AI BASED ALGORITHMS FOR THE DETECTION OF (IR)REGULARITY IN MUSICAL STRUCTURE

LORENA MIHELAČ a, JANZ POVH b, c, *

aSciDrom Scientific Lab
School Center Novo Mesto
Šegova ulica 112, 8000 Novo Mesto, Slovenia
e-mail: lorena.mihelac@sc-nm.si

bFaculty of Mechanical Engineering
University of Ljubljana
Aškerševa ulica 6, 1000 Ljubljana, Slovenia
e-mail: janez.povh@fs.uni-lj.si

cInstitute of Mathematics, Physics and Mechanics
Jadranska ulica 19, 1000 Ljubljana, Slovenia

Regularity in musical structure is experienced as a strongly structured texture with repeated and periodic patterns, with the musical ideas presented in an appreciable shape to the human mind. We recently showed that manipulation of musical content (i.e., deviation of musical structure) affects the perception of music. These deviations were detected by musical experts, and the musical pieces containing them were labelled as irregular. In this study, we replace the human expert involved in detection of (ir)regularity with artificial intelligence algorithms. We evaluated eight variables measuring entropy and information content, which can be analysed for each musical piece using the computational model called Information Dynamics of Music and different viewpoints. The algorithm was tested using 160 musical excerpts. A preliminary statistical analysis indicated that three of the eight variables were significant predictors of regularity ($E_{cpitch}$, $IC_{cpintfref}$, and $E_{cpintfref}$). Additionally, we observed linear separation between regular and irregular excerpts; therefore, we employed support vector machine and artificial neural network (ANN) algorithms with a linear kernel and a linear activation function, respectively, to predict regularity. The final algorithms were capable of predicting regularity with an accuracy ranging from 89% for the ANN algorithm using only the most significant predictor to 100% for the ANN algorithm using all eight prediction variables.

Keywords: regularity, musical structure, perception, AI algorithms.

1. Introduction

1.1. Motivation. Regularity exists in natural and human-made objects, including biology, physics, engineering, architecture, and art, and plays an important role in human life. The detection of repeated structures (patterns) is important, as it governs our recognition and understanding of the world (Pauly et al., 2008). Therefore, finding patterns that are repeated and form a regular structure can help understand and analyze abnormalities in the structure due to some criteria (e.g., unexpected use of chords in the harmonic progression and its impact on the listener enjoyment). Formally, we can define regularity as a subset $X_R$ of the set of all configurations $X$ which have some structure that an observer tends to utilize or recognize (Feldman, 1997, p. 3).

In music, regularity is experienced as a strong structured texture with dominant periodic patterns and strong neighboring relationships (Manjunath et al., 2000), where musical ideas are arranged in a shape appreciable by the human mind (Pole, 2014). Conversely, irregularity is experienced in a non-structured or weakly structured
musical piece, where the relationship between patterns can rarely be detected, and the enjoyment is affected due to the large mental space required for processing the musical content full of novelties (Kramer, 1988).

In our recent study (Mihelač and Povh, 2020), we examined the listener acceptability of music based on the complexity of harmony and proposed three objective measures for complexity: (i) complexity of harmony measured based on the presence of basic tonal functions and parallels in the harmonic flow, (ii) unigram and bigram entropy measuring the complexity in the harmonic progression, and (iii) regularity in terms of the order (regularity) and disorder (irregularity) found in the harmonic progression.

The latter was detected by a human expert (considering also comments of two additional musical experts) based on the presence of particular circumstances in musical excerpts (e.g., the occasional appearance of chords in harmony, non-detectable functions in harmony due to the use of figured chords, emphasized rhythm or melody capturing the attention of the listener and placing the harmony in the background, etc.). Specifically, if at least one of 10 situations described previously by Mihelač and Povh (2020) appears within the harmonic flow, the musical example was labelled as irregular; otherwise, it was regular. A dataset containing 160 musical excerpts with all three complexity measures evaluated, is available at [http://kt.ijs.si/data/DATA_HARMCOMP.zip](http://kt.ijs.si/data/DATA_HARMCOMP.zip).

This process of detecting irregularity has two drawbacks: (i) it demands high expertise of a listener/analyst and (ii) it is difficult to repeat the results (i.e., different experts can differ regarding some marginal examples in the labels that they assign). This motivated us to search for artificial intelligence (AI) algorithms capable of replacing human experts to explain and predict the (ir)regularity of musical excerpts. We built support vector machine (SVM) and artificial neural networks (ANN) classifiers that accept as input eight variables measuring entropy and information content. These variables were selected, because they can be evaluated for each musical excerpt automatically by using the computational model called Information Dynamics of Music (hereafter referred to as IDyOM) and various features, i.e., viewpoints (Pearce, 2005; 2018).

