Exploring the perceptual effects of performers’ interpretations

Jennifer MacRitchie and Hubert Eiholzer
University of Applied Sciences and Arts of Southern Switzerland

Background in performance analysis and perception. Empirical studies of performance attributes such as tempo variations, dynamics (Palmer, 1989; Repp, 1997, 1999) and even visual gestures (Vines et al., 2006) posit that structural emphases can be measured from these aural and visual streams. Recording continuous perceptual responses from audience judges on tasks such as shaping perceived phrasing and tension (Vines et al., 2006) provides insight into the communicative function of performance and allows the exploration of relationships between performance parameters and their “communicative” outcomes.

Background in philosophy of musical performance. To perform a piece of music is to present and portray the patterns of that piece of music (Walton, 1988). The nature and extent to which performers may actually articulate patterns is unclear. Empirical studies in this field may help clarify questions such as the artistic value of the performer’s contribution in performing a musical work and the comparability of different performances (what Levinson calls “performative results”) of the same piece.

Aims. This study examines whether a performer’s unique interpretation can be conveyed to the point where it changes how the listener “hears” a piece, extending the work of Spiro (2007).

Main contribution: Musically-trained audience judges were presented with three performances each of Chopin’s B Flat Minor Sonata, final movement, and of Chopin’s Prelude in E minor and asked to indicate phrasing continuously using a slider. The effects across participants for the one performer and also across performances are examined, exploring how these relate to measured aural features such as tempo and dynamics. Results across all participants show a range of responses indicating that there is not an ‘agreed’ perceived interpretation for each performance. Collating the phrase boundary responses for each performance, the strength of responses at particular points in the score vary slightly across performances of the same piece, suggesting that performers can control the extent to which a structural boundary is emphasised.

Conclusions: These preliminary results show that despite the fact that the act of listening to music can be personal and dependent on a number of a priori factors, different interpretations can have an effect on how we hear phrasing according to the emphasis or de-emphasis of certain structural features.

Implications. These results imply that performers can have an influence on the audience perception of phrasing even in well-known pieces of repertoire. This may help in considering how audiences respond in comparing performances and could potentially redefine the role and artistic value of performers in music making.

Keywords: interpretation; perception; continuous response; performance analysis

*Correspondence: Jennifer MacRitchie, Conservatorio della Svizzera Italiana, Via Soldino 9, 6900 Lugano, Switzerland; tel: +41 91 960 2363, e-mail: jennifer.macritchie@conservatorio.ch
*Received: 30th March 2012; Revised: 13 May 2013; Accepted: 23 July 2013
*Available online: 10 March 2014
*doi: 10.4407/jims.2014.02.004
Introduction

With a large part of the classical music industry involved in producing recordings of canonical repertoire performed by different artists, we could feasibly ask how different are these performances, and can an artist really provide a “fresh” interpretation of a well-known piece to a listening audience? Laying aside certain performance traditions of different time-periods and focusing on current recordings, what makes these different to each other in terms of how the audience understands the piece? It is acknowledged that performers can create different interpretations from the same musical material. So how are their unique performance choices conveyed within the same music and to what extent do these have an effect on how the audience “hears” the music? The present study examines the potential effects of different performances of the same piece of music on the perception of phrasing.

Background

In defining an interpretation of a piece of music, there are some difficulties as to whether this can be solely defined as differences from the literal interpretation of the score. For example, it may often be seen that “doing nothing at all” to parts of the music is viewed as one particular interpretation. For the purposes of the empirical part of this study, we look at “interpretation” by examining the differences between the selected performances in tempo and dynamics and their relation to the musical structure.

Background in Philosophy of Musical Performance

Most musicians seem to agree that a musical work is or implies a sound structure with different patterns related to each other in various ways, a performance of a musical work counts as a performance of that work only if it presents the work’s sound structure and to understand a musical work is to hear it in terms of the music’s structure. Philosophers of music, however, have subjected these claims to discussion, refinement or criticism (see Kania, 2010, for an overview).

To perform a piece of music is to present the sound patterns - melodic, harmonic, rhythmic, their combinations and relationships - that constitute the piece’s structure. But performers not only present sound patterns, they also present them in certain ways (Walton, 1988). A performer may emphasize certain similarities between patterns and obscure others, may present one pattern as a restatement or as a variation or a development of another, may present a pattern as a new idea or, alternatively, as the continuation of an old one and so on. Furthermore, performers may present patterns not only in different structural or functional roles; they may also present them as having different expressive features, for instance as being graceful or sad or as changing from hope to despair.
To say that the performance of a musical work is the presentation of the work’s sound structure is not sufficient (Levinson, 1996). A typical concert performance differs from a prima vista performance or a student’s imitation of his model interpreter’s rendition of that work in two respects. In a concert performance the performer intends to present the structure of the work, but he or she intends to do so by playing the work in a certain way. A performance is a “considered way of playing a piece of music, involving highly specific determinations of all the defining features of the piece as given by the score and its associated conventions of reading” (Levinson, 1996, p63). Although a performance is typically driven by a conception of the musical work, by what its main elements are and how they hang together at various levels, it is neither equivalent to such a conception nor does it necessarily express such a conception held by the performer. If this is correct, two questions remain: What is it that performers can do to present the structure in such a way that the audience can perceive it accordingly? And wherein lies the artistic value, if any, of such presentation?

Empirical studies in this field may help clarify questions such as these through comparing different performances (performative results) of the same piece and how these in turn can affect audience perception.

Background in Performance Analysis and Perception Studies

That a composition can inspire so many different interpretations is one of the most interesting aspects of musical performance. An interpretation is a very individual entity, personal to the performer. It may reflect, or be perceived as reflecting aspects of the performer’s identity and most certainly their thoughts on the composition at hand. Palmer (1989) found that the most striking changes of tempo (e.g., from fast to slow and vice versa) in six performances of a Brahms Intermezzo occurred at locations identified before performance by each pianist as phrase boundaries. The amount of overlap measured between the amplitude envelopes of each two consecutive notes in the piece was also found to be largest for notes that performers had circled on the score as belonging to the melody line. This suggests that the act of mapping the performer’s thoughts to performance follows a strict set of expressive rules, but it is the way in which each performer uses these rules according to their own interpretations of the score which can often produce different sounding performances. Further empirical studies of musical performance have identified some of the rule-sets that performers can use for articulation, expressive timing and shaping phrases (for one example of such a rule-set see Friberg, Bresin, & Sundberg, 2006). However, despite the existence of rule-sets, there is not a simple path between an interpretation and a resulting performance. Repp (1997, 1999), for example, has measured timing and dynamics in hundreds of audio recordings, looking at how performances of Schumann’s Träumerei and of the beginning of Chopin’s Etude in E major differ from each other. Results suggest that although structural interpretations could be considered somewhat limited in the set of distinct possibilities they offered, the expressive shaping strategies used by performers was on more of a continuous scale of possibility. The most distinct differences found between performances were in fact those on a microstructural level. As a musical performance can be treated as a
“direct stimulus to the listener” (Gabrielsson, 1988), providing the audience with certain impressions concerning the composed music, the question arises: to what extent can a performer influence the audience’s impression of a piece of music?

Examining what information audiences gain from a performance, Krumhansl (1996) presented various synthesised performances of an expert performance of Mozart’s piano sonata in E flat major to audience judges, and asked them to perform several tasks. The first performance was not modified, the second kept dynamics constant, and the third maintained a constant tempo. After hearing the entire piece once, tasks undertaken on successive hearings by the judges included identifying phrasing segmentation at varying levels, tension and the beginnings of new musical ideas. The data suggested consensus on judgements of larger sectional boundaries and correspondences between local tempo and local phrasing judgements. Looking at comparisons of real musical performances and MIDI-synthesised ones, Spiro’s (2007) work continued this idea of examining different interpretations by looking at the effects of structural and performance features on audience phrase perception. Two different performances of the cor anglais solo from Wagner’s Tristan und Isolde were analysed for tempo and sound intensity. Audiences were presented with the original audio versions and a processed “deadpan” MIDI version. In two consecutive sessions, participants were asked in the first to press and hold down a key to indicate the start of a phrase, releasing the key with the end of the phrase, and in the second session, to press a key to indicate the beginning of their expectation of each phrase ending. The results showed that participants were more likely to agree on where phrases ended in the real performances than the MIDI versions, but disagreements as to exactly where the phrases ended were small. Differences were also found in the lengths of time between the point at which participants expected phrases to end and when the phrases actually ended. These differences varied by performance suggesting that participants’ expectations were influenced by the specific performance to which they were listening.

Using the Continuous Response Methodology to Measure Perception of Phrasing

For the purposes of this current study it was decided to assess audience judgment by using continuous responses. This has been verified as an accurate method of collecting audience responses (see Sloboda, 2001; Schubert, 2005) having also been used particularly to investigate the way phrasing is perceived (Vines, Krumhansl, Wanderley, & Levitin, 2006; MacRitchie, Buck, & Bailey, 2009). Although key presses may provide more precise locations at which audience members perceive phrase boundaries, the information as to where they expect phrase boundaries to occur changes over time and so would only be captured by a continuous response. MacRitchie et al.’s (2009) study concerned piano performances of Chopin Preludes where the performers’ views of structure largely agreed with one another, and this was detected by the audience judges. However, Vines et al. (2006) used performances differing slightly in interpretation and the differences were also reflected in audience responses. For example one performer of Stravinsky’s second Piece for Solo Clarinet interpreted the second phrase as a continuation of the first,
while the other interpreted it as a separate phrase. The first performer used increasing dynamics and constant tempo across the phrase boundary to reflect his choice, compensating for the short breath taken between phrases. The second performer took a larger breath between the two phrases and started the second phrase at a slightly different tempo than before. Audience responses reflected the difference between the two interpretations irrespective of whether presentation mode was audio only or audiovisual. However, in both modes of presentation, judges may have been influenced by perceptual cues provided by the sound and motion of the performers’ breathing. The present article investigated whether audience reactions are still influenced by different interpretations when these breathing cues are not as pertinent to the sound production by using the piano as the main instrument.

