The relative importance of aural and visual information in the evaluation of Western cannon music performance by musicians and non-musicians.

Noola K. Griffiths and Jonathon L. Reay

School of Social Sciences, Humanities and Law, Teesside University

Abstract

Aural and visual information have been shown to affect audience evaluations of music performance (Juslin, 2000; Griffiths, 2010); however, it is not fully understood which modality has the greatest relative impact upon judgements of performance or if the evaluator's musical expertise mediates this effect.

An opportunity sample of thirty-four musicians (8 male, 26 female M_{age} = 26.4 years) and 26 non-musicians (6 male, 20 female, M_{age} = 44.0 years) rated four video clips for technical proficiency, musicality and overall performance quality using seven-point Likert scales. Two video performances of Debussy's *Clare de lune* (one professional, one amateur) were used to create the four video clips, comprising two clips with congruent modality information, and two clips with incongruent modality information. The incongruent clips contained the visual modality of one quality condition with the audio modality of the other. It was possible to determine which modality was most important in participants' evaluative judgements based on the modality of the professional quality condition in the clip that was rated most highly.

The current study confirms that both aural and visual information can affect audience members' experience of musical performance. We provide evidence that visual information has a greater impact than aural information on evaluations of performance quality, as the incongruent clip with amateur audio + professional video was rated significantly higher than that with professional audio + amateur video. Participants' level of musical expertise was found to have no effect on their judgements of performance quality.

Keywords: aural information, visual information, multimodality, evaluation, performance

Introduction

Music and movement are intrinsically linked in live performances of music from the Western cannon and form the basis for audience evaluations which allow differentiation between performances. This paper aims to probe the nature of the relationship between aural and visual features in the evaluation of musical performance. Aural and visual modes of expression are dependent upon musical genre, with some genres emphasizing visual modes of expression more than others (Thompson, Graham and Russo, 2005); therefore, for the purposes of this study, the focus of the discussion is restricted to performances of music from the Western cannon. Within this tradition, there has long been the cultural practice of a focus on the sonic properