Music Visualisation and its Short-term Effect on Appraisal Skills

Ali Nikrang[1,2][0000−0002−2453−6091], Maarten Grachten[3], Martin Gasser[4], Harald Frostel[5], Gerhard Widmer[6][0000−0003−3531−1282], and Tom Collins[7,8][0000−0001−7880−5093]

[1] Ars Electronica Futurelab, Linz, Austria
[2] University of Music and Theatre, Munich, Germany
[3] Independent Machine Learning Consultant, Cologne, Germany
[4] Muse Group, Limassol, Cyprus
[5] Independent Machine Learning Consultant, Linz, Austria
[6] Institute of Computational Perception, JKU, Linz, Austria
[7] University of Miami, Coral Gables, USA
[8] MAIA, Inc., Davis, USA
tomthecollins@gmail.com

Abstract. Music psychologists have long been concerned with phenomena such as repetition and tonality – both the internal representations formed by exposure to these phenomena, and how these representations vary with expertise. A question arises of whether less expert listeners can gain proficiency in perceiving a particular phenomenon by being exposed to representations that are known to be employed by more expert listeners. The current paper addresses this question within the domain of music appraisal. Participants with varying levels of musical expertise interacted with visualizations of two excerpts from Beethoven's symphonies. One visualization (ScoreViewer) showed the staff notation of the music, synchronized to an orchestral recording. The other (PatternViewer) also depicted the notes synchronized to the recording, as well as representations of the music's repetitive and tonal structure. Participants' appraisal skills were assessed via multiple-choice questions on instrumentation, dynamics, repetition, and tonality. Results indicated that interacting with the PatternViewer visualization led to a significant improvement in listeners' appraisal of repetitive and tonal structure, compared to interacting with the ScoreViewer. The size of this effect was well predicted by amount of formal musical training, such that less expert listeners exhibited larger improvements than more expert listeners. While further work is required to determine whether the observed effects transfer beyond the pieces studied or into long-term learning, these findings for appraisal skills indicate that carefully chosen representations from models of expert behavior can, in turn, help less expert individuals to improve their understanding of musical phenomena.

Keywords: Music · Visualisation · Appraisal · Cognition · Learning.

1 Introduction

Whether at the concert hall, in front of a computer, or holding a phone, we may question: what does one look at when listening to music, and does it make any difference to our perception of the music?

Some listen to music with their eyes shut. In a live setting or music video, some people study the performers' movements, others the setting itself, and some even study fellow audience members. When listening to recorded music, computer/phone applications also offer a choice of visualisation modes. An uncompressed sound (or audio) file typically consists of 44,100 samples (values between -1 and 1) per second, which displace the membrane of the speaker(s) and give the listener the impression of hearing the music as it was originally recorded. Some visualisations operate on this data too to provide animations that appear to be synchronised with the music as heard. A screenshot of such a visualisation is given in Fig. 1A.

The demand for and popularity of looking whilst listening to music is difficult to overestimate: a visualisation application by [19] called the *Music Animation Machine*, which plays an audio file synchronised to a symbolic representation of the music (i.e., the note information present in staff notation), has over 30,000,000 views on YouTube.[9] Notes appear as horizontal oblongs in the Music Animation Machine, as shown in Fig. 1B. Time runs along the x-axis such that the starting and stopping time of each note corresponds to an oblong's left- and right-most extremities, respectively. Thus a note's duration corresponds to the oblong's horizontal extent. The pitch height of each note is represented by the oblong's height on the y-axis, with lower-sounding notes appearing lower down on this axis. The representation is known as *piano roll*, because holes punched in paper rolls according to the same (x, y)-axes were used to drive an autonomous musical instrument called the *player piano*, which reached peak popularity in the early twentieth century (see Fig. 1C).

Other visualisation apps offer a higher level of interactivity and have a stronger didactic emphasis. For instance, the *Orchestra App* offers multi-perspective video recordings of orchestral concerts, interviews with the conductor and performers, information about orchestral instruments, audio-synchronised staff notation and piano roll, and more (see Fig. 1D). The user is able to switch between these display modes whilst the piece is playing. Despite the abovementioned visualisations, for many, the cultural facade of Western classical music remains imposing and impenetrable. Some listeners do not derive pleasure from listening to the music and/or they are unable to *appraise* it, by which we mean *to gain an explicit understanding of the inner workings of a piece*. Examples of music appraisal include identification of: orchestral instruments and their interplay; changes in dynamics and tempo, such as *crescendo* and *accelerando*; short repeated musical ideas called *motives*, which build to make *themes* and in turn larger *sections*; tonal structure such as changes in key (the note collections used by the composer over the course of a piece).

[9] `https://www.youtube.com/user/smalin/`

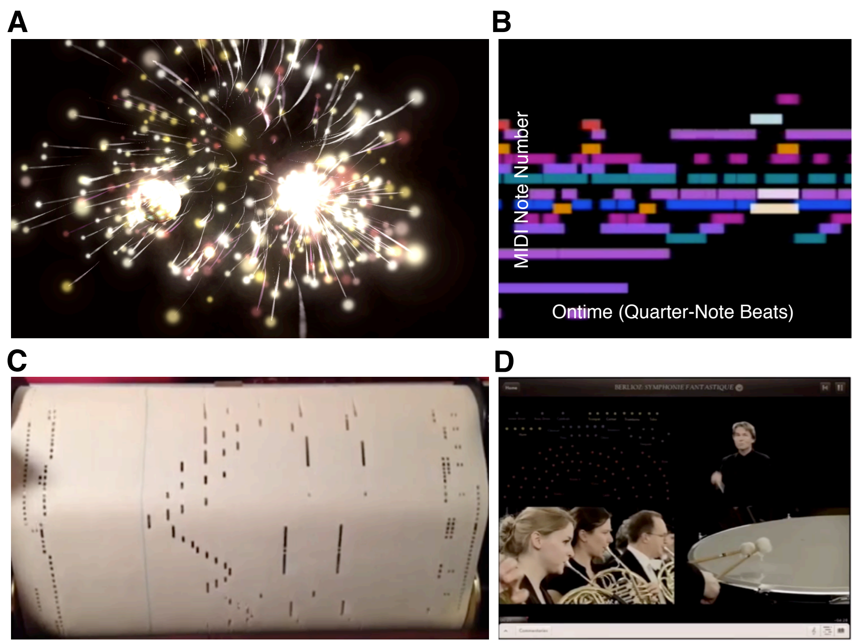

Fig. 1. A) Screen capture of a visualisation mode available with Apple's iTunes software; (B) Piano-roll representation of the Music Animation Machine [19], with axes added to indicate note on and offtimes (x-axis) and MIDI note number (y-axis); (C) An actual piano roll, where for mechanical reasons the x- and y-axes are usually flipped compared with Figure 1B; (D) Screen capture of The Orchestra app, showing multi-perspective videos sychronised to the music audio.