### 1.2. Main contributions

The main contributions of this paper are as follows:

- We supplement the dataset from our other work Mihelač and Povh (2020) with eight new variables (features): IC\_cpitch, IC\_cpint, IC\_cpint\_ref, IC\_cpitch\_dur, E\_cpitch, E\_cpint, E\_cpint\_ref, E\_cpitch\_dur. The first four variables describe different viewpoints of information content, whereas the last four variables describe different viewpoints of entropy. These additional features were computed by IDyOM. The supplemental dataset (HARMCOMP\_2) with MIDI files, and sheet music is available to the research community at [http://kt.ijs.si/data/DATA_HARMCOMP.zip](http://kt.ijs.si/data/DATA_HARMCOMP.zip).

- We analysed the relevance of the new features for irregularity using statistical analysis, including a Mann-Whitney-Wilcoxon test, which showed that the distributions of E\_cpitch, IC\_cpint\_ref, and E\_cpint\_ref in the dataset displayed statistically significant differences regarding the subgroups of regular and irregular music excerpts, therefore suggesting them as good candidates for predicting regularity.

- We built SVM and ANN models to predict regularity based on the eight new features. Using 10-fold cross-validation, we showed an accuracy ranging from 89 % (achieved with an ANN using only the most significant predictor) to 100 % (achieved with an ANN using all eight prediction variables). These results suggest that AI-based algorithms can potentially replace expert-based detection of musical (ir)regularity.

All statistical results were obtained using the software \( R \) version 3.6.0.

### 1.3. Glossary

In this subsection we present several musical terms used in this study that might be unclear to readers unfamiliar with music. *Melody* (Fig. [1](#)) is a horizontal appearance of notes representing a unique ordering of notes from a specific scale (Solomon, 2019, p. 24) and having an organized and recognizable shape. On its lowest level at the music surface (Jackendoff, 2009), it can be defined as a sequence of events (hereafter referred to as \( e_1, e_2, \ldots \)), in which each event \( e_i \) is associated with pitch (highness or lowness; the frequency of a sound), duration (the length of a sound), loudness (intensity; loudness or softness of a sound), and timbre (quality or colour of a sound).

*Harmony* is a vertical presentation of tones (Fig. [1](#)), which represents the structure of music with respect to...
chord composition and progression. A chord (Fig. 1) is a combination of at least three tones (sounds) performed simultaneously. A non-chord tone is a tone that does not belong to the prevalent (predominant) chord (Solomon, 2019).

Key designates a certain pitch as the tonal center of a musical piece and considers other pitches as scale degrees (the position of a note in a scale) around that center. Scale is a collection of discrete pitch relationships (Burns, 1999, p. 215) comprising a pattern of whole and half steps, whereas scale-degree indicates the position of a note within a scale (Solomon, 2019). The keys are simply named by the scale (see Fig. 2) on which they are based, e.g., (key) C major, C minor, etc. (Benward and Saker, 2008). Function in music does not have the same meaning as in mathematics. In music, a function describes the role a chord plays with respect to the root (tonic function) of a key. We can define in each scale (key) three main or basic harmonic functions: Tonic is the first scale degree (I) (i.e., the base of the key), Subdominant is the fourth scale degree (IV) of the key, and Dominant is the fifth scale degree (V) of the key.

All other scale degrees in a scale are defined as parallels (secondary degrees) of the three main functions (Dahllaus, 2014): subdominant parallel (Sp; II degree), dominant parallel (Dp; III degree), and tonic parallel (Tp; VI degree).

Each tone in a melody can imply a certain harmony. In some melodies, the tones can be organized in such a way that merging them together results in a very strong implied harmony (Fig. 3). In this example, the noteheads, presented as crosses, form implied harmonies on the pedal point “G” (presented as noteheads between brackets), a tonic (T) in the first bar, an implied subdominant in the second bar, an implied dominant in the third bar, and again a tonic in the fourth bar.

A pedal point or pedal tone is a bass note (presented as notehead between brackets in Fig. 3), usually the tonic or dominant used through a sequence, including some chords (harmonies), which shifts around it.