This paper aimed to extend Spiro’s (2007) research methods exploring the differences between performers’ interpretations combined with the continuous response method (Schubert, 2004, used by Vines et al., 2006) in order to find out how audiences listen to performances and undertake a phrase segmentation task while doing so. These results were then compared across different performances to evaluate the similarities and differences between them to determine the extent to which the performer’s emphasis on different patterns in the notated score influences the audience’s chunking of groups of musical notes into phrases.

Method

Materials. Three performances of Chopin’s Prelude in E minor Op.28 No.4 and the final movement of Chopin’s B Flat minor piano sonata Op.35 were given by three masters students at the Royal Conservatoire of Scotland and University of Glasgow and three professional pianists respectively. As the performers were different for each piece, we denote them as P1, P2 and P3 for the performances of the Prelude, and F1, F2 and F3 for performances of the sonata’s final movement. These performances were recorded live by the first author in order to emulate real performance environments as part of previous projects. The Prelude is well known by performers and audiences, and although its structure is clear, it is still capable of being interpreted in different ways (for example, one recording that has been analysed for its individual sense of expressive timing is the 1975 recording by Martha Argerich, see Senn [2009]). The final movement of the sonata (referred to as the Finale), despite being by the same composer, has a potentially ambiguous structure. Despite this, it is a piece that has been recorded many times. This piece is a little less well known, and completely unfamiliar to some listeners. These two pieces provided a contrast between familiar and unfamiliar pieces, with comparisons also available across composed style and different tempi.

Musical Analysis. Scores of both pieces are included in the appendices. The Prelude was chosen as a stimulus piece as, despite its relative simplicity, it offers a multitude of interpretations with some pianists choosing to emphasise the left-hand chordal progression, the bass-line as it descends or even to portray the melody line as
something of a lament, separate from the chordal accompaniment (Barolsky, 2008). Clarke’s (1995) analysis of this piece takes note of the antecedent-consequent structure of the melody that effectively splits the piece into two halves around Bar 13. Clarke’s voice-leading analysis follows the descent of the melody beginning on the B at Bar 1, descending to A at Bar 5 with a transfer in the register occurring at Bar 9. The midpoint at Bar 13 marks the repetition of the opening material, continues through the dynamic climax at Bar 17 featuring the octave Bs and is followed by an interrupted cadence in Bar 21, finishing with the final perfect cadence in Bars 24-25.

The Finale on the other hand is something that confounds traditional analysis even as to its general form (Barolsky, 2007). Many analyses point out the repetition of the opening bars at Bar 39 that suggests it is something of a binary sonata form, as implied by Charles Rosen (1995). The unrelenting quaver octaves present changes of harmony quickly to the listener, however the following landmarks are noted: following a melodic contour that reaches its lowest point at Bar 23 a new D flat theme is introduced. This theme is modulated and experiences a melodic climax at Bar 51 with the return to the tonic at Bar 57. The composition is marked sotto voce and no other dynamics are notated except where a single crescendo and decrescendo is marked at Bars 14-15.

Performance Analysis. To explore the potential aural cues audiences may use to make these phrasing judgements, we examined the measured values of tempo and dynamics for each performance. Melody line was also considered as a potential cue for phrasing perception, but this is regarded as a constant across interpretations of the same piece. Global tempo was estimated by taking the median tempo, and the amount of variability within the sample calculated by the standard deviation was used to represent the amount of rubato used by the performer. Global dynamics for the melody was also estimated by the median key velocity, and variability again represented by the standard deviation. The values for the performances of the Prelude can be seen in Table 1. Here we see that the three performers for the Prelude used similar general tempi and dynamics, with a large degree of rubato and dynamic shaping, as may be expected in this piece.

Table 1. Global tempo and dynamics measurements calculated as the median and standard deviation of the tempo calculations and of the right hand key velocity values respectively for all three performances of the Prelude.

<table>
<thead>
<tr>
<th>Performer</th>
<th>Global Tempo</th>
<th>Tempo Variation</th>
<th>Global Dynamics</th>
<th>Dynamics Variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>64.0023</td>
<td>20.6633</td>
<td>68</td>
<td>14.7756</td>
</tr>
<tr>
<td>P2</td>
<td>69.7404</td>
<td>13.562</td>
<td>61</td>
<td>15.551</td>
</tr>
<tr>
<td>P3</td>
<td>74.3028</td>
<td>16.1663</td>
<td>69.5</td>
<td>25.3419</td>
</tr>
</tbody>
</table>

Figure 1 shows tempo and key velocity information for performances of the Prelude as it develops over time. The overall shape of tempo variation for all three pianists was very similar, all with a small ritardando at the halfway point at Bar 13, a large increase in tempo at the harmonic and melodic arrival in Bar 17 and a small decrease in tempo at Bar 21 emphasising the return to the tonic. However, at the level of...
microstructural differences, P3 was the most consistent in shaping the timing of each bar in the same way, whereas P1 began in a similar manner but later extended tempo shapings over bar lines, in effect joining bars together. P2 made the fewest tempo variations across the first 8 bars, with most of the tempo variation occurring in the second half of the piece. In terms of dynamics, all three performers made a crescendo towards the climax at Bar 17. This was the major structural landmark for P1 and P2 but P3 made another large crescendo in the run-up to Bar 9. Although these performances appear similar, it is these underlying microstructural differences that we hypothesise will influence the audience perception of phrasing.