As relatively little is known about the effect of attending to visualisations while listening to music, the purpose of this paper is to investigate whether an interactive visualisation application, based on 40 years' worth of psychological research into listeners' cognition of repetitive and tonal structure, is capable of enhancing music appraisal skills. We report an experiment where listeners interacted with two visualisation applications in order to answer questions that were adapted from music-appraisal examinations. The music excerpts come from the symphonies of Ludwig van Beethoven (1770-1827). One visualisation application, called the *ScoreViewer*, presents the staff notation of the music synchronised to audio; the other, called the *PatternViewer*, is also synchronised to audio and presents a piano-roll representation such as in Fig. 1B, as well as representations of repetitive and tonal structure.

The remainder of the paper is organised as follows. The ScoreViewer and PatternViewer are introduced, the next sections describe the setup and results of an experiment investigating the effect of visualisation applications on music appraisal skills, and a final discussion section sets our findings in the wider context of interacting with and understanding intellectual artifacts, and the catalytic role that experimental-psychological research can play in this process.

2 ScoreViewer and PatternViewer

2.1 ScoreViewer

ScoreViewer is a web-based application that shows the staff notation (score) of the music, synchronised automatically to an orchestral recording [13]. At the top of the ScoreViewer, there is a media control bar for play/pause and skipping to different parts of the score/audio file (see Fig. 2). The current bar is always indicated by a semi-transparent gray box. ScoreViewer is a system for listening to orchestral music accompanied by the score. In the current paper it acts as a kind of control condition for normal music listening with score, while the second visualisation presents listeners with extra information that might help them to appraise the inner workings of a piece.

2.2 PatternViewer

The second visualisation is called PatternViewer.[10] Like ScoreViewer, it also depicts the notes (in piano-roll notation) synchronised to the recording. Additionally, it contains representations of the music's repetitive and tonal structure. These representations are based on [11]'s model of cognition of repetitive structure and [17]'s model of tonal cognition.

There are five elements to the PatternViewer's display, as shown in Fig. 3:

1. As with the ScoreViewer, the PatternViewer contains a media control bar for play/pause and skipping to different parts of the piece. The control bar

[10] `https://tomcollinsresearch.net/research/PatternViewer/`

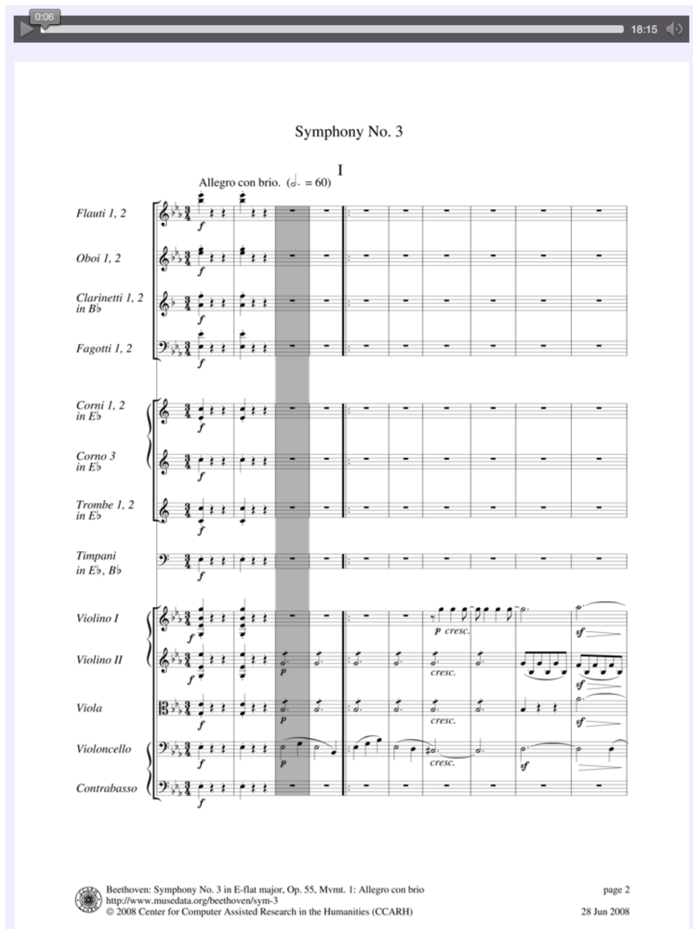

Fig. 2. The Scoreviewer visualisation. An audio file is embedded towards the top of the browser page. Below this is the staff notation corresponding to the point in the audio file, with the current bar indicated by a gray box. In this example the beginning of the main theme from Beethoven's Symphony no.1 first movement is highlighted.