In the continuation, we define the texture of the melody. In music, the texture of a melody is how the melodic material is combined in a composition and its effect on the overall quality of the sound in a musical piece (Benward and Saker, 2008). Homophonic music is music comprising melody and harmony. The texture of a melody can be presented in different ways (see Fig. 4). These can have either only one line or two or more lines and be written in a very independent way from the perspective of pitch and rhythm or as an example of chordal melody, when all the voices below the upper line have similar rhythmic material.

2. Scientific background and related work

During the listening process, the listener applies models acquired through the learning of regularities found in musical structure via exposure to music (Pearce, 2018), with this exposure either long-term (e.g., entire lifetime) or short-term (during the listening of a single composition).

During music processing, predictions of music are generated in order to organize and process the perceived musical content (Pearce, 2018). The ability to predict a subsequent event that meets listener expectations is important in the listening process, as it affects (among other things) the aesthetic experience. For example, if musical structure is somehow manipulated (e.g., by composer or performer), it creates in the listener the feeling of enjoyment if the event has happened or disappointment if not, tension if the event is delayed, or the feeling that the structure is ambiguous when the event is missing (Meyer, 1957; Narmour, 1990).

The idea that the structure of the musical content in a musical piece can affect the listener perception of a musical piece was proposed by Meyer in his seminal book Emotion and Meaning in Music (Meyer, 1957), where he outlined how some structures in musical pieces create higher or lower perceptual expectations for subsequent events depending on how the structure is manipulated by a composer.

The structure of a musical piece can be manipulated by deviating the form. Too complicated or even amorphous music, which includes a succession of...
heterogeneous ideas, without any relation between them affects the acceptance of music by a listener, as in the end, music should be structured in an understandable and (to a certain degree) predictable way in order to be properly enjoyed (Pole, 2014). The structure of musical dimensions, such as melody, harmony, and rhythm, can be deviated, as well. An unexpected chord in a harmonic progression, unexpected relationships between two chords in a chord progression that mismatches the rules of musical syntax, unexpected intervals in a melody, or unusual rhythm are examples of deviations (Rohrmeier, 2011; Rohrmeier and Pearce, 2018).

Listeners tend to use a set of basic perceptual principles that are applied to different musical styles and dependent on the type of music to which the listeners are exposed (Krumhansl, 2004). If the content of a musical piece does not meet these principles, it might confuse the listener based on the unusual use of musical dimensions or elements (e.g., unexpected use of pitch, intervals, rhythm, etc.) in the musical content. In such a case, the information provided is unrecognized (Finnas, 1989; Edmonds, 1995), and the regularity, sometimes posited as a mid-point between order and disorder (Grassberger, 2004), is perceived as disorder.

According to Edmonds (1995), the problem of perceiving a structure as regular (ordered) or irregular (disordered) lies in the fact that no language of representation of a structure is provided. From the perspective of music, this suggests that the listener, in the absence of an inherent language, has to impose one.

Music is multidimensional, and musical dimensions are never found in isolation but rather constantly interacting (more or less) with each other (Prince et al., 2009b). The perception of these dimensions depends on how they are presented in the musical shape, which can be vertical, when the relationships between notes are presented simultaneously (e.g., harmony), or horizontal (e.g., melody), when notes are presented in a sequence one after another. A specific example is pitch, which can be presented either vertically (as chord) or horizontally as a sequential presentation of pitches (Loui, 2012). However, vertical and horizontal presentation of music (encompassing all the discussions of what exactly should be considered as vertical or horizontal), when presented together in a musical piece, forms a unit, in which the musical content is deposited (Busch, 1985). These two dimensions are inexorably connected, with horizontal an extension of vertical and vice versa (Williams, 2005). Findings from different studies (Lerdahl and Jackendoff, 1983; Butler and Brown, 1994; Platt and Racine, 1994) suggest that even when only one dimension is presented (e.g., melody), listeners tend to imply structures found in another dimension (e.g., harmony).

According to Sloboda and Parker (1985), each tone in a single melodic line can imply a harmony as a mental model of the underlying structure, with similar findings reported in other studies (Thompson and Cuddy, 1989; Platt and Racine, 1994; Holleran et al., 1995). When melody and harmony are presented together in a musical example (when the vertical and horizontal dimensions are presented together), a harmonic frame is established (Povel and Jansen, 2002), which can have two aspects: a global aspect (key and mode) and a local aspect defined as a region within the key, which is assigned to a harmony and defined as a function (e.g., tonic, subdominant, dominant, etc.). According to Povel and Jansen (2002), a listener establishes first global and then local aspects, although the processes of the establishment of these two aspects is not well understood, as they are usually conceived as hierarchical (Bharucha, 1987; Tillmann et al., 2000). For AI based analysis of recorded speech, see the work of Piotrowska et al. (2019).

