was encoded as 1–2–3–1 and corresponded to an IC of 3.47). While the training material included chord inversions, due to the minimal perceived difference between chord inversions in such a short stimulus, the four-chord progressions in the stimuli encoded only the root and the quality (major, minor, augmented, diminished) of the chord, with the possibility of a chordal seventh where applicable. Other IDyOM parameters were as above. As shown in Figure 2, average harmonic information content for each complexity level ranged from 3.47 to 9.10 ($SD = 2.07$). Melodic information content varied in a relatively narrow range (range = 2.35–4.04, $SD = 0.66$) while rhythmic information content remained constant at 1.52 across all levels of harmonic complexity. The close relationship between melody and harmony makes complete independence impossible, though as illustrated in Figure 2, a degree of separation was achieved. This relationship will affect interpretation of the results, since a significant effect of harmonic information content implies a related influence of melodic information content and vice versa.

It is important to note that each musical parameter in this study is manipulated separately to yield varying ranges of average information content for each stimulus, as described above. The varying range is primarily a result of the size of each parameter's alphabet, where rhythm contains the smallest alphabet and harmony the largest. Only a relatively small number of different interonset intervals are possible for the rhythmic parameter when compared to the number of chromatic pitches available to the melodic parameter. Harmony has the largest alphabet as the Montreal Billboard Corpus includes four-note chords and their inversions, producing a much larger alphabet for the harmonic parameter than the melodic parameter. These differences will be accounted for in the analysis and incorporated into the interpretation of the results.
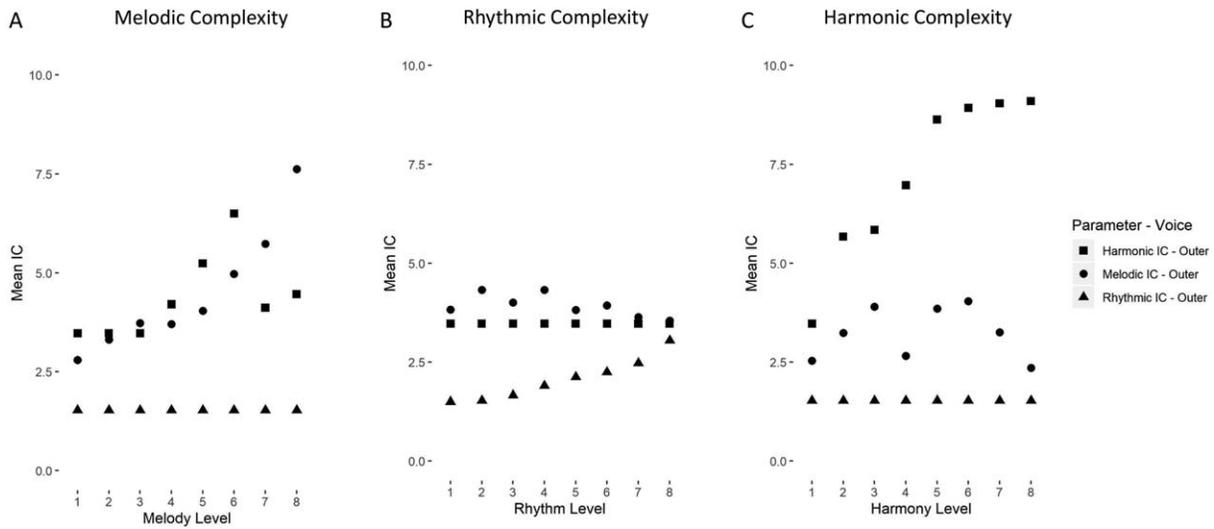
FIGURE 2. The mean information content (IC) of the outer voices plotted for each of the 24 stimuli by type of manipulation (melody, rhythm, harmony). IC for melodic complexity levels is illustrated in panel A, where melodic IC (circles) increases with level. IC for rhythmic complexity levels is illustrated in panel B, where rhythmic IC (triangles) increases with level. IC for harmonic levels is illustrated in panel C, where harmonic IC (squares) increases with level. Note that rhythmic IC is constant in panels A and C while harmonic IC is constant in panel B. However, due to the close relationship between melody and harmony, the manipulation of harmonic IC in panel C is only partially independent from melodic IC and the manipulation of melodic IC in panel A is only partially independent from harmonic IC.

EV$_3$

For the feature-based EV$_3$ model (Eerola, 2016), the MIDI toolbox (Toiviainen & Eerola, 2016) was used to calculate tonal ambiguity, pitch proximity, and rhythmic variability for each outer voice and averaged for each stimulus. These measures were also computed for each target melody.

PROCEDURE

Data were collected via the online survey tool Qualtrics. Participants first read through the information sheet and provided consent before reading the instructions and answering two practice trials to familiarize themselves with the type of stimuli and form an idea of their complexity. Participants were simply asked to rate the complexity of each stimulus. They were encouraged to use the full range of the complexity scale presented, a Likert scale ranging from 1 (*not complex*) to 7 (*very complex*), and judge complexity of a stimulus in relation to the other stimuli rather than in relation to other music. For each participant, two of the four target melodies were randomly selected resulting in 48 stimuli for each participant. The 48 stimuli were divided into three blocks of 16 stimuli, one for each of the melodically, harmonically and rhythmically manipulated stimuli. Within blocks, the 16 stimuli were randomized for each participant and the presentation order of the blocks was also randomized across participants.