appears in the top-right corner, and can be clicked or dragged to navigate to different time points;
2. In the top panel is a global-scale, piano-roll representation of the piece, which is synchronised to the audio. Notes light up in the piano-roll as they are heard in the audio. Different colours indicate changing key estimates (see point 4 for further explanation). There are two reasons for the choice of piano-roll representation as opposed to ordinary staff notation. First, it is simpler in piano-roll for listeners to follow the notes currently playing in a piece [19]. Second, it is easier to render piano-roll events in different colours, which is useful for conveying key estimates;
3. In the bottom-right corner is a local-scale, piano-roll representation of active (currently playing) pattern occurrences (see point 5 for further explanation).
4. In the bottom-left corner is a coloured circle of fifths, which provides a reference for the key estimates shown in the global-scale piano-roll representation. For instance, the notes currently playing in Fig. 3 are coloured red, which represents the key of C major (red segment labeled C in the coloured circle of fifths). A modern version of Newton's colourwheel [20], as used in experiments on visual cognition [28,29], was a natural choice for the reference. The circle of fifths, in itself, is not a music-psychological representation but a music-theoretic diagram. [18] showed that the important hierarchical relations between pitch classes and keys enshrined in the circle of fifths are a psychological reality, however, and later [17] specified a method for estimating key based on empirical key profiles. Therefore, as it appears in the PatterViewer, the circle of fifths and perceived key estimates represent a model of tonal cognition.
5. In the top-left corner is a pendular graph, where each node represents a note collection that repeats (called pattern occurrence). As the piece plays, nodes in this graph will appear red when a pattern occurrence is active. The user has the option to hear/see the pattern occurrence that a node represents—by clicking on the nodes in the graph, the audio and piano-roll representations will skip to and play the pattern occurrence automatically. The pendular graph in Fig. 3 is inspired by [11], and therefore the pendular graph models the cognition of repetitive structure.

In a pendular graph, a node's horizontal (x-axis) position corresponds to the time in the piece at which the pattern occurrence begins [7]. (If there are a lot of occurrences in a short period of time, nodes may be shifted slightly to the right.) All members of a specific pattern occupy the same vertical (y-axis) position. Vertically, different pattern occurrences are arranged in ascending order of duration. Therefore, large repeated sections tend to appear higher up and have labels nearer the start of the alphabet, whereas motives and themes tend to appear lower down in these graphs and have later alphabetic labels. If an edge (line) joins the pattern occurrence X_i to the higher pattern occurrence Y_j, this means that all the notes belonging to X_i belong also to Y_j. In Fig. 3, the nodes A_1 and C_1 are active (red). Node C_1 represents the first theme in the symphony movement, and its note content is displayed in the local-scale piano-roll repre-

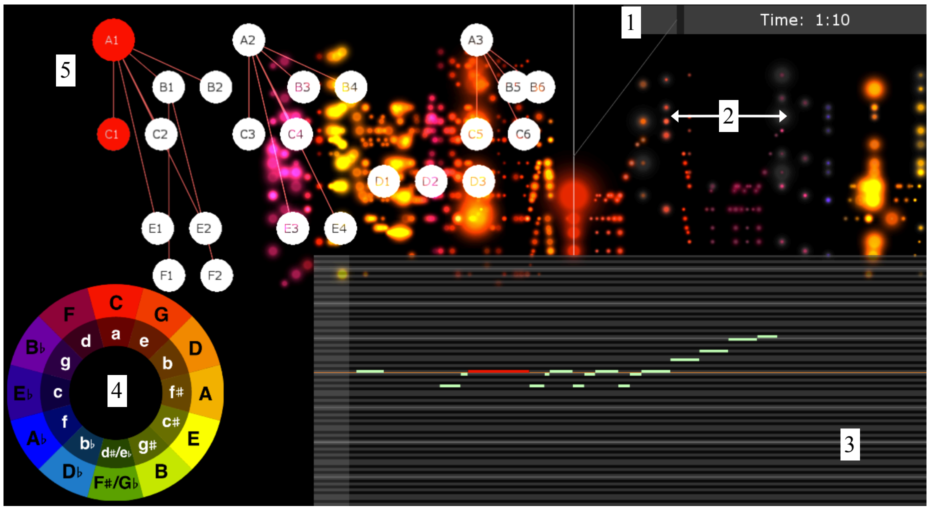

Fig. 3. The PatternViewer visualisation. Top left: a pendular graph representing the hierarchical, repetitive structure of the current piece. For instance, C_1 is the first occurrence of the main theme, whose notes are a subset of the large repeated section A_1 known as the exposition. Top right: a global-scale point-set representation, coloured by the current key of the music. A white vertical bar indicates the position within the audio, which can be moved back or forward by clicking to the left or right. Bottom left: a coloured circle of fifths, which can be used to identify the key colour shown in the top right (currently C major). Bottom right: a local-scale piano-roll representation, showing the contents of a pattern occurrence from the top-left graph (currently the main theme C_1).

sentation in the bottom-right corner. In total, there are six occurrences of the theme in this movement, labeled $C_1, C_2, \ldots, C_6$ in the pendular graph. Node A_1 represents a large repeated section. The first movement of this symphony follows the conventional *sonata form*, in which A_1 is the *exposition* section. The edge (line) connecting C_1 to A_1 indicates that this first theme occurrence is part of the exposition section (as is the second theme occurrence C_2). When the exposition repeats (A_2), there are subsequent occurrences of the theme (C_3 and C_4).

In summary, the ScoreViewer and PatternViewer provide interactive , audio-synchronised visualisations of a piece of music. Additionally, the PatternViewer contains representations of the piece's repetitive and tonal structure, which are based on prominent models of music cognition [11,18].

3 Experimental hypotheses

The main hypothesis that we investigate concerns how appraisal of the repetitive and tonal structure of a piece of music varies with use of the ScoreViewer and PatternViewer visualisations:

- **Null hypothesis.** There is no signfiicant difference in terms of listeners' appraisal of repetitive and tonal structure, when using the ScoreViewer or PatternViewer visualisation.
- **Alternative hypothesis.** Interacting with the PatternViewer visualisation leads to a significant improvement in listeners' appraisal of the repetitive and tonal structure of a piece of music, compared to interacting with the corresponding, audio-synchronised staff notation (as in the ScoreViewer).

A secondary hypothesis centers on two aspects of music that might still be appraised effectively by studying the staff notation—instrumentation and dynamics. The following hypothesis is intended to balance the first hypothesis, allowing for the possibility that the PatternViewer is not always preferable to staff notation for appraising music:

- **Null hypothesis.** There is no significant difference in terms of listeners' appraisal of instrumentation and dynamics, when using the ScoreViewer or PatternViewer visualisation.
- **Alternative hypothesis.** Interacting with the ScoreViewer visualisation leads to a significant improvement in listeners' appraisal of the instrumentation and dynamics of a piece of music, compared to interacting with the PatternViewer.