The fact that music can be presented vertically and/or horizontally can explain why it is insufficient to capture only one musical dimension (e.g., harmony in a musical piece) when seeking answers what causes higher/lower feelings of (ir)regularity in listeners of music. Findings of Prince et al. (2009b) suggest that the perception of a musical dimension alone can differ from the perception of the same musical dimension when interacting with another dimension. In the latter case, even seemingly small alteration of structure in a particular dimension (e.g., harmony) can affect the the perception of structure in another dimension (e.g., melody). Depending on the stimulus and task properties, the importance of one musical dimension can be magnified (Prince, 2011) depending on the informative value, and a certain dimension, having a greater informative value, is likely to dominate other dimensions (Melara and Algom, 2003; Prince et al., 2009a).

Approaches to measuring musical regularity include evaluation of regularity in music by listeners, which can be very subjective and dependent on long- or short-term exposure and acquisition of formal/informal musical knowledge (Steinbeis et al., 2006; Herbert, 2012). Regardless of the subjective nature of such measurements, they can highlight peculiarities in musical structure, which should be considered when analysing deviations in musical structure (Mihelač and Povh, 2020).

Another example of measuring musical regularity involves mismatch negativity (MMN). Previous studies (Bouwer and Honing, 2012; Bader et al., 2017) indicate that an MMN response depends on the magnitude of the violation (Schröger and Winkler, 1995; Näätänen et al., 2007) and is elicited when the previously established regularity is violated. Because the auditory system is considered predictive in nature (Pearce, 2005; Pearce et al., 2010b; Bouwer and Honing, 2012) and creates expectations based on extracted regularities found in a musical information, any incoming information that does
not match a prediction represents an error signal in the form of an MMN (Bendixen et al., 2009).

Another approach to measuring regularity in musical structure is computational simulation of human perception. By enforcing a set of rules, this method examines how well human perception is captured in order to verify theoretical principles and generate quantitative predictions when applying the model to novel circumstances (Plaut, 2000).

We used the latter approach in the present study to examine the regularity of melody in a dataset comprising 160 musical excerpts. We previously studied this dataset in detail (Mihelač and Povh, 2017; 2019; Mihelač et al. 2018) and showed that the regularity of the harmonic progression affected user perception while highlighting other musical dimensions. Therefore, we expanded our focus to melody in order to elucidate how regularity in melody and the interplay between melody and harmony affects listener perception.

3. IDyOM

3.1. Modelling musical structure with IDyOM.

The computational model IDyOM is based on $n$-gram models, which are frequently used in statistical language modelling. An $n$-gram can be defined as a sequence of $n$ symbols, where the sequence itself can include anything (e.g., characters, words, etc.), with an $n$-gram model which is simply a collection of such sequences. The basic $n$-gram model used in music is the unigram model, where $n = 1$, and the occurrence of each tone (event) in a sequence is counted.

Because the quantity of $n − 1$ is known as the order of the model, this implies that in the case $n = 1$, we have a simple zeroth-order, where each tone is treated as an independent event, i.e., not dependent on the preceding context. For $n = 2$ (a bigram model), the prediction of forthcoming events is governed by a first-order model, and two adjacent tones are considered in the prediction of a forthcoming tone, which depends on the preceding tone.

To capture as much information as possible from musical structure, IDyOM uses a variable-order $n$-gram model. The use of low-order $n$-gram models, such as unigram or bigram models, might not adequately explain the statistical regularity in musical structure or the effect of context on expectations, and the use of higher order models could prevent the capture of statistical regularity, (see, e.g., Wiggins et al., 2009).

IDyOM learns in an unsupervised manner from the musical structure and generates predictions about forthcoming events in musical sequences (Pearce, 2005). Listeners are sensitive to statistical regularities and irregularities in musical structure, which are progressively internalized during long- or short-term exposure to music and then generalized to new musical examples.

IDyOM is capable of simulating both instances, the long- or short-term exposure to music with long- and short-term modelling. Long-term exposure to music is simulated by using the long-term model (LTM), which is trained on a large corpus of music, whereas short-term exposure is simulated using the short-term model (STM), which learns about repeated patterns in a particular musical sequence/piece dynamically (Pearce, 2018). These models have been tested on different tasks and shown to be accurate predictors of melodic expectancy (Pearce and Wiggins, 2006), behaviour, and neural measures (electroencephalography) of melodic expectedness (Pearce et al., 2010c; Agres et al., 2018), as well as accurate identifiers of phrase boundaries (Pearce et al., 2010a; 2010b).

3.2. Viewpoints. IDyOM enables the perception of a sequence of events (e.g., a melody consisting of notes) from different angles. Whenever a sequence of events is given, functions and viewpoints are defined to accept initial sub-sequences of a sequence and select a specific feature (e.g., pitch, duration, relationships between tones, etc.) in the sequence (Pearce and Wiggins, 2012). The two crucial dimensions (viewpoints), in which events in a sequence are described in IDyOM, are pitch (cpitch) and time (dur). IDyOM also offers derived viewpoints, such as cpint (the distance between two pitches) and cpintref, with the latter representing how close/far an event in a sequence is from the tonic. The use of this viewpoint is motivated by the fact that the regularities in pitch relative to the tonic influence the melodic structure (Pearce, 2005; 2018; Arthur, 2018).

According to Volk (2016), there is still a gap existing in the understanding between temporal based dimensions and other musical dimensions. Previous studies (Krumhansl, 2000; Justus and Bharucha, 2003) are treating melodic (pitch-based) and temporal (time-based) relations separately, empirically, and theoretically; however, their independence from the perspective of processing has been questioned (Jones and Boltz, 1989; Boltz, 1999; Griffiths et al., 1999), in that melody and rhythm are perceived as a unified dimension by listeners. Therefore, we have decided to link two viewpoints together to create a compound of the viewpoints representing a cross-product of two viewpoints, viewpoint cpitch$dur$ in the present study, because it remains unclear whether these two dimensions can be considered separately or unified.

The decision to use the viewpoint cpint to examine intervals and relationships between two adjacent tones is based on the fact that pitch relationships evoke a particular scale and affect the feeling of stability between scale tones, as all scale tones are not equivalent from the perspective of their importance and are hierarchically organized depending on how distant or closely related
they are to the tonic (Tillmann et al., 2000; Peretz and Zatorre, 2005).

Another option involves exploring a sequence and its events from multiple viewpoints (viewpoint selection in IDyOM) using flexible views of abstract musical objects. In this case, a hill-climbing procedure is used that combines different viewpoints and identifies the combination of source viewpoints in order to minimize the mean information content of the dataset (Pearce, 2005).

In the present study, we used four viewpoints cpitch, cpint, cpintfref, and cpitchdur. These were labelled in our dataset and in the results from Section 4 as ICcpitch, ICcpint, ICcpintfref, and ICcpitchdur for information content and as Ecpitch, Ecpint, Ecpintfref, and Ecpitchdur for entropy.

### 3.3. Entropy and information content. IDyOM uses a complex methodology to compute estimates of probabilities of an event to appear in the sequence that we are considering. It is encapsulated into a lossless data compression algorithm PPM*, an improved version of PPM (Prediction by Partial Match), originally introduced by Cleary and Witten (1984). The classic PPM algorithm, where the maximum context length is a fixed constant, compresses sequences of symbols, one by one, and learns gradually about context-dependent conditional probability distributions (Steinruecken et al., 2015). An improved variant of this sequence prediction model, the PPM* algorithm, able to process contexts of unbounded length compared (Cleary et al., 1995), has been used in IDyOM, and combined with interpolated smoothing and update exclusion, and long- and short-term models as well (Pearce, 2005). This methodology has been well explored and justified in the last three decades. The reader can find many details in the dissertation by Pearce (2005), in the works of Gold et al. (2019). IDyOM computes the corresponding sequence of values for the probability, entropy, and information content as Ecpitch, Ecpint, Ecpintfref, and Ecpitchdur for entropy.

The information-theoretic measures used in IDyOM are entropy (H) and information content (IC). Shannon’s entropy (Shannon, 1948) is used as baseline theory for quantifying the uncertainty in the prediction of a musical event before it is heard, using specific viewpoint. If $X_i$ is a set of all possible continuations of a given musical event $e_i$, IDyOM first computes probability $p_i$ that event $e_i$ happens, using the very complex methodology roughly explained in Section 3.1. Then for each $x \in X_i$ it computes probabilities $p_x$ that $e_i$ will continue with $x$.

Once it has $p_i$ and $\{p_x \mid x \in X_i\}$, it computes

$$H(e_i) = - \sum_{x \in X_i} p_x \log_2(p_x), \quad (1)$$

$$IC(e_i) = - \log_2(p_i), \quad (2)$$

The methodology to compute $H(x_i)$ and $IC(x_i)$ is nicely depicted in Fig. 1 from Gold et al. (2019).

Figure 5 shows 14 events in the well-known melody for The X-Files, a musical excerpt included in the dataset examined in this study and used to examine the probability, information content, and entropy of each event.

Using the viewpoint cpitchdur, which examines the feature pitch and duration of each event in the sequence, IDyOM computes the corresponding sequence of values for the probability, entropy, and information content (Table 1 and Fig. 6 as cpitchdur, depicted as a dashed line).

As shown in Table 1 the events found to be highly expected by the model were events $e_{10}$, $e_{11}$, and $e_{14}$ according to their high probability values and low information content. Conversely, events found to be the most unexpected were $e_7$ and $e_{12}$, with very low probabilities and high information content.

When analyzing the information content in all four chosen viewpoints in the same musical excerpt, Fig. 6 shows that the highest and lowest information contents were obtained with the viewpoints cpitch (black full
4. Exploring and predicting regularity with entropy and information content

4.1. Data. The dataset used in our investigations was introduced by Mihelač and Povh (2020) and comprises 160 musical examples, including 141 popular musical pieces and 19 classical music examples, collected and evaluated for their complexity, musical style, and acceptability (Mihelač and Povh, 2020). These musical examples were shortened to musical excerpts from 14 s to 18 s in duration, resulting in 8 to 12 bars. In each musical excerpt, chords in the harmonic progression were located, and entropy (unigram and bigram) was calculated. All of the musical excerpts were labeled as “regular” or “irregular” based on clear criteria which patterns must be present in the excerpt to label it as irregular. When the situation was clear, the label was assigned by the main evaluator, while in ambiguous situations, two additional experts were involved. These values of (ir)regularity were taken as a ground truth in the present research.

For the purpose of this study, we extracted melodies from all these 160 musical excerpts in order to obtain pure monophonic musical excerpts and using only the very first upper lines in all of the melodies. This was done, because some of the melodies in this dataset were written in a polyphonic or chordal manner (see Section 1.3), which would make the analysis of multiple lines in these melodies beyond the scope of this study. The new dataset, with 4692 events in total and an average of 29.33 events per excerpt, comprises all the variables used in the previous study, and additionally it contains eight new variables based on entropy and information content. These eight variables represent mean values of ICpitch, ICpint, ICpintfref, ICpitchdur, Epitch, Epint, Epintfref, and Epitchdur computed for the musical excerpts using computational model IDyOM (for details, see Sections 1.1 and 3.2).

We summarise the distributions of all eight viewpoints in Table 3 which contains the values for min, median, arithmetic mean and max for each viewpoint, separately for regular and irregular excerpts.

4.2. Classification models for regularity. In this subsection, we propose two AI algorithms to predict regularity, as introduced by Mihelač and Povh (2019), which take the information content and entropy variables as inputs and are trained on the dataset from that work (Mihelač and Povh, 2019).

First, the principal component analysis (PCA) was performed on the variables ICpitch, ICpint, ICpintfref, ICpitchdur, Epitch, Epint, Epintfref, and Epitchdur. Figure 7 shows that regular and irregular musical excerpts were linearly separated into a 2-dimensional subspace spanned by the first two principal components, which motivated a search for classification models promoting the linear separation.

Before developing classification models, we explored the relevance of the eight predictive variables using a Mann–Whitney–Wilcoxon test (R script wilcox.test) to determine which variables showed significantly different distributions on the subgroups of
Table 3. Main characteristics of distributions for all eight viewpoints, separately for regular and irregular excerpts.

<table>
<thead>
<tr>
<th>Viewpoint</th>
<th>Regular</th>
<th></th>
<th></th>
<th></th>
<th>Irregular</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>min</td>
<td>med</td>
<td>mean</td>
<td>max</td>
<td>min</td>
<td>med</td>
<td>mean</td>
<td>max</td>
</tr>
<tr>
<td>E_{cpitch}</td>
<td>1.78</td>
<td>3.28</td>
<td>3.17</td>
<td>6.10</td>
<td>1.63</td>
<td>3.48</td>
<td>3.33</td>
<td>5.62</td>
</tr>
<tr>
<td>E_{cpint}</td>
<td>2.61</td>
<td>3.28</td>
<td>3.26</td>
<td>4.04</td>
<td>2.33</td>
<td>3.45</td>
<td>3.44</td>
<td>4.32</td>
</tr>
<tr>
<td>IC_{cpitch}</td>
<td>1.71</td>
<td>3.10</td>
<td>3.11</td>
<td>6.33</td>
<td>1.69</td>
<td>3.02</td>
<td>3.13</td>
<td>5.71</td>
</tr>
<tr>
<td>IC_{cpint}</td>
<td>2.49</td>
<td>3.12</td>
<td>3.12</td>
<td>4.01</td>
<td>2.56</td>
<td>3.06</td>
<td>3.05</td>
<td>3.56</td>
</tr>
<tr>
<td>IC_{cpintfref}</td>
<td>3.37</td>
<td>4.39</td>
<td>4.33</td>
<td>5.55</td>
<td>3.66</td>
<td>4.81</td>
<td>4.73</td>
<td>5.74</td>
</tr>
<tr>
<td>E_{cpintfref}</td>
<td>4.08</td>
<td>4.58</td>
<td>4.56</td>
<td>5.00</td>
<td>4.54</td>
<td>5.04</td>
<td>5.01</td>
<td>5.28</td>
</tr>
<tr>
<td>IC_{cpitch_dur}</td>
<td>1.59</td>
<td>3.38</td>
<td>3.37</td>
<td>6.10</td>
<td>1.28</td>
<td>3.51</td>
<td>3.50</td>
<td>6.84</td>
</tr>
<tr>
<td>E_{cpitch_dur}</td>
<td>2.30</td>
<td>3.39</td>
<td>3.40</td>
<td>4.61</td>
<td>2.30</td>
<td>3.60</td>
<td>3.52</td>
<td>4.42</td>
</tr>
</tbody>
</table>

Fig. 7. First two principal components obtained by PCA demonstrated linear separation of the regular and irregular excerpts specifically due to the second principal component.

regular and irregular musical excerpts, according to the definition of regularity by Mihelač and Povh (2019). This test was performed because the assumptions requiring the use of two samples t-test were not met. Table 4 shows that the variables E\_{cpitch}, IC\_{cpintfref}, and E\_{cpintfref} had significantly different distributions on the sets of regular and irregular examples. Therefore, they were natural candidates for building blocks of the classification models. Figure 8, which depicts distributions of E\_{cpintfref} on regular and irregular excerpts, additionally supports the decision to use this variable as a feature in classification models.

We built classification models using subsets of one to eight variables, where in subsets with k variables, the k most relevant variables were used (i.e., the k variables with the lowest p-values in Table 4) which means that the model with one prediction variable included only variable E\_{cpintfref}, the model with two prediction variables included E\_{cpintfref}, IC\_{cpintfref}, etc.).

We used SVM and ANN as classification methods. For the SVM, we used the R library caret and function train with a linear kernel. The training part and evaluation of the models were done using 10-fold cross-validation.

For the ANN, we used the R package neuralnet and the function with the same name. Additionally, we used the logistic activation function. The 10-fold cross-validation was used to build and evaluate the models. Other input settings for neuralnet function were set to default, see the manual (Fritsch et al., 2019) for details about default settings.

Regarding hidden layers, we tested ANN with no hidden layer, with one hidden layer (having 1 to 4 neurons) and with 2 hidden layers (having 4 neurons each). The results obtained with zero hidden layers were already very good and increasing the number of layers and the number of neurons did not improve the models significantly (in terms of accuracy and Cohen’s Kappa), whereas the computational complexity increased significantly. Therefore, we decided to keep and report the results for ANN with no hidden layers. This actually means that we could replace ANN with logistic regression.
In this study, we used 160 musical excerpts previously used to evaluate the effect of harmony on musical acceptability (Mihelač, 2017; Mihelač and Povh, 2017; 2020; Mihelač et al. 2018; 2019). Given that our focus here was on melody, we extracted from each musical excerpt only the first upper line in order to obtain 160 monophonic musical excerpts, which were then used to evaluate eight variables, including the pitch, interval, scale degree, and duration of the information content (IC\_cpitch, IC\_crint, IC\_crintfref, and IC\_cpitch\_dur) and the entropy (E\_cpitch, E\_crint, E\_crintfref, and E\_cpitch\_dur) by adding them to the original dataset.

To identify irregularities in 53 of the 160 musical excerpts, which was detected during the previous study and were taken as ground truth in the present study, we analyzed the relevance of the new features using the Mann–Whitney–Wilcoxon test, revealing that three variables (IC\_cpitch, IC\_crintfref, and E\_crintfref) differed significantly between regular and irregular musical excerpts.