ANALYSIS

All analyses were performed using R (3.3.2). In addition to compiling descriptive statistics, the degree of correlation between mean participant ratings and melodic, harmonic, and rhythmic IC was evaluated using Pearson's correlation coefficient. Next, multiple linear regression analyses were performed using the *lme4* package (Bates, Maechler, Bolker, & Walker, 2015). Models were constructed to predict complexity ratings averaged across participants for each stimulus. Models were evaluated using Pearson's correlation between the model's predictions and the data. Variance explained by each model was measured by calculating the coefficient of determination R^2. The overall F-statistic of the model was also recorded and Cohen's F^2 effect size reported using each model's adjusted R^2. Statistical significance of each predictor was tested by a likelihood-ratio test between a null model (intercept only) and a model containing the single evaluated predictor. Statistical significance of each individual factor level for a given predictor was evaluated using 95% confidence intervals, where an interval not including zero indicates a significant predictor.

The influence of music training on complexity ratings was evaluated by adding a fixed effect for training, reflected by Gold-MSI scores for each participant. In contrast to the previous analysis, here the models are applied to the complexity rating for each trial. To include

TABLE 1. *Descriptive Statistics (Means and Standard Deviations) for Participant Complexity Ratings for Each Parameter and Level*

	Level								
	1	2	3	4	5	6	7	8	Mean
Melody	2.96 (1.51)	3.32 (1.44)	3.73 (1.38)	4.10 (1.39)	3.78 (1.23)	4.58 (1.17)	4.32 (1.34)	4.48 (1.48)	3.90 (0.60)
Harmony	3.62 (1.32)	3.07 (1.34)	3.96 (1.30)	3.25 (1.27)	4.71 (1.13)	4.58 (1.10)	4.25 (1.39)	4.42 (1.34)	3.98 (0.65)
Rhythm	3.62 (1.57)	3.19 (1.40)	3.85 (1.32)	3.67 (1.37)	3.35 (1.36)	3.46 (1.40)	3.85 (1.44)	4.65 (1.30)	3.70 (0.48)
Mean	3.39 (0.38)	3.19 (0.35)	3.82 (0.43)	3.66 (0.40)	3.93 (0.60)	4.21 (0.61)	4.16 (0.40)	4.49 (0.23)	3.86 (0.59)

in the model maximal random effects in accordance with the experimental design (Barr, Levy, Scheepers, & Tily, 2013), random intercepts on participant and stimulus number were added to all models including music training to create multiple linear mixed effects models. To evaluate the significance of music training, mixed effects models with and without the predictor were compared using a log likelihood ratio test. The multiple linear mixed-effects model was evaluated using Pearson correlation between the model's predictions and the data and the coefficient of determination R^2.

Results

Descriptive statistics can be found in Table 1. Multiple linear regression analyses were carried out to address the relationship between the complexity ratings and three aspects of the stimuli: first, the categorical experimental manipulations of stimulus complexity; second, the quantitative information-theoretic measures of predictability corresponding to the experimental manipulations; and third, descriptive features of the stimuli. The influence of music training (as a fixed effect) is also evaluated for each of these three analyses. An additional analysis was conducted to evaluate the predictive power of Eerola's (2016) EV$_3$ model.