4 Method

4.1 Participants

We recruited eighteen students from Johannes Kepler University Linz (mean age = 26.3 years, SD = 3.8, see also Fig. 4A), via the general student email list, offering €20 compensation, and mentioning that an interest (but not proficiency) in classical music was desirable. Thus our participants had varying levels of musical expertise, as shown in Fig. 4B-D. Mean years of formal musical training (instrumental or vocal) was 6.0 (SD = 4.7, see Fig. 4B). We also collected data on participants' musical activities, such as regularity of playing an instrument or singing (Fig. 4C) and concert attendance (Fig. 4D).

4.2 Stimuli

During the experiment, participants heard two excerpts from Beethoven's symphonies. The first excerpt was from Symphony no.1 in C major op.21, the opening 6:40 of the first movement (bars 1–225). The second excerpt was from Symphony

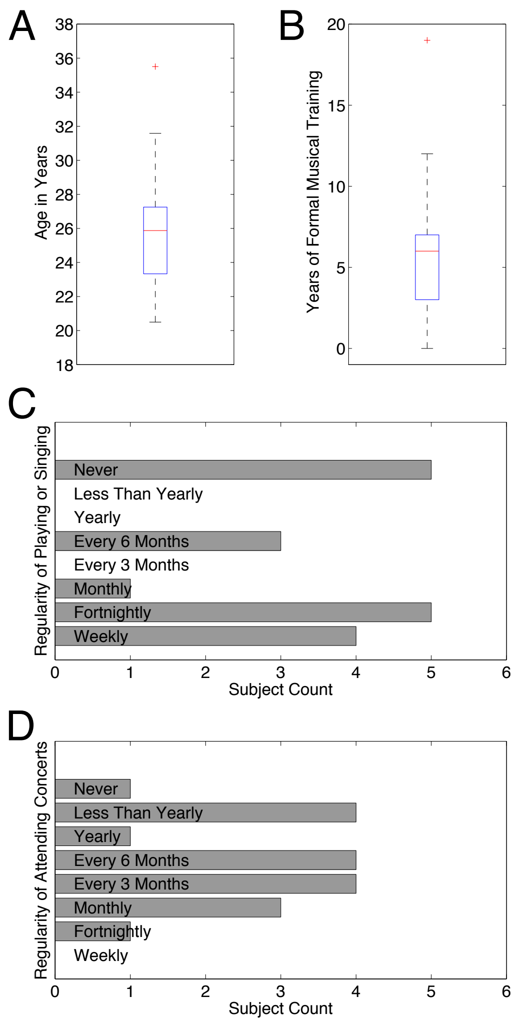

Fig. 4. Age, formal music training, and musical activities of the listeners in the study. One mature student is the cause of the outliers in Figs. 4A and B.

no.3 in E♭ major op.55, the opening 6:40 of the first movement (bars 1–158).[11] The excerpts will be referred to hereafter as Excerpt A and Excerpt B respectively. They were chosen to be matched in terms of tempo (Allegro con brio for both, following a short Adagio molto introduction in Excerpt A), overall key (major), and familiarity level amongst the listening public. The excerpts were the same length to avoid the possibility of participants listening to one excerpt more times than the other. In both cases the opening 6:40 encapsulated the statement and repetition of the large repeated section known as the exposition. Excerpt A included some of the next sections as well (development and partial recapitulation), but most questions focused on the exposition. Thus the overall form of the music heard by participants was similar also.

4.3 Procedure

In order to test our hypotheses, we need to operationally define *music appraisal skills*. Two sets of twenty questions were devised (one set for Excerpt A, the other for Excerpt B). The questions were based on past examination papers on appraising music from the UK GCSE Music qualification [1,12,21,27]. Each of the four main exam boards offer a Music qualification, which is taken as an elective course by students aged 15-16 years. Among other areas, the syllabi for these courses identify (1) instrumentation, (2) dynamics (including tempo), (3) repetitive structure, and (4) tonal structure as being relevant to the appraisal of music, and so our twenty questions were constructed by devising five questions from each of these four categories. Each question was multiple-choice, with three possible answers. An example question for each category is given below:

1. **Instrumentation.** Which of the following instruments is playing in the time window 1:46-1:49?
2. **Dynamics.** Which time window contains the most powerful (loudest) occurrence of the main theme?
3. **Repetitive structure.** Typically the first movement of a symphony begins with a large section called the exposition, which is repeated. In the current piece, at what time does the repeat of the exposition begin?
4. **Tonal structure.** To which key is there a brief modulation at 2:58?

For the purposes of analysis, the four question categories exemplified above are grouped into two topics: (A) instrumentation and dynamics (hereafter, Instr.Dynam or Topic A); (B) tonal and repetitive structure (hereafter, Reptn.Tonal or Topic B). These question categories and topics were selected on the basis that Instr.Dynam questions may be easier to answer with the ScoreViewer, where the names of the instruments and dynamic markings are stated (see Fig. 2), whereas Reptn.Tonal questions may be easier to answer with the PatternViewer, where representations of repetitive and tonal structure are provided (see Fig. 3). The

[11] Both were from recordings of the Royal Concertgebouw Orchestra. The first was conducted by David Zinman and the second by Iván Fischer.

topics, showing equal favor to the two visualisation applications, enable our experimental hypotheses to be investigated in a balanced manner.

After giving informed consent, receiving task instructions, and undertaking some training to gain familiarity with the layouts of the ScoreViewer and PatternViewer, a participant was presented with Excerpt A and the corresponding question set. To prevent ordering effects, the twenty questions appeared in a random order. After answering these questions, the participant was presented with Excerpt B and its corresponding question set (also in a randomised order). Thirty minutes were allocated for listening to and answering questions on each excerpt.