The significant difference between regular and irregular musical excerpts demonstrated by IC\_crintfref and E\_crintfref, suggests the presence of implied harmonies, and confirms the salience of implied harmonies in the perception of music in homophonic music (Sloboda and Parker, 1985; Thompson and Cuddy, 1989; Platt and Racine, 1994; Holleran et al., 1995).

Applying the concepts of “global” and “local” establishment of the harmonic frame from Povel and Jansen (2002) to previous and new data revealed that while listening to a particular monophonic musical excerpt, listeners first generate a key and mode, after which implied harmonies are “created” for each note. In some cases, these implied harmonies do not “fit” in the existing harmonic framework (according to the rules of the harmonic syntax), which is (re)created when the same melody is combined with its underlying harmony.

Therefore, when “horizontal” (melody) and “vertical” (harmony) musical content are presented together, a “fusion” of different tones occurs (Parncutt, 1989; Huron, 2001) to generate different harmonies. These harmonies can be emphasized depending on the information value either in melody or harmony, which could explain the feeling of higher complexity in musical excerpts with a simple harmonic progression (e.g., T-D-T), as the focus is placed on the melody and its harmonies. This agrees with previous findings reported by Melara and Algol (2003) or Prince et al. (2009a).

The significance of the variable E\_cpitch can be explained by the higher degree of diversity in pitches and the higher number of non-chordal tones found in irregular musical excerpts. Specifically, non-chordal tones appear to affect the listener enjoyment of music, as well as the expectation for forthcoming events, according
to our previous results (Mihelač et al., 2018; Mihelač and Povh, 2020). The successful harmonic analysis of a musical piece by a listener is clearly dependent on how successfully the non-chordal tones are resolved and assigned to a harmony and how the tones distributed in a melody are perceived as an implied harmony, which agree with findings from a previous study (Povel and Jansen, 2002).

In our previous work (Mihelač and Povh, 2020), we identified 10 peculiarities found in irregular musical excerpts and based on the musical expertise of three experts according to their evaluation and perception of complexity. In the present study, we used two classification methods (SVMs and ANNs), with both algorithms using 10-fold cross-validation to confirm a high level of accuracy (>97%) in predicting (ir)regularity in our dataset. These results indicated that expert-based detection of (ir)regularity in musical structure can be replaced by AI algorithms.

Because only eight variables were included in the analysis of regularity of musical structure in this study, future work should focus on analysis of additional viewpoints (variables) in IDyOM. This was previously found useful in the perception of musical structure and prediction of forthcoming events (Pearce, 2005), with additional study potentially offering deeper insight into the (ir)regularity of musical structure. Additionally, the approaches used in the present study to analyse (ir)regularity could be applied to other datasets, especially to those comprising non-complex/less complex musical examples (e.g., children’s songs, children’s folk songs, folk songs, etc.), and used as an additional clarification of the listener’s acceptance/rejection of musical pieces/genre.

Acknowledgment

This research work was supported by Šolski Center Novo Mesto, Slovenia, and by the Slovenian Research Agency (ARRS) through the research project J1-8155 and the research program P2-0256.

References


Lorena Mihelač currently works as a VET professor and a supervisor in the SciDrom Scientific Lab at the School Center Novo Mesto, Slovenia. She holds a BA from the University of Music in Zagreb, Croatia, an MA from the Academy of Arts in Novi Sad, Serbia, as well as a BSc and an MSc from the Faculty of Information Studies of the School Center Novo Mesto, Slovenia. She also holds a PhD from the Postgraduate School ZRC SAZU, Ljubljana, Slovenia. Now she is finishing her second PhD at Jožef Stefan International Postgraduate School (Information and Communication Technologies) in Ljubljana.

Her research covers different areas, such as music education, music psychology, music therapy, and lately modelling of human music perception. She has published many educational textbooks, several chapters in monographs, journal papers, and numerous works in conference proceedings.

Janetz Povh is an associate professor and a senior researcher at the Faculty of Mechanical Engineering, University of Ljubljana. He holds BSc, MSc and PhD degrees in mathematics from the University of Ljubljana. His research activities vary from pure mathematical optimization topics, like developing new methods for (non-commutative) polynomial optimization problems, to more applied areas such as applications of mathematical optimization methods and tools in data science and complex networks analysis. He has published 35 journal papers (17 in top mathematical optimization journals), numerous works in conference proceedings, several book chapters and a book with Springer.

Received: 7 January 2020
Revised: 4 May 2020
Accepted: 20 June 2020