EXPERIMENTAL MANIPULATIONS

First, does the manipulation of information-theoretic predictability across 8 levels per musical parameter predict perceived complexity ratings? For this model, *complexity* (1–8), *musical parameter* (melody, harmony, or rhythm) and *version* (four types of middle voice) were included as predictors of mean complexity ratings. Figure 3 shows the overall increase in mean complexity ratings with complexity level for all musical parameters. Complexity was treated as a continuous numeric factor while for musical parameter, factor levels were compared to rhythm, and for version, comparisons were made to Version A. This model has $R^2 = .43$ and $F(6, 89) = 13.28$, $p < .0001$, $F^2 = 0.77$ and $r = .68$. Overall, *complexity* was a significant predictor of mean complexity ratings, $F(1, 94) = 67.46$, $p < .0001$, $F^2 = 0.69$. while *musical parameter* and *version* were not, $F(2, 93) = 1.89$, $p = .15$, $F^2 = 0.01$, and $F(3, 92) = 0.47$, $p = .69$, $F^2 = 0.01$, respectively. More specifically, only harmony was significantly different from rhythm, $t(89) = 2.49$, $p = .01$ for harmony and $t(89) = 1.78$, $p = .07$ for melody, while no level of version was significantly different from Version A (all $p > .05$). Modelling the interaction between complexity and musical parameter yielded a significant effect, $R^2 = .45$, $F(3, 92) = 27.91$, $p < .0001$, $F^2 = 0.84$ and $r = .69$, but adding the interaction
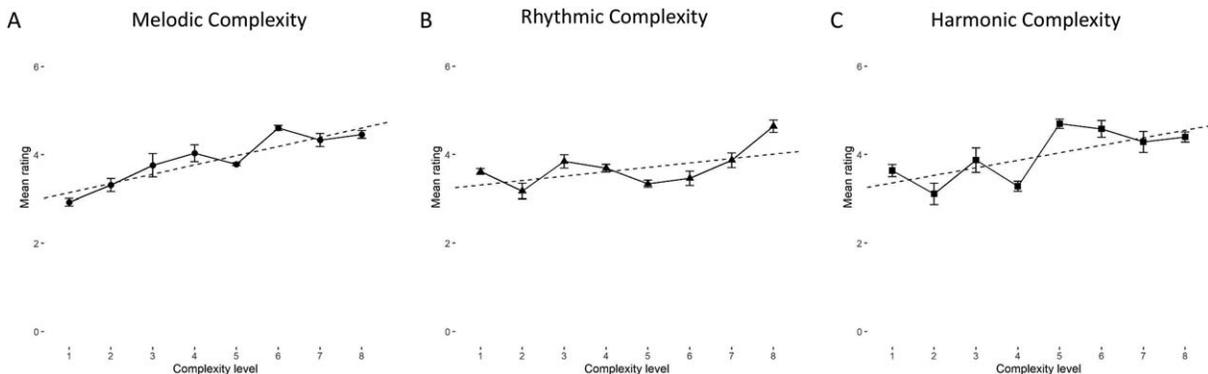

FIGURE 3. Mean complexity ratings are plotted by complexity level for each musical parameter: melodic complexity (A), rhythmic complexity (B) and harmonic complexity (C). Dotted lines illustrate lines of best fit. Error bars illustrate standard error of the mean.

TABLE 2. *Summary of the Fixed Effects Manipulation Model and Fixed Effects Interaction Model*

Manipulation model			
Predictor	Coefficient	*SE*	R^2
(Intercept)	2.97	0.14	–
Complexity	0.16	0.01	.41
Melody	0.19	0.11	.02
Harmony	0.27	0.11	
Version B	-0.13	0.12	.00
Version C	-0.02	0.12	
Version D	0.07	0.12	
Interaction model			
Predictor	Coefficient	*SE*	R^2
(Intercept)	3.11	0.09	–
Complexity: Melody	0.18	0.23	
Complexity: Rhythm	0.12	0.23	.45
Complexity: Harmony	0.18	0.23	

Note: Total R^2 = .43 and .45 respectively ($p < .0001$). Each factor of parameter is in relation to rhythm; and each factor of version is in relation to version A.

to the previous model with main effects yielded little improvement in fit to the data, $F(2, 87) = 2.52$, $p = .08$, $F^2 = 0.83$.

For the single trial analysis, the addition of *music training* marginally improved a model without it, $\chi^2(1) = 3.58$, $p = .05$, and the resulting correlation to the data was $r = .27$, $t(2686) = 14.68$, $p < .0001$, $R^2 = .06$. A summary of the fixed effects model predicting mean ratings and the complexity and musical parameter interaction model can be found in Table 2.

INFORMATION-THEORETIC MEASURES

Is perceived complexity accurately simulated by information content? Table 3 presents the correlation matrix of perceived complexity and information theoretic complexity values. The model included the following predictors: *mean melodic information content*, *mean rhythmic information content,* and *mean harmonic information content*, based on the mean IC of the outer voices for these parameters (see *Complexity Measures* for details). As the range of mean information content for these predictors varies (mean melodic IC range = 2.35–7.62; mean rhythmic IC range = 1.27–3.04; mean harmonic IC range = 3.47–9.10), each predictor was transformed into z-scores (mean = 0, *SD* = 1) so that the mean melodic IC range became −1.42–3.48, mean rhythmic IC range became −0.52–3.49, and mean harmonic IC range became −0.75–2.00. Mean melodic IC of the middle voice was also included as a predictor but rhythmic IC was not included because it does not vary.

TABLE 3. *Correlation Matrix Between Participant Complexity Ratings and Information-theoretic Measures of Melodic, Harmonic and Rhythmic Complexity.*

	Participant Ratings	Melody IC	Rhythm IC	Harmony IC
Participant Ratings	–			
Melody IC	.43 **	–		
Rhythm IC	.06	−.27 *	–	
Harmony IC	.44 **	.32 *	.24 *	–

Note: Ratings and measures are transformed into z-scores prior to correlation test.
$^*p < .05$ $^{**}p < .001$

Melodic, rhythmic, and harmonic IC were significant predictors of mean complexity ratings, but only melodic and harmonic IC were significant when compared to a null model, $F(1, 94) = 11.41$, $p = .001$, $F^2 = 0.09$ and $F(1, 94) = 35.09$, $p < .0001$, $F^2 = 0.35$, respectively, while rhythmic or middle voice IC were not, $F(1, 94) = 0.34$, $p = .91$, $F^2 = 0.00$ and $F(1, 94) = 0.15$, $p = .69$, $F^2 = 0.00$, respectively. The resulting model accounted for a large proportion of the variance in the complexity ratings, $R^2 = .54$, $F(4, 91) = 30.68$, $p < .0001$, $F^2 = 1.20$ and $r = .75$.

To further assess the independent contributions of melodic and harmonic IC, models leaving out one of these predictors at a time were fitted. A model with melodic and rhythmic IC only yielded an R^2 of .08 while a model with harmonic and rhythmic IC yielded an R^2 of .34.

For the single trial analysis, the addition of *music training* marginally improved the model, $\chi^2(1) = 3.58$, $p = .05$, and the resulting correlation to the data was $r = .26$, $t(2686) = 14.35$, $p < .0001$, $R^2 = .06$. A summary of the fixed effects model predicting mean ratings can be found in Table 4.

DESCRIPTIVE FEATURES

What musical properties correspond to the variations in perceived complexity? In other words, how can the differences in complexity perception be characterized in musical terms? For this model, the following features were calculated: the *mean interval size* (in semitones; range 0.41–8.41) of the two outer voices of each stimuli; the *mean pitch* (MIDI note numbers; range 72.14–78.14 for the upper voice and 43.85–51.00 for the lower voice) of each of the two outer voices separately; the *mean note duration* (in ms, where 1 beat = 24 ms; range 16.00–27.42) of the two outer voices; the proportion of out-of-key notes in relation to the total number of pitches in the two outer voices, *key proportion* (range 0–0.38); and

TABLE 4. *Summary of the Fixed Effects Information Content Model*

Predictor	Coefficient	SE	R^2
(Intercept)	3.86	0.04	–
Melody IC	0.27	0.04	.09
Rhythm IC	0.22	0.04	.00
Harmony IC	0.43	0.04	.45
Middle voice IC	0.02	0.04	.00

Note: Total R^2 = .54.

TABLE 5. *Summary of the Fixed Effects Musical Properties Model*

Predictor	Coefficient	SE	R^2
(Intercept)	3.86	0.04	–
Key proportion	0.40	0.05	.34
Mean Duration	-0.41	0.13	.07
Syncopation Score	-0.23	0.12	.01
Mean Interval Size	-0.04	0.06	.00
Mean Pitch – Upper voice	-0.03	0.05	-.01
Mean Pitch – Lower voice	0.12	0.07	.02

Note: Total R^2 = .43

TABLE 6. *Summary of the Fixed Effects EV_3 Model*

Predictor	Coefficient	SE	R^2
(Intercept)	8.50	1.45	-
Tonal ambiguity	1.19	0.28	.18
Pitch proximity	0.14	0.05	.04
Rhythmic variability	-0.22	0.70	.00

Note: Total R^2 = .22

syncopation score, where a lower score equates to more syncopation (Lerdahl & Jackendoff, 1983). All predictors were scaled to have mean = 0 and standard deviation = 1. *Mean duration* and *key proportion* were significant predictors in this model, though only *key proportion* was significant when tested against a null model, $F(1, 94) = 50.63$, $p < .0001$, $F^2 = 0.52$. *Mean duration* makes a significant contribution only when the other predictors are specified in the model, $t(89) = -3.21$, $p = .001$. The model overall has a total R^2 of .43 and $F(6, 89) = 12.96$, $p < .0001$, $F^2 = 0.75$ and $r = .68$.

For the single trial analysis, the addition of *music training* marginally improved the model, $\chi^2(1) = 3.59$, $p = .05$, and the resulting correlation to the data was $r = .27$, $t(2686) = 14.75$, $p < .0001$ and $R^2 = .06$. A summary of the fixed effects model can be found in Table 5.

Finally, a fixed effects model with the three feature-based components from Eerola's (2016) EV_3 model was tested. The model overall has an R^2 of .22 and $F(3, 92) = 10.04$, $p < .0001$, $F^2 = 0.28$ and $r = .49$. Tonal ambiguity and pitch proximity were significant predictors, $F(1, 94) = 21.94$, $p < .0001$, $F^2 = 0.22$ and $F(1, 94) = 10.97$, $p = .0013$, $F^2 = 0.10$, while rhythmic variability was not, $F(1, 94) = 0.70$, $p = .40$, $F^2 = 0.00$. A summary of this model can be found in Table 6.

Discussion

In order to test the hypothesis that information-theoretic predictability can account for perceived musical complexity, complexity ratings were collected for a series of stimuli manipulated in terms of melodic, rhythmic, and harmonic information content as calculated by IDyOM. The results support the hypothesized relationship: both the categorical experimental conditions, which are based on information content, and raw information content itself successfully predicted mean ratings, explaining 43% and 54% of the variance in the data respectively. When random effects are added to account for individual and stimulus differences, the fixed effects lose much of their predictive power, but

the resulting models do not explain more variance in the data. This will be discussed in more detail below. While the information content model demonstrates a correlational relationship between information content and complexity, the explicit manipulation of information content in the experimental design crucially also provides causal evidence. The coefficients and the relative R^2 of these models can be interpreted to draw some conclusions.