Figure 1. Tempo (left) and key velocity data (right) for each performance of the Chopin Prelude. Tempo is calculated from the inter-onset intervals of matched MIDI notes in the left hand, as the constant quaver rhythm accompaniment provides a more accurate measure of tempo throughout the piece. Key velocities for the right hand melody are used to provide an indication of the dynamic changes in the melody. Vertical dotted lines at Bars 5, 9, 13, 17 and 21 mark features discussed in the musical analyses.

Table 2. Global tempo and dynamics measurements calculated as the median and standard deviation of the tempo calculations and of the right hand key velocity values respectively for all three performances of the Finale.

<table>
<thead>
<tr>
<th>Performer</th>
<th>Global Tempo</th>
<th>Tempo Variation</th>
<th>Global Dynamics</th>
<th>Dynamics Variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>164.0995</td>
<td>9.1635</td>
<td>99</td>
<td>20.7001</td>
</tr>
<tr>
<td>F2</td>
<td>203.5002</td>
<td>18.2247</td>
<td>91</td>
<td>18.9312</td>
</tr>
<tr>
<td>F3</td>
<td>205.3938</td>
<td>28.8099</td>
<td>80</td>
<td>21.3098</td>
</tr>
</tbody>
</table>

For performances of the Finale, Table 2 shows the calculations for the performers’ tempo and dynamics, showing the much slower overall tempo taken by Performer F1 as opposed to F2 and F3. However, the largest variations in tempo were produced by F3.

Figure 2 shows the tempo and dynamics information displayed over time for all performances of the Finale. F3’s wider variation in tempo than F1 or F2 can be seen
clearly, with an additional ritardando around Bar 17 and emphases at Bars 39, 51 and 57 with smaller ritardandi occurring either on or slightly before each junction similar amount of variation in dynamics can be seen for all three performers. The dynamic shaping for each performance shows particular differences between bars 23 and 39, the section where the new D flat theme is introduced and makes its way back to the repetition of the opening bars at Bar 39.

With reference to the phrase arch rule proposed by Friberg et al. (2006) suggesting that performers emphasise phrases using arch-like shapes for both tempo and dynamics, we hypothesised that the particular phrase junctions indicated in the musical analyses, likely to emphasised by the performer with both a ritardando and a decrescendo, would be the phrase boundaries most commonly perceived by judges. On the basis of our observation of the differences between all the performers’ audio-recordings as measured above, we hypothesised that the instances of larger overall variations in tempo combined with larger overall variations in dynamics (as seen in performances F3 and P3) would influence the audience to segment these performances into more phrases.

Participants. Fifteen musically-trained judges were recruited by email from the Conservatorio della Svizzera Italiana. Each participant completed a short questionnaire designed to ascertain their level of musical training, background and listening and performing preferences. Participants comprised 10 females and 5 males, ages ranging from 24 to 58 years. All were formally trained in Western traditional classical music with a mean of 15.2 years of musical training. Three participants stated piano as their main instrument and four further participants noting they had
received some training on the piano even though it was not their primary instrument'. Fourteen participants were familiar with the Prelude but only four were familiar with the Finale. Of these four, three participants were pianists, and one of these had studied both pieces.

**Apparatus.** Participants listened to each performance through Sennheiser HD-500 headphones and were simultaneously asked to indicate phrasing continuously using a vertical slider. The slider was presented on a computer screen for the duration of each performance and its position was sampled every 117ms via the stimulus delivery and experimental control software Presentation 16.0.

**Procedure.** Participants listened to three performances of each of the two Chopin pieces described above in a pseudo-randomised order to control for order effects. A practice trial involving two separate performances of another Chopin Prelude (Op. 28 No.7) were used to familiarise participants with the process. The directions presented on screen are shown in Figure 3. These assumed that a phrase was defined as a chunk of music with a beginning and an end, with a climax occurring at some point in between them. Participants were directed to shape the movement of the phrasing in each performance as “the performer is telling you”. Extra guidance was given to counteract the potential ambiguity of the word “phrasing” by describing the point in the music where a singer would take a breath. Continuous judgements were used to gather information as to whether phrase endings could be predicted for a long time preceding their occurrence, or if the listener experienced them as occurring suddenly. Obviously this would have been affected by their past musical experiences and how well they knew the piece.

**Results**

We examined participants’ judgements of phrasing across performances of the same piece by three different performers, first for the three performances of the Prelude, and then for the three performances of the Finale. Judgements were assessed in relation to the measured tempo and dynamics of the audio-recordings.

---

You will hear several performances of different pieces. For each of these, you are to describe how the performer is phrasing the piece by moving the mouse up and down on the slider on the screen in front of you.