We used a within-subjects design that was counterbalanced across visualisation condition. That is, all participants heard Excerpt A first, followed by Excerpt B. One of two visualisation conditions (ScoreViewer or PatternViewer) was assigned to Excerpt A, and the other to Excerpt B, such that half of participants interacted with Excerpt A via the ScoreViewer, and half interacted with Excerpt A via the PatternViewer (and similarly for Excerpt B). Therefore, conditions were counterbalanced across pieces, and any learning effect that caused participants to perform better on Excerpt B than A would not bias our comparison of the ScoreViewer and PatternViewer.

At any stage, participants were able to ask questions of clarification, they were able to revise answers, and they were encouraged not to worry if they were unsure of the answer to a question. Instructions and questions were presented in English via a web interface, which contained links to Wikipedia and Youtube when helpful (e.g., definition of a tonal sequence or example of a clarinet sound). There was also an experimenter on hand to provide German translations of terms when necessary.

4.4 Apparatus

Participants sat at a laptop in an $12\,\mathrm{m}^2$ room, and read and answered the question sets on this machine. The laptop was connected to a projector, which showed the visualisations on a $2 \times 3\,\mathrm{m}$ wall section. The audio was presented by stereo speakers at a level of $-20\,\mathrm{dB}$.

5 Results

All participants had above-chance performance on the task, so no data were excluded from the following analyses. A two-way within-subject ANOVA on accuracy, with factors for Topic and Visualization type, revealed a significant main effect of Topic ($F_{1,17} = 22.64,\ p < .001$, see also Fig. 5), and a significant interaction effect of Topic and Visualization ($F_{1,17} = 6.84,\ p < .05$, indicated by the lines having different gradients in Fig. 5). The signifcant interaction suggests that the effect of Topic on accuracy varies as a function of Visualisation, with Reptn.Tonal questions being answered more accurately in the PatternViewer and Instr.Dynam questions being answered more accurately in the ScoreViewer.

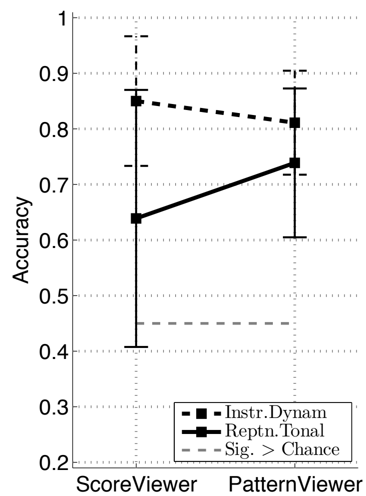

Fig. 5. Plot of accuracy varying with Topic (Instr.Dynam or Reptn.Tonal) and Visualization type (ScoreViewer or PatternViewer). Chance performance is indicated at $18/40 = .45$ by the dashed gray line. Under a binomial distribution with parameters $n = 40$ trials (questions) and $p = 1/3$ chance of success (multiple-choice question with three options), the probability of scoring more than 18 correctly is less than .05.

To investigate the interaction effect more thoroughly, a planned comparison was performed on the accuracy of Reptn.Tonal answers in the PatternViewer versus Reptn.Tonal answers in the ScoreViewer. The result of a one-sided, paired t-test was significant at the .05 level ($t(17) = 1.77$, $p < .05$). Thus there is evidence for rejecting our main, null hypothesis (that there is no signfiicant difference in terms of listeners' appraisal of repetitive and tonal structure, when using the ScoreViewer or PatternViewer visualisation), in favor of the alternative hypothesis—using the PatternViewer visualisation leads to a significant improvement in listeners' appraisal of the repetitive and tonal structure of a piece of music, compared to interacting with the corresponding, audio-synchronised staff notation (as in the ScoreViewer).

Further analysis reveals that the size of this effect is not uniform across participants, but varies as a function of musical training. Figure 6A shows Reptn.Tonal accuracy in the PatternViewer minus Reptn.Tonal accuracy in the ScoreViewer for each participant, plotted against years of formal musical training. The plot suggests that the effect size is a function of musical training, with participants' accuracy on Reptn.Tonal questions being higher in the PatternViewer than in the ScoreViewer (positive difference) if they have received approximately five years or less of formal musical training. If they received more than five years of musical training, the effect is either small or small in the opposite direction (negative difference).[12] Restricting the comparison of accuracy on Reptn.Tonal answers to those participants having five years or less of formal musical training, a one-sided, paired t-test is more significant still ($t(7) = -3.12,\ p < .01$).

Switching focus to the secondary hypothesis, concerning how participants' appraisal of instrumentation and dynamics varies with visualisation, a planned comparison was performed on accuracy of Instr.Dynam answers in the PatternViewer versus Instr.Dynam answers in the ScoreViewer. The result of a one-sided, paired t-test was not significant at the .05 level ($t(17) = -1.51,\ \text{ns}$), suggesting we cannot reject the null hypothesis of no signfiicant difference in terms of listeners' appraisal of instrumentation and dynamics, when using the ScoreViewer or PatternViewer visualisation.

Again, we performed a subsequent analysis to investigate whether the size of this effect is uniform across participants or varies as a function of musical training. Figure 6B shows Instr.Dynam accuracy in the PatternViewer minus Instr.Dynam accuracy in the ScoreViewer for each participant, plotted against years of formal musical training. The distribution appears to be more even about zero than in Fig. 6A. For the Reptn.Tonal differences in Fig. 6A, it was possible to identify an exponential decay curve, $y = \alpha + \beta \exp \gamma x$, which provided a significant fit to the observations, where x is years of formal musical training, y is the Reptn.Tonal accuracy difference between PatternViewer and ScoreViewer, and α, β, γ are parameters estimated from the data. A corresponding analysis for the Instr.Dynam differences in Fig. 6B did not lead to idenfitifcation of any signficant linear or exponential trends. Thus the intuitive inspection that the size of the visualisation effect for Instr.Dynam does not vary as a function of musical training is supported by the regression analysis results.