First, the intercepts are both close to the centre of the rating scale, indicating a slightly lower than mid-scale baseline rating. For the model based on experimental manipulations, the relationship between the *complexity* predictor and mean ratings is positive, as well as the relationship between both factor levels of the *musical parameter* predictor and mean ratings, indicating that ratings increase with complexity level and are higher overall for both melodically and harmonically manipulated groups of stimuli. The first effect is clearly visible in Figure 3, while the latter effect suggests that listeners consider unpredictable harmonic progressions to be more complex than unpredictable melodies (e.g., large leaps or out-of-key notes), and these more complex than unpredictable rhythms. The differences in ratings between parameters differ slightly in pattern rather than in magnitude although the explanatory power of parameter is negligible compared to complexity level and the interaction between complexity and musical parameter did not improve the fit of the model over and above the main effects. As shown in Figure 3, at

complexity levels five and six, all parameters diverge from the expected increasing pattern. Ratings in the harmonic parameter are especially high while ratings in the rhythmic parameter are especially low and ratings in the melodic parameter are low for level five but high for level six. This pattern mirrors the jump in harmonic information content from level four to five and from level five to six in melodic information content (see Figure 3), which illustrates how the greater level of detail provided by information content over complexity level yields a more accurate model of perceived complexity. Finally, *version* was not a significant predictor, supporting the assumption that the four versions of each stimulus would be rated similarly.

The information content model provides converging evidence for the interpretation of the above predictor parameter, where harmonic IC carries the most explanatory power, followed by melodic IC and finally rhythmic IC, with the same pattern of magnitude seen in their coefficients. Compared to the complexity levels, the more detailed measure of complexity provided by information content accounted for a greater proportion of variance in the complexity ratings. To summarize thus far, the manipulation of rhythmic complexity had a lower impact on perceived complexity ratings such that rhythmic information content accounted for a negligible proportion of the variance in the ratings while manipulations of harmonic and melodic complexity had a significant impact on perceived complexity.

Several raw musical properties of the stimuli were also examined for potential predictive power to characterize the musical features corresponding to differences in perceived complexity. Mean interval size for both outer voices, mean pitch of each outer voice, mean note duration for both outer voices, the proportion of out-of-key notes among both outer voices, and the degree of syncopation of both outer voices were calculated to represent melodic, rhythmic, and harmonic dimensions of music. The results revealed an effect of note duration and proportion of out-of-key notes on the perceived complexity ratings. Both are in the expected direction, where higher proportion of out-of-key notes leads to higher complexity ratings, and longer note duration (which characterized stimuli in the lower levels of complexity) results in slightly lower complexity ratings. The importance of the out-of-key versus in-key proportion is consistent with the two previous models where experimental manipulations and information content are predictors: the unexpected harmonic progressions contain more out-of-key pitches, which is also reflected in melodic information content since these have low probability in context.

There was, perhaps surprisingly, no effect of music training on ratings, where it might be expected that increased exposure to music would yield better models and therefore lower information content, and lower perceived complexity. However, all participants were fairly musically sophisticated, all scoring more than 50% on the Gold-MSI music training sub-scale and perhaps there was not enough variation to detect such an effect. Additionally, when random effects for participants were included in the mixed effects models, this explained the majority of the variance accounted for by the model, 4% in each case. Random effects on the stimuli accounted for 2% of the variance in these models, the second largest contributing predictor. Together, the random effects explain all (though overall little) variance accounted for by the mixed effects models. Thus, it can be concluded that individual differences supersede any effects of experimental manipulation, information content, musical features, or music training on ratings. However, when results are averaged across participants and more general patterns are considered, these same experimental manipulations, measures of information content, and musical features explain up to half the variance in the data in each case, up to as much as 54% in the case of information content. It would be useful to replicate these results using a larger sample of participants varying more widely in music training. As Margulis (2016) raises in her commentary on Eerola's (2016) work, it would also be worth conducting a similar study with participants from different cultural backgrounds, as this is known to be an important influence on music perception, including musical complexity (Eerola, Himberg, Toiviainen, & Louhivuori, 2006). The contribution of random effects on stimuli also highlight the specific nature of the stimuli used here; however, due to the relatively small percentage of variance explained and the strong predictive power of measures of information content across participants, the relationship between information content and perceived complexity is expected to generalize to other, more ecological stimuli. Testing the generalizability of the relationship between information content and perceived complexity would be very valuable future work.

We asked participants to rate their subjective experience of complexity without measuring other factors such as stylistic familiarity or aesthetic appeal. Since approximately half of the variance of the data are left unexplained by our analytical models, it is possible that these factors may have contributed to listeners' complexity ratings. However, a recent study replicating the present finding that IDyOM's information content accounts well for perceived complexity (Clemente

et al., 2019) lends confidence that the present results are not artefactual. Clemente et al. used carefully controlled and stylistically congruent melodic stimuli, though they are in some ways more limited than those used here, making the two studies complementary in conceptually replicating the finding. Furthermore, a pair of studies have recently addressed the relationship between IDyOM's information content and aesthetic appreciation of music (genuine melodies from a range of styles in the case of (Gold, Pearce, Mas-Herrero, Dagher, & Zatorre, 2019), and pop harmonies in the case of (Cheung et al., 2019), both of which find evidence for non-linear relationships between information-theoretic complexity and pleasure (cf. Berlyne, 1974).

The task demands of independently manipulating melodic, harmonic, and rhythmic structure according to monotonically varying degrees of information content meant that the stimuli lack ecological validity in certain ways. While the hypothesis tested in the present research does require that the stimuli possess specific musical features or adhere to specific music-theoretic or stylistic principles, we outline these potential limitations here. First, in the melodic dimension, many of the leaps present in the upper levels of complexity are not idiomatic in Western tonal music. In the rhythmic dimension, the stimuli are less stylistically unusual; however, the prevalence of repeated pitches and the increase in syncopation as rhythmic complexity levels increase could be considered somewhat non-idiomatic. Furthermore, in the harmonic dimension, the challenge of creating eight distinct levels of complexity in four-chord progressions in only two measures of music makes it very difficult, if not impossible, to create ecologically valid stimulus variations. This means that the harmonic style of these stimuli does not perfectly match the Montreal Billboard Corpus used for training the harmonic complexity model in IDyOM, but the fact that the harmony model, trained on the Billboard corpus, accounted for a significant proportion of the variance in the complexity ratings suggests that it provides an accurate simulation of the perception of the stimuli even if it cannot predict the stimuli themselves very accurately. It is also worth noting that the harmony model does not explicitly account for voice-leading, which may make additional contributions to perceived complexity that are not capture by the IDyOM simulation. Finally, as noted previously, it is impossible to completely separate melodic and harmonic aspects of music, leaving great scope for future research to explore more accurate ways to model human perception of these linked musical parameters.

It is important to replicate these findings in real-world musical stimuli, which would provide additional evidence of the link between information-theoretic measures of melodic, harmonic, and rhythmic predictability, and perceived complexity. This experimental approach has the advantage of being more ecologically valid than that followed in the present experiment but the disadvantage of being more correlational and less controlled. At the other extreme, it would be possible to achieve very high experimental control with completely artificial, stylistically unfamiliar stimuli at the expense of ecological validity (e.g., Loui et al., 2010). It is also possible to design new stimuli that balance the two, such as two-voice stimuli where harmony is implied or longer stimuli that can follow idiomatic musical patterns. The experimental approach used here achieves a balance of ecological validity with experimental control, but future research should corroborate the relationship between information-theoretic predictability and perceived complexity with other experimental designs that have more extreme complementary advantages and disadvantages in terms of this balance.

Finally, tonal ambiguity, pitch proximity, and rhythmic variability as defined in Eerola's (2016) EV_3 model were tested on these new data. These predictors explained 22% of the variability in the data. This is not particularly surprising as these parameters might be expected to capture similar variability as the key proportion, pitch interval, and mean duration metrics of the musical features model, which together explained 41% of the variance in the data. It is possible that the nature of the stimuli led to these differences: the current stimuli are short and relatively extreme in term of complexity variation, while the EV_3 model was originally tested on a mixture of specially constructed experimental stimuli and folk music containing many more melodies that vary more subtly in terms of complexity. The EV_3 model may therefore generalize to melodic music better than the musical features model. While the current data were better explained by experimental manipulations and information content as computed by IDyOM, model comparisons should continue to be made to test the generalizability of perceptual models of perceived complexity, including models based on musical features. At the very least, a feature-based analysis is helpful in providing an interpretation of complexity that can be understood in musical terms.

In summary, we find that the results demonstrate a strong causal link between information content and perceived complexity. Specifically, we find that harmonically manipulated stimuli, melodic information content and out-of-key notes—all closely related—have the largest influence on complexity ratings. This is followed by

melodic manipulations and harmonic information content and finally rhythmic manipulations and information content, where other melodic and rhythmic musical features have negligible influence. While different types of measures of complexity have been proposed (Eerola, 2016; Narmour, 1992; Vuust & Witek, 2014), this is the first time that a causal relationship between information content generated by IDyOM and perceived complexity has been empirically tested. This information-theoretic measure of predictability can serve current and future research by providing a general-purpose, quantitative indicator of perceived musical complexity that can be applied consistently to more than one musical parameter.

Author Note

Correspondence concerning this article should be addressed to Sarah Sauvé, Division of Community Health and Humanities, Faculty of Medicine, Memorial University of Newfoundland, St. John's, Newfoundland and Labrador, A1C 5S7, Canada. E-mail: sarah.sauve@mun.ca

References

Agres, K., Abdallah, S., & Pearce, M. (2017). Information-theoretic properties of auditory sequences dynamically influence expectation and memory. *Cognitive Science*, *42*, 43–76. https://doi.org/10.1111/cogs.12477

Balkwill, L.-L., & Thompson, W. F. (1999). A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music Perception*, *17*, 43–64. https://doi.org/10.2307/40285811

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Begleiter, R., El-Yaniv, R., & Yona, G. (2004). On prediction using variable order Markov models. *Journal of Artificial Intelligence Research*, *22*, 385–421.

Berlyne, D. E. (1974). *Studies in the new experimental aesthetics: Steps toward an objective psychology of aesthetic appreciation.* Oxford, England: Hemisphere.

Birbaumer, N., Lutzenberger, W., Rau, H., Braun, C., & Mayer-Kress, G. (1996). Perception of music and dimensional complexity of brain activity. *International Journal of Bifurcation and Chaos*, *6*(02), 267–278. https://doi.org/10.1142/S0218127496000047

Bunton, S. (1997). Semantically motivated improvements for PPM variants. *The Computer Journal*, *40*, 76–93. https://doi.org/10.1093/comjnl/40.2_and_3.76

Burgoyne, J. A., Wild, J., & Fujinaga, I. (2011). An expert ground truth set for audio chord recognition and music analysis. *ISMIR*, *11*, 633–638.

Burke, M. J., & Gridley, M. C. (1990). Musical preferences as a function of stimulus complexity and listeners' sophistication. *Perceptual and Motor Skills*, *71*, 687–690. https://doi.org/10.2466/pms.1990.71.2.687

Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008). Moving on time: Brain network for auditory-motor synchronization is modulated by rhythm complexity and music training. *Journal of Cognitive Neuroscience*, *20*, 226–239. https://doi.org/10.1162/jocn.2008.20018

Cheung, V. K. M., Harrison, P. M. C., Meyer, L., Pearce, M. T., Haynes, J.-D., & Koelsch, S. (2019). *Uncertainty and surprise jointly shape musical pleasure and modulate amygdala activity.* Manuscript submitted for publication.

Clemente, A., Vila-Vidal, M., Pearce, M. T., Aquiló, G., Corradi, G., & Nadal, M. (2019). *A set of 200 musical stimuli varying in balance, contour, symmetry, and complexity: Behavioral and computational assessments.* Manuscript submitted for publication.

Conklin, D., & Witten, I. H. (1995). Multiple viewpoint systems for music prediction. *Journal of New Music Research*, *24*, 51–73. https://doi.org/10.1080/09298219508570672

Eerola, T. (2016). Expectancy-violation and information-theoretic models of melodic complexity. *Empirical Musicology Review*, *11*, 2–17. https://doi.org/10.18061/emr.v11i1.4836

Eerola, T., Himberg, T., Toiviainen, P., & Louhivuori, J. (2006). Perceived complexity of western and African folk melodies by Western and African listeners. *Psychology of Music*, *34*, 337–371. https://doi.org/10.1177/0305735606064842

Egermann, H., Pearce, M. T., Wiggins, G. A., & McAdams, S. (2013). Probabilistic models of expectation violation predict psychophysiological emotional responses to live concert music. *Cognitive, Affective, and Behavioral Neuroscience*, *13*, 533–553. https://doi.org/10.3758/s13415-013-0161-y

GINGRAS, B., PEARCE, M. T., GOODCHILD, M., DEAN, R. T., WIGGINS, G., & MCADAMS, S. (2016). Linking melodic expectation to expressive performance timing and perceived musical tension. *Journal of Experimental Psychology. Human Perception and Performance, 42*, 594–609. https://doi.org/10.1037/xhp0000141

GOLD, B. P., PEARCE, M. T., MAS-HERRERO, E., DAGHER, A., & ZATORRE, R. J. (2019). *Predictability and uncertainty in the pleasure of music: A reward for learning?* Manuscript submitted for publication.

HANSEN, N. C., & PEARCE, M. T. (2014). Predictive uncertainty in auditory sequence processing. *Frontiers in Psychology, 5*(1052). https://doi.org/10.3389/fpsyg.2014.01052

HARRISON, P. M., & PEARCE, M. T. (2018). Dissociating sensory and cognitive theories of harmony perception through computational modeling. In R. Parncutt & S. Sattmann (Eds.), *Proceedings of ICMPC15/ESCOM10* (pp. 194–199). Graz, Austria: Centre for Systematic Musicology.

HURON, D. (2006). *Sweet anticipation: Music and the psychology of expectation*. Cambridge, MA: MIT Press.

LARGE, E. W., FINK, P., & KELSO, S. J. (2002). Tracking simple and complex sequences. *Psychological Research, 66*, 3–17. https://doi.org/10.1007/s004260100069

LERDAHL, F., & JACKENDOFF, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.

LOUI, P., WESSEL, D. L., & KAM, C. L. H. (2010). Humans rapidly learn grammatical structure in a new musical scale. *Music Perception, 27*, 377–388.

MACKAY, D. J. (2003). *Information theory, inference and learning algorithms*. Cambridge, UK: Cambridge University Press.

MADSEN, S. T., & WIDMER, G. (2007, January). A complexity-based approach to melody track identification in MIDI files. *Proceedings of the International Workshop on Artificial Intelligence and Music*. Hyderabad, India: IWAIM.