Each time, the mouse will start at the bottom of the slider.

As the performer BEGINS a new phrase, move the slider UP.

As the performer ENDS a phrase, move the slider DOWN.

You will have two practice trials to try this out. To continue and start these practice trials, press ENTER.

**Figure 3.** Instructions given to participants regarding the mouse slider.
Results of Continuous Responses. Results from the 15 participants for each performance can be seen in real-time along with the musical stimuli in the videos attached as appendices. These videos provide a novel representation of the results of the phrasing-segmentation task and allow readers to follow from moment to moment how each participant followed the phrase structure of each performance. Readers can also view the differences between the group of participants in terms of the level of phrasing they are attending to, whether this was small chunks or larger sections of the piece. The videos also allow observation of how each participant’s expectation of these phrase boundaries developed over time. Two performances of both the Prelude (P1 and P3) and the Finale (F1 and F3) are shown. The responses of each participant changed slightly over the course of listening to the three performances, while the patterns of response (and attending phrase level) remained similar. The responses of the participants as a group to the same performance, however, were different.

An example of the participants’ continuous responses produced for P1’s performance of Chopin’s Prelude in E minor is seen in Figure 4. Comparing participants’ responses, we can see that some listeners attend to the large sections of the material whereas others attend to shorter segments.

**Figure 4.** 15 Participant responses to P1’s performance of Chopin’s Prelude in E minor Op.28 No.4. Y axis measures the position of the mouse slider (range -300 to 300).

Between-Participant Comparison. To calculate the rate of agreement between participants undertaking the phrase-segmentation task, comparisons were made between results from all three performances of the Prelude and all three performances of the Finale. To control for serial ordering effects (see Schubert, 2001), correlations were calculated using the first-order difference measure of each of the participants’ responses for each performance of each piece. These were then subjected to Spearman’s correlation tests to determine how well the participants agreed with each other about each performance. From 630 correlations calculated for each set of performances, for performances of the Prelude, there were 392 positive significant correlations and 52 negative significant correlations ranging from -0.29 to 0.55. For performances of the Finale, there were 356 positive significant correlations and 22
negative significant correlations ranging from -0.13 to 0.45. This would suggest that there was slightly higher participant agreement for the phrase-segmentation task for performances of the Prelude than performances of the Finale. As the rate of agreement was fairly low for both pieces, it was decided not to calculate an average of the participants’ responses as this would not be reflective of the overall group response.

**Within-Participant Comparison.** In order to compare the responses of the same participant for different performances, the data was time-normalised using the timing of the start of each notated bar in the audio as benchmarks. This process mapped the responses from the time domain into a warped time domain that spanned from 0 to 1000 data points. From this we could say that each response at a particular time point would relate to the same position in the score, despite differences in timing, i.e. rubato in each performance. Figure 5 shows these time-normalised responses of participants 1 through 6 for the set of Prelude performances as an example. Correlations were calculated using the same procedure as for the between-participant comparison, i.e. using the first-order differences to control for serial correlation effects. 3x15 correlations for participants ranged from -0.013 to 0.577 for the Prelude and from -0.099 to 0.717 for the Finale. The participants showing the highest correlation between their responses for the Prelude included the pianist who had previously studied both pieces (Participant 4). The responses of the participants shown in Figure 5 shows that there was some agreement (e.g. between participants 3 and 4) on the structure of all three performances, i.e. the number of sections, and the location of phrase boundaries, but other participants’ responses were different for each performance.

![Figure 5](image)

*Figure 5.* Comparison of responses for each performance of the Prelude for Participants 1-6. The red line denotes P1’s performance, the blue line P2 and the green line P3. This data is time-normalised according to the onset time of the first note of each bar.
Figure 6 shows the time-normalised responses of participants 1 through 6 for the set of performances of the Finale. The responses of all the participants were different for each performance in terms of the number of sections indicated, and the location of perceived phrase boundaries. Some participants consistently indicated that they perceived the entire piece as one long extended phrase, so their responses were strongly correlated ($r = 0.7$). The next highest correlation was $r = 0.4515$. These findings show variable agreement between judgements for different performances of the same piece.

Phrase Boundary Analysis. When examining the correspondences between perceived phrasing judgements, it was necessary to collate information on the position of the “troughs” in each response that corresponded to a perceived phrase boundary.

Phrase boundaries were identified in terms of the bar of notated music in which they appeared, although of course delayed reaction times may have resulted in some differences between the score and the participant’s perception as recorded using the slider. For this reason and accounting for differences between performers’ overall tempi for the piece, responses were collated for each bar of the Prelude, and every two bars of the Finale. Figure 7 shows the percentage of participants who responded to each bar for the Prelude or every 2 bars for the Finale, indicating a phrase boundary at this point in the performance. High peaks in the graph denote locations where a large majority of the participants agreed that there was a phrase boundary. The highest percentage of agreement for performances of the Prelude was approximately 87% whereas the highest agreement for performances of the Finale was approximately
33%. There was therefore a higher rate of agreement on phrase boundary location in the Prelude.