6 Discussion

Of long-standing interest in psychology and neuroscience is how structured information in our environment passes through low-level and high-level representations, being transformed by as well as transforming the mind and brain. Two eminent examples of this type of information are repetitive and tonal structures in music, with listeners being exposed to music over many years, and behavioural experiments revealing that implicit sensitivities develop to these aspects of music's structure, even in the absence of formal training [24,26]. While pieces of music, like other cultural artifacts, will always be open to different interpretations by different individuals, the models of repetitive and tonal structure proposed by [11] and [18] respectively are particularly parsimonious. Therefore, we investigated whether presenting these parsimonious representations to listeners can help them to appreciate the inner workings of a piece, more readily than this

[12] Only one participant was at ceiling for repetitive/tonal questions in the ScoreViewer, and two other participants were at ceiling for repetitive/tonal questions in the PatternViewer, so these comparisons are not being confounded by ceiling effects.

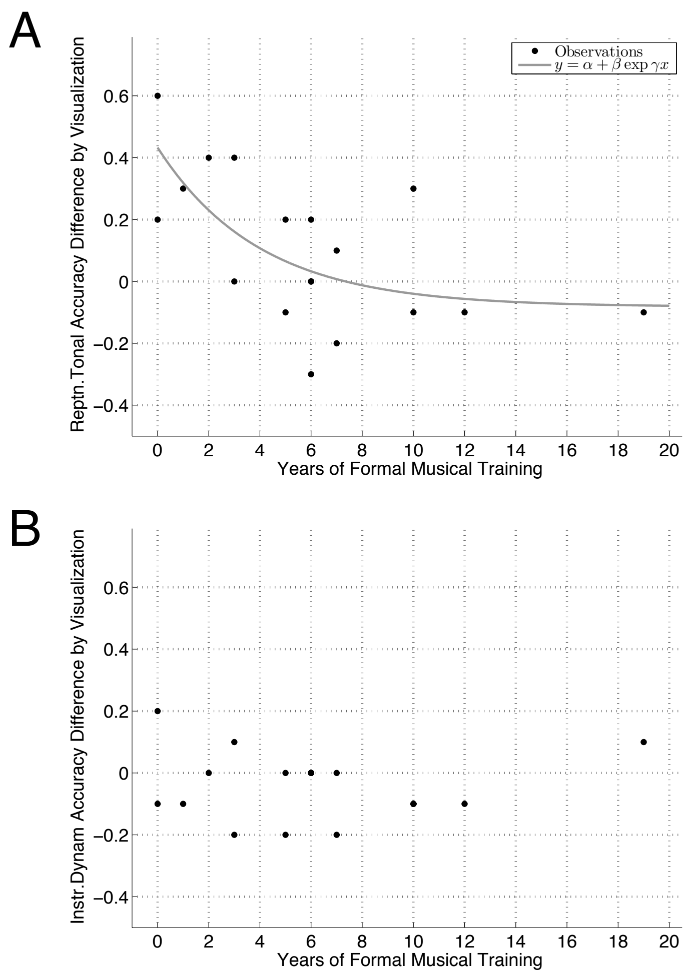

Fig. 6. (A) Plot of Reptn.Tonal accuracy in the PatternViewer minus Reptn.Tonal accuracy in the ScoreViewer against years of formal musical training. Parameters for the curve $y = \alpha + \beta \exp \gamma x$ were identified by iterating over choices for γ and performing linear regression to select optimal values of α and β. The regression of accuracy difference on years of training was significant, $F(1, 16) = 12.31, p < .01, s = .19, R^2 = .43$, and gave parameter values $\alpha = -0.08,\ \beta = 0.51,\ \gamma = -0.25$; (B) Plot of Instr.Dynam accuracy in the PatternViewer minus Instr.Dynam accuracy in the ScoreViewer against years of formal musical training. A corresponding regression analysis did not identify any significant linear or exponential trends.

appreciation might emerge simply by continued music listening. Our contribution was to determine how appreciation or appraisal of a piece of music varied with question topic (instrumentation, dynamics, repetition, tonality) and visualisation: we used the ScoreViewer, which automatically aligns a playing audio file to the relevant position in the staff notation; and the PatternViewer, which is automatically aligned to the audio as well, but includes representations of repetitive and tonal structure also. As such, the first visualisation (ScoreViewer) acted as a kind of control condition for normal music listening with score, while the second (PatternViewer) presents listeners with parsimonious representations that might help them to appraise the inner workings of a piece.

6.1 Outcome of the experiment

The results of our experiment demonstrated that listeners' appraisal of repetitive and tonal structure was improved when using the PatternViewer, compared to using the ScoreViewer. While this overall observation holds, it was nuanced by a further analysis, which suggested participants with five years or less of formal musical training showed the greatest appraisal improvements for repetitive and tonal structure questions, when using the PatternViewer compared to ScoreViewer. On the other hand, more musically-expert users showed moderate or no improvement. When modeling this accuracy-difference-by-visualisation for repetition and tonal questions (Fig. 6A), we found a proportion of variance $R^2 = .43$ was explained by just one variable: years of musical training. Our main finding invites the conclusion that exposure to carefully chosen representations derived from behavioural experiments and accompanying models can help listeners to gain an appreciation of the inner workings of pieces of music, over a relatively short period of time.

That more expert listeners displayed moderate or no improvement on repetition/tonality questions when using the PatternViewer versus ScoreViewer suggests that their habitual methods for answering such questions may have been disrupted in the PatternViewer. Most music students are taught to estimate the key of an excerpt of staff notation by reading the pitch classes of notes and determining to which key they fit best. The temporal order and rhythmic structure of the notes plays a role as well. While it is possible for trained musicians to read the pitch classes of notes in the PatternViewer, it is not as straightforward a task as in the ScoreViewer—the latter using the five staves to which musicians are highly accustomed. Being forced to put their habitual technique for key estimation to one side and to adopt the method provided by PatternViewer instead could have resulted in a processing cost, leading to the moderate or no improvements observed. A similar explanation might be offered for repetitive structure. Processing costs of this kind, where participants have to unlearn a previously successful strategy, have been proposed as an explanation for similar patterns of results in other task domains [2].

Questions from two further categories were presented (instrumentation and dynamics), which were intended to complement repetitive and tonal structure questions. That is, instrumentation and dynamics questions might well have been easier to answer in the ScoreViewer than in the PatternViewer, and so the question categories were balanced with respect to visualisation mode, allowing for the possibility that the PatternViewer is not unilaterally preferable to staff notation for appraising music. Results indicated that there was no significant difference in terms of listeners' appraisal of instrumentation and dynamics, when using the ScoreViewer or PatternViewer visualisation. Subtracting a participant's accuracy for instrumentation and dynamics questions in the ScoreViewer from their accuracy in the PatternViewer, this difference did not appear to be a function of years of formal musical training. Based on this finding, we conclude that in terms of improving music appraisal, the PatternViewer's benefit is limited to matters of repetition and tonality. Staff notation is just as effective for conveying matters of instrumentation and dynamics.

The results regarding instrumentation and dynamics should be set in the context of (1) staff notation's longevity, and (2) its explicit specification of instrumental and dynamic information. (1) With some of the earliest surviving polyphonic music dating from 1150-1250, and maintaining broadly the same format in today's notation, it would be unlikely for a new representation of music to be unilaterally preferable to staff notation for conveying musical information. (2) Instrument names are written on the score, and so if noteheads appear in the flute part, then it is clear that a flute plays during a certain time window. Similarly, opening "<" and closing ">" hairpins that signify crescendos and diminuendos respectively are written into the score at appropriate points. It might be suggested that instrumentation and dynamics questions are straightforward to answer in the ScoreViewer, because the relevant information is made explicit. It should be said, however, that these questions can be made more complex (and sometimes were in our experiment), by asking about combinations or families of instruments, or subtler changes in loudness or tempo.

Unlike instrumentation and dynamics, only rarely are aspects of repetition and tonality marked explicitly in the score (and by extension, ScoreViewer): a key signature is given at the beginning of a piece, but still it may be in a major key or the relative minor, and usually the key signature is not modified to reflect local key changes; since the Classical period, repeated sections have been marked in the score rather than all the notes being written out a second time, but still, motives, themes, and other local repeated figures are not annotated. The PatternViewer, then, makes aspects of repetitive and tonal structure more explicit, providing the cognitive representations via which it is possible for a less expert listener to relate technical terms such as *motif* or *key change* to observable and interactive elements/events on the screen, such as nodes in a pendular graph or distinct colour changes in piano roll respectively.

Guiding the listener in this way could invite the criticism that any improvement in appraisal (such as that observed in our experiment) may be transient rather than permanent. For instance, does reading off estimates of local keys from a lookup index or diagram (such as the circle of fifths in the PatternViewer) help a less expert user to *understand* tonality any more than using a calculator to work out 6×7 fosters an understanding of multiplication? Our answer is that it depends on the topography of the index and the recurrence of its use. After interacting with the PatternViewer for a longer period of time (though substanially less time than would be required in the absence of its repetitive and tonal structure representations), a user may notice that pieces tend to begin and end in the same key, that tonicsisations and modulations tend to occur between proximal keys on the circle of fifths, that the second theme in the recapitulation of a sonata movement switches key compared to the exposition, from the dominant (or relative major) to the tonic, etc. The more parsimonious the topography of the index or diagram, the more readily these tendencies are likely to become apparent. Even for the more expert listener, it can be helpful to highlight such tendencies, since [9] demonstrated that music students are surprsingly insensitive to global key relations when listening with the score.

6.2 Experimental psychology as a catalyst for understanding intellectual phenomena

Recent years have seen a huge increase in educational games/apps. Some of the names and claims attributed to these applications (e.g., "brain fitness workouts that can help your mind process information more quickly" [3]) imply that proficiency gains will go hand in hand with use. These claims, largely untested, have been called into question by the scientific community [22]. At a more moderate pace than the rapid growth of the applications market, educational psychologists and technologists are developing and testing electronic games that appear to enhance learning in domains such as mathematics, the natural and social sciences, and engineering [5]. The domain of music appraisal is ripe for investigation, therefore. It has been shown that meaningful interactions with structured information such as music can have a positive impact on intellectual ability [16]. In this paper, we have incorporated two parsimonious models of music cognition [11,18] into a visualisation application and tested its ability to increase proficiency for certain aspects of music listening and appraisal. The models-come-visualisations, based on behavioural data from previous repetition and tonality experiments with more expert listeners, appear to help less expert listeners, inviting the interpretation that these models act as a catalyst or go-between: a less expert listener's intuitions about musical structure are promoted to the level/clarity of understanding required to link musical events to the correct technical terms synonymous with proficient appraisal. According to this interpretation, the role of cognitive psychology is to act as a catalyst, as a less expert individual gains deeper understanding and appreciation of intellectual phenomena [14].

6.3 Limitations and future work

We have been careful to set out the limitations and scope of the PatternViewer's ability to enhance music appraisal skills. In part, this is to avoid the kind of controversy that was caused by non-replicability and partial reporting of the so-called *Mozart effect* [4], where incomplete information was provided on what questions appeared to be answered more accurately after ten minutes of listening to Mozart [23]. Here we are not claiming an increased intellectual ability beyond music—only increased appraisal abilities for *certain aspects* of musical structure. Rather than showing a general proficiency effect after mere exposure to music, of more immediate relevance to the present work would be to devise and test analogous visualisation applications within alternative task domains.

With regards the PatternViewer, future work will consist of (1) investigating whether there are other aspects of musical structure (e.g., [25,15]) for which beneficial representations can be developed, and (2) expanding the number of pieces for which we have the necessary data encodings. With regards (1), other aspects of musical structure could include harmonic progressions, textures, musical tension, etc. This work in turn may motivate new psychological experiments of the type that [10] and [18] performed in the 1980s, because, for instance, models of sensitivity to different chord functions in a progression, and sensitivity to different textural categories, are still to be fully developed [8].

We are cautious about including too many representations in the visualisation, however. First, the visualisation contains five sources of information already (media control bar, global and local piano roll, coloured circle of fifths, and pendular graph). Introducing more representations could lead to information overload. Second, there is a balance to be struck between the number of pieces for which the PatternViewer can be used and the depth/variety of representations on offer for any one piece. We continue to advance methods for automating aspects of the PatternViewer display, such as alignment of the audio and piano roll [13], or specification of the pendular graph [6], but there is still a bottleneck in terms of the relatively small number of pieces for which we have high-quality symbolic encodings. The developers of the Orchestra App face an even bigger challenge in this respect, since in the Orchestra App there are even more display modes for any given piece.