MARGULIS, E. H. (2016). Toward a better understanding of perceived complexity in music: A commentary on Eerola (2016). *Empirical Musicology Review, 11*, 18–19. http://dx.doi.org/10.18061/emr.v11i1.5275

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

NARMOUR, E. (1990). *The analysis and cognition of basic melodic structures: The implication-realization model*. Chicago, IL: University of Chicago Press.

NARMOUR, E. (1992). *The analysis and cognition of melodic complexity: The implication-realization model*. Chicago, IL: University of Chicago Press.

NORTH, A. C., & HARGREAVES, D. J. (1995). Subjective complexity, familiarity, and liking for popular music. *Psychomusicology: A Journal of Research in Music Cognition, 14*, 77. http://dx.doi.org/10.1037/h0094090

OMIGIE, D., PEARCE, M. T., & STEWART, L. (2012). Tracking of pitch probabilities in congenital amusia. *Neuropsychologia, 50*, 1483–1493. https://doi.org/10.1016/j.neuropsychologia.2012.02.034

OMIGIE, D., PEARCE, M. T., WILLIAMSON, V. J., & STEWART, L. (2013). Electrophysiological correlates of melodic processing in congenital amusia. *Neuropsychologia, 51*, 1749–1762. https://doi.org/10.1016/j.neuropsychologia.2013.05.010

PEARCE, M. T. (2005). *The construction and evaluation of statistical models of melodic structure in music perception and composition* (Doctoral dissertation). City University London, United Kingdom.

PEARCE, M. T. (2018). Statistical learning and probabilistic prediction in music cognition: Mechanisms of stylistic enculturation. *Annals of the New York Academy of Sciences, 1423*(1), 378–395. https://doi.org/10.1111/nyas.13654

PEARCE, M., & MÜLLENSIEFEN, D. (2017). Compression-based modelling of musical similarity perception. *Journal of New Music Research, 46*, 135–155. https://doi.org/10.1080/09298215.2017.1305419

PEARCE, M. T., MÜLLENSIEFEN, D., & WIGGINS, G. A. (2010). The role of expectation and probabilistic learning in auditory boundary perception: A model comparison. *Perception, 39*, 1365–1389. https://doi.org/10.1068/p6507

PEARCE, M. T., RUIZ, M. H., KAPASI, S., WIGGINS, G. A., & BHATTACHARYA, J. (2010). Unsupervised statistical learning underpins computational, behavioral, and neural manifestations of musical expectation. *NeuroImage, 50*, 302–313. https://doi.org/10.1016/j.neuroimage.2009.12.019

PRINCE, J. B., THOMPSON, W. F., & SCHMUCKLER, M. A. (2009). Pitch and time, tonality and meter: How do musical dimensions combine? *Journal of Experimental Psychology. Human Perception and Performance, 35*, 1598–1617. https://doi.org/10.1037/a0016456

SAUVÉ, S. A., SAYED, A., DEAN, R. T., & PEARCE, M. T. (2018). Effects of pitch and timing expectancy on musical emotion. *Psychomusicology: Music, Mind, and Brain, 28*, 17–39. http://dx.doi.org/10.1037/pmu0000203

SAUVÉ, S. A., STEWART, L., & PEARCE, M. T. (2014). The effect of music training on auditory grouping. In M. Kyoung Song (Ed.), *Proceedings of the ICMPC-APSCOM 2014 Joint Conference* (pp. 293–296). Seoul, Korea: ICMPC.

SCHELLENBERG, E. G. (1997). Simplifying the implication-realization model of melodic expectancy. *Music Perception, 14*, 295–318. https://doi.org/10.2307/40285723

SEARS, D. R., PEARCE, M. T., CAPLIN, W. E., & MCADAMS, S. (2018). Simulating melodic and harmonic expectations for tonal cadences using probabilistic models. *Journal of New Music Research*, *47*, 29–52. https://doi.org/10.1080/09298215.2017.1367010

SHMULEVICH, I., & POVEL, D.-J. (2000). Complexity measures of musical rhythms. In P. Desain & L. Windsor (Eds.), *Rhythm perception and production* (pp. 239–244). Lisse, Netherlands: Swets & Zeitlinger.

SONG, C., SIMPSON, A. J. R., HARTE, C. A., PEARCE, M. T., & SANDLER, M. B. (2013). Syncopation and the score. *PLoS ONE*, *8*(9), e74692. https://doi.org/10.1371/journal.pone.0074692

TOIVIAINEN, P., & EEROLA, T. (2016). *MIDI Toolbox 1.1.* Available online at https://github.com/miditoolbox/1.1

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*, 1111. https://doi.org/10.3389/fpsyg.2014.01111

ZATORRE, R. J., CHEN, J. L., & PENHUNE, V. B. (2007). When the brain plays music: Auditory–motor interactions in music perception and production. *Nature Reviews Neuroscience*, *8*, 547. https://doi.org/10.1038/nrn2152

Appendix
Scores of excerpts for complex target stimuli for Levels 1-8