Figure 7. Percentage of responses indicating a phrase boundary with a trough in the continuous response measure for Performer P1’s, P2’s and P3’s performances of Chopin’s E minor Prelude Op.28 No.4 (Column a) and Performer F1’s, F2’s and F3’s performances of Chopin’s B Flat minor sonata Op.35, Final Movement (Column b).
The three graphs in Figure 7(a) representing responses for the Prelude show differences between the numbers of phrase boundaries indicated for each interpretation. The phrase boundaries at Bar 13 and at the imperfect cadence in Bar 21 were consistently indicated in responses for all performances, particularly the latter in the second and third performances. Thus participants were experiencing the same piece of music differently when it was performed by different pianists. This is illustrated by the strength of each boundary as calculated from the percentage of responses. The graphs in Figure 7(b) representing responses for the Finale show a relatively low rate of agreement on perceived boundaries, no more than 34%, but those on which there was most agreement were in Bars 19-24 and 38-46. When the perceived strength of each boundary is disregarded, it is clear that the temporal locations of phrase boundaries are consistent in different performances of Prelude but not in the Finale.

Participants varied in their judgements of where phrase boundaries occurred even though it might be thought that, with a piece as familiar as Chopin’s Prelude, they would agree strongly on its general structure. The variability of judgements was constrained by the compositional structure of the piece (supporting Spiro, 2007); judgements of phrase boundary location in the Finale, with its more ambiguous structure, varied to a greater extent.

**Discussion of Results with Relationship to Audio Cues**

The three performances of each piece used as stimuli are at first glance not obviously different from one another, except for the measurements of overall tempo and dynamics, which in themselves only give us an idea of each performance. As it is only at the microstructural level that we are likely to find the majority of differences between performances of classical music, particularly Romantic piano music, these small variations in tempo and dynamics at locations of particular structural importance in the composition are discussed with respect to the results of the analysis of perceived phrase boundaries.

Two performances of this Prelude by the same performer were analysed by Eric Clarke (1995), one of which, amongst other expressive features, emphasised the division of this piece into two at Bar 13 by utilising a large ritardando in Bar 12, and the other of which accelerated into Bar 13 and in doing so made less of this half-way division. We arrive at a similar conclusion to Clarke’s, that despite there being small local differences in tempo and dynamics, it is the placement of these in the context of the performance that can lead towards an overall different interpretation. This in turn is what appears to be perceived by the audience, although as a group, they appear to have attended to different levels of phrase structure.

There was much less agreement between participants (< 34%) as to the locations where phrase boundaries occurred in performances of the Finale. All pianists performed in accordance with the presto marking, with F3’s performance showing the
highest amount of tempo variation, whereas the other two performances maintained a more steady pulse throughout. F1 was the performance with the least amount of tempo variation. Despite being marked *sotto voce* by Chopin, our three performances all used very varied dynamics. The introduction of the D flat theme and the lowest pitch of the melody at Bar 23 was characterised in all three performances with a large decrescendo but tempo variations were minimal, with only F3 showing a slight ritardando. This may explain why this particular boundary was not strongly perceived, and was only marginally more so for the F3 performance (33% audience agreement on this boundary in F3 as opposed to 20% and 27% for F1 and F2 respectively). The half-way point of the Finale at Bar 39 is characterised in all three performances with a large decrescendo, but only in performances F2 and F3 was this accompanied by a marked decrease in tempo. Looking at the results of the analysis of perceived phrase boundaries, judges make a phrase boundary response around Bar 17 of the Finale. This is not indicated in the analysis as a particular turning point in the composition. Nevertheless, descending chromatic arpeggios are succeeded by rising scales. This would suggest that listeners’ responses were influenced in the absence of expressive markers (other than the ritardando produced by F3 at this point) by the changing direction of the melodic line. That being said, the rate of agreement on this boundary was lowest for F3’s performance (20%) than those of F1 (27%) or F2 (33%).

Barolsky’s (2007) analysis of performances of the Chopin Finale by Cortot and Pogorelich show a contrast in the treatment of the recapitulation at Bar 39, such that the former uses minimal dynamics in order to mask the changing harmonies and the latter produces a dramatic crescendo to highlight the return of the initial material. The three performers in the present study also used dynamics to highlight particular passages while maintaining a constant tempo; when these passages involved quickly-changing harmonic progressions listeners may have been confused. This might of course be intentional on the part of pianists who view this piece as one long unrelenting passage.

Our main hypothesis, that phrase boundaries would be emphasised in performances with both ritardando and diminuendo (as in Friberg et al.’s [2006] phrase-arch rule) and as such would be perceived by all participants, was not upheld for all performances of both pieces. It appears that some phrase boundaries in the Prelude on which there was considerable agreement, particularly those in Bar 21, were often characterized by a ritardando and crescido. Yet those phrase boundaries in the Finale characterized the same way were not identified by all participants; agreement rates were below 34% for all phrase boundaries. This suggests that musical context is important, as well as familiarity: listeners have to recognize the feature of the composition that is being indicated by the performer’s variation of tempo and/or dynamics if they are to agree on its structural importance. Our final hypothesis was that performances varying tempo and dynamics simultaneously would be segmented by participants into more phrases than those varying tempo and dynamics independently. This was indeed the case for P3’s performance of the Prelude, for which there were high rates of agreement on phrase boundaries in Bars 5 and 9 as
well as those identified in Bars 13 and 21. Low rates of agreement between participants’ responses preclude support for the hypothesis in relation to the Finale.

**General Discussion**

There was a range of responses to each individual performance indicating that participants did not perceive the music the same way. However, most participants’ responses changed slightly as they listened to successive performances of the same piece. Their style of response was constant in that they attended to the same hierarchical level of phrasing in each of the different performances. Learning effects could explain the change from identifying longer phrases when the piece was heard for the first time to segmenting into smaller sections on subsequent hearings of the same piece. This finding was not consistent, however: some participants began by segmenting into small sections and ended by identifying longer phrases. Thus listeners can choose to focus their attention on different aspects of the music from small motifs to its overarching structure. They also used the slider in different ways, although each participant’s style of using the slider was stable across different performances: it is clear from the continuous responses shown in Figures 5 and 6 that most participants, with the exception of Participant 5, made gradual movements of the slider.

According to listeners’ responses to the different performances, the performers did not differ in their overall interpretation of the compositional structure of the E minor Prelude of Chopin – in other words they segmented it more or less into the same phrases – but they did emphasize different elements by varying tempo and dynamics, supporting the findings of Spiro (2007). They did not all, however, use ritardando and decrescendo to emphasise phrase endings. Rather, listeners identified major sectional boundaries where both tempo and dynamics were varied simultaneously at particular points of harmonic or melodic interest. Perceptions of the location of phrase boundaries differed even though the differences between the three performances were at the microstructural level and therefore attributable to very small differences between the performers’ interpretations of the same pieces. It might have been interesting to compare listeners’ responses to strikingly different interpretations, particularly of the E minor Prelude, such as those analysed by Barolsky (2008) and Clarke (1995) but the present results indicate the subtlety of performers’ use of tempo and dynamics not just locally, as suggested by Krumhansl (1996) but in relation to the larger context. As we have seen, participants did not agree that there was a phrase boundary in Bar 17 of the Finale, where F3 made a large ritardando; they were more in agreement on the phrase boundary in Bar 9 of the Prelude where P3 combined a ritardando and diminuendo at a change of register than in P1’s performance. It may also be the case that, as shown by Vines et al. (2006) and MacRitchie, Buck and Bailey (2009) the use of gesture in performance helps audiences identify phrase boundaries.
How do the findings contribute to the philosophical discussion of musical performance? The results are consistent with the view that performers, intending to present the musical work in compliance with the score, may present the same structure or the same aspects of a given structure in different ways by varying tempo and dynamics (there are other parameters, but these two are commonly considered to be the most powerful means of expression) so as to emphasize different elements, such as phrases and their boundaries, and thereby strengthening or weakening the relationships between them. Since this interpretation of the findings derives from listeners’ responses to performances, and these differ, we have to assume that audiences perceive musical structure at different hierarchical levels: shorter phrases at lower levels and longer phrases at a higher level.

How can different performances be performances of the same piece? This is not a trivial question for those interested in the artistic value of a performance qua performance of a particular musical work and in comparing the artistic value of different performances of the same work. To say that performers intend to present the structure of a work is not to specify the elements of that structure. Are phrases structural elements? Or are they better construed as the performer’s way of presenting the structure? It may not be possible to answer this question. Phrasing is often considered part of the performer’s toolkit for presenting the structure of a piece but conversely many scores include explicit instructions to the performer (not just recommendations) and in some cases the phrases are so securely embedded in the configuration of the notes that no phrase markings are necessary: even a hurried prima vista reading will reveal them.

If the artistic value of a performance qua performance of a musical work depends on how clearly and accurately the performance presents the work’s structure, and if phrasing is an important tool to present the structure, our evaluation of a given performance will likely look at the phrases’ effects within the larger musical context. For example, if the preferred conception of the final movement of Chopin’s op. 35 is that of a “long unrelenting piece”, Pogorelich’s highlighting of the recapitulation may be deemed artistically inferior to Cortot’s underplaying the formal boundary. Thus the question of the phrasing’s coherence with the rest of the musical piece needs to be considered, rather than whether it is done with emphasis or not at all. Apart from musical coherence, the other consideration of artistic value may be originality. If empirical research into listeners’ responses to the phrase segmentation of a given performance is to go beyond finding out which types of performance act (e.g. variations in tempo, dynamic, articulation, timbre, balance) make phrases more or less audible, it might profitably look into the question which phrase segmentation in a given piece of music is mainstream and which is a deviation from the norm. Engaging in this kind of research might help us understand more clearly what it is that contributes to the artistic value of a performance of a musical work.

This study explored audiences’ perceptions of the locations of phrase boundaries by using a continuous response method while listening to different interpretations of two pieces of music. However, this method uses a vertical scale, which allowed them to indicate the climax of each phrase but may have undermined listeners’ ability to
indicate their perception of phrases at different hierarchical levels. So as to give participants the freedom to interpret the size of a “phrase” unit in their own terms, they were not given explicit instructions. So, despite the participants all having a high level of musical training, their understanding of the task may have differed. The use of a vertical scale may also have encouraged participants to follow increases and decreases in the dynamic levels of the audio-recording. While the method worked well in that it could be used for recording participants’ judgements in real time, there could be no certainty that different participants were not focusing on different levels of structural hierarchy when listening to the same piece or that they shared the same understanding of the task. It might not have been ideal to assume the Chopin Finale was unfamiliar, nor to use a structurally-ambiguous piece of music intended to be played at such a fast tempo (it is notated in quavers throughout and performances F1 to F3 had median tempos of one dotted crotchet equaling 164-205bpm). However, the results from the analysis of responses to the Prelude showed that, even when the structure of a piece of music is explicit, performers can still have an influence on how the phrasing is heard. Future studies will use a greater variety of pieces and more participants. All the pieces will have an overall structure on which performers agree, but which are nevertheless open to different interpretations. In this way each performance is likely to include the same landmarks, in relation to which participants’ responses can be analysed.

Conclusions

Three performances were given of each of two pieces of music. The performances were relatively similar in terms of tempo and dynamics but contained microstructural differences that may have had an influence on where listeners indicated that they perceived phrase boundaries to occur, using a slider as they listened to each set of performances. Different participants attended to phrasing at different hierarchical levels but were consistent across performances. There was more agreement as to the locations of phrase boundaries in performances of one piece than the other. Phrase boundaries appear to have been associated with each performer’s use of tempo and dynamics generally and also specific changes of tempo and dynamics within the musical context. These preliminary results show that while the act of listening to music is personal and may depend on a number of factors such as general listening history and preferences, different interpretations affect the perception of phrasing. For a piece as familiar as Chopin’s Prelude in E minor, with a clear structure on which listeners might be expected to agree, there were nevertheless differences between the way phrase boundaries were perceived, suggesting that performers’ interpretations are conveyed by different emphases on particular structural features. This has implications for performers who wish to present fresh interpretations of canonical works and communicate them effectively to an audience. Studies such as this show how audiences respond when comparing performances of the same piece, thus potentially redefining the role of performers and the artistic value of different performances.
Exploring the Perceptual Effects of Performers’ Interpretations

References


1 Kendall Walton (1988) calls this ‘portrayal’ of sound patterns; see his illuminating discussion of what it means to perform a composition.

2 By composed style, this article refers to the melody, accompaniment style of the Prelude versus the homophonic octave-apart melody of the Finale.

3 Dynamics and tempo were taken from the key velocities and inter-onset intervals respectively, calculated from the MIDI recordings. MIDI information was gathered from these performances using the Moog Piano Bar for the Finale performances, and the MIDI output of the instrument used for the Prelude performances.

4 This could be seen as otherwise if the piece considered were thick in texture, and choices in melody line would have to be made.

5 The participants represented a wide range of instruments: piano (3), oboe (3), flute (2), harpsichord (1), double bass (1), cello (1), viola (1), violin (1), clarinet (1) and a composer. Four of the participants who did not identify themselves as pianists noted that they had had some additional training on the piano, although this was typically for a low number of years compared to their primary instrument.

6 Some may share Alfred Brendel’s (2010, 264) reservations about Glenn Gould who, according to Brendel, is at times insightful, but often fails to respect the structural and expressive features that constitute the performed work.

7 Consider for example Leonard B. Meyer’s (1973, ch. 2) explanation of why the phrasing of the theme in the Peters edition of Mozart’s piano sonata in A-major K331 is wrong.
Appendix 1: Score of Chopin’s Prelude in E minor Op.28 No.4, lilypond files edited from Mutopia project (www.mutopiaproject.org)
Appendix 2: Chopin’s B Flat minor sonata Op.35, Final Movement, lilypond files edited from Mutopia project (www.mutopiaproject.org)
Biographies

**Jennifer MacRitchie** is a postdoctoral researcher in performance science at the Conservatorio della Svizzera Italiana, Lugano, Switzerland. Her main fields of research lie in performance analysis, the connection between movement and sound and the consequent effects of these in the perception of musical features. She gained her PhD at the Science and Music Research Group, Glasgow University (2011) and an M.Eng in Electronics with Music, Glasgow University (2006). Jennifer also performs regularly on both piano and viola.

**Hubert Eiholzer** is vice director and head of the research department of the university school of music at the Conservatorio della Svizzera Italiana, Lugano, Switzerland. In the past ten years of working at this institution, he has been essential in the school’s transformation to a university school and has created, among other things, the research department focusing on research in music directed at musical production, performance and teaching. He has taught courses in music history, music theory and music philosophy, supervised students’ BA and MA dissertations and has been involved in a variety of ways in the department’s research projects. He has a licencees lettres (1983, philosophy) and diploma in instrumental music teaching (1984, piano) and a doctorate in philosophy (1994, music aesthetics).