In future, we envisage this work could have a wider impact on people's listening experiences. New immersive user experiences in which the user can interact with music structures in a visualisation space could further promote the understanding of the inner workings of a piece. At the more compact end of the scale, another possibility could be the development of mobile apps for use in school music lessons, or for use by music enthusiasts in preparing to attend the concert hall.

Acknowledgments

Gerhard Widmer's work is supported by the European Research Council (ERC) under the EU's Horizon 2020 research and innovation programme, grant agreement No. 101019375 ("Whither Music?").

References

1. AQA: Listening to and appraising music: Past paper for GCSE music unit 1. AQA, Manchester, UK (2012), retrieved 13 January, 2014 from http://www.aqa.org.uk/

2. Brady, T.F., Konkle, T., Alzarez, G.A.: Compression in visual working memory: Using statistical regularities to form more efficient memory representations. Journal of Experimental Psychology: General **138**(4), 487–502 (2009)
3. Brainmetrix: Brain training, brain exercises, brain fitness games, brain metrix (2014), retrieved 1 August, 2014 from http://www.brainmetrix.com/
4. Chabris, C.F.: Prelude or requiem for the Mozart effect? Nature **400**, 826–827 (1999)
5. Chen, N.S., Hwang, G.J.: Transforming the classrooms: Innovative digital game-based learning designs and applications. Educational Technology Research and Development **62**, 125–128 (2014)
6. Collins, T., Arzt, A., Flossmann, S., Widmer, G.: SIARCT-CFP: Improving precision and the discovery of inexact musical patterns in point-set representations. In: Proceedings of the International Society for Music Information Retrieval Conference. pp. 549–554. Curitiba, Brazil (2013)
7. Collins, T., Meredith, D.: Maximal translational equivalence classes of musical patterns in point-set representations. In: Proceedings of Mathematics and Computation in Music. pp. 88–99. Montreal, Canada (2013)
8. Collins, T., Tillmann, B., Barrett, F.S., Delbé, C.: A combined model of sensory and cognitive representations underlying tonal expectation: From audio signals to behavior. Psychological Review **121**(1), 33–65 (2014)
9. Cook, N.: The perception of large-scale tonal closure. Music Perception **5**(2), 197–205 (1987)
10. Deutsch, D.: The processing of structured and unstructured tonal sequences. Perception and Psychophysics **28**, 381–389 (1980)
11. Deutsch, D., Feroe, J.: The internal representation of pitch sequences in tonal music. Psychological Review **88**(6), 503–522 (1981)
12. Edexcel: Listening and appraising: Past paper for GCSE music unit 3. Pearson Education Ltd, London, UK (2012), retrieved 13 January, 2014 from http://www.edexcel.com/
13. Grachten, M., Gasser, M., Arzt, A., Widmer, G.: Automatic alignment of music performances with structural differences. In: Proceedings of the International Society for Music Information Retrieval Conference. pp. 607–612. Curitiba, Brazil (2013)
14. Hmelo-Silver, C.E., Pfeffer, M.G.: Comparing expert and novice understanding of a complex system from the perspective of structures, behaviors, and functions. Cognitive Science **28**, 127–138 (2004)
15. Katsiavalos, A., Collins, T., Battey, B.: An initial computational model for musical schemata theory. In: Proceedings of the International Society for Music Information Retrieval Conference. pp. 166–172. Delft, The Netherlands (2019)
16. Kraus, N., Martin, M.: Biological impact of musical training in at-risk children. In: Proceedings of the American Psychological Association's 122nd Annual Convention. APA, Washington, DC (2014), retrieved 8 August, 2014 from http://www.apa.org/news/press/releases/2014/08/musical-training.aspx
17. Krumhansl, C.L.: Cognitive foundations of musical pitch. Oxford University Press (1990)
18. Krumhansl, C.L., Kessler, E.J.: Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. Psychological Review **89**(4), 334–368 (1982)
19. Malinowski, S.A.: Time-line of the Music Animation Machine (and related experiments) (1985), retrieved 23 May, 2014 from http://www.musanim.com/mam/mamhist.htm

20. Newton, I.: Opticks. Dover (1952), (Original work published 1704)
21. OCR: Listening: Past paper for GCSE music unit 4. OCR, Coventry, UK (2012), retrieved 13 January, 2014 from http://www.ocr.org.uk/
22. Owen, A.M., Hampshire, A., Grahn, J.A., Stenton, R., Dajani, S., Burns, A.S., Howard, R.J., Ballard, C.G.: Putting brain training to the test. Nature **465**(7299), 775–778 (2010)
23. Rauscher, F.H., Shaw, G.L., Ky, K.N.: Music and spatial task performance. Nature **365**(64478), 611 (1999)
24. Saffran, J.R., Johnson, E.K., Aslin, R.N., Newport, E.L.: Statistical learning of tone sequences by human infants and adults. Cognition **1**(70), 27–52 (1999)
25. Sutcliffe, R., Hovy, E., Collins, T., Wan, S., Crawford, T., Root, D.L.: Searching for musical features using natural language queries: the C@merata evaluations at MediaEval. Language Resources and Evaluation **53**(1), 87–140 (2019)
26. Tillmann, B., Bharucha, J.J., Bigand, E.: Implicit learning of tonality: A self-organizing approach. Psychological Review **4**(107), 885–913 (2000)
27. WJEC: Appraising: Past paper for GCSE music unit 3. WJEC CBAC Ltd, Cardiff, UK (2012), retrieved 13 January, 2014 from http://www.wjec.co.uk/
28. Zhang, W., Luck, S.J.: Discrete fixed-resolution representations in visual working memory. Letters, Nature **453**(7192), 233–235 (2008)
29. Zhang, W., Luck, S.J.: The number and quality of representations in working memory. Psychological Science **22**(11), 1434–1441 (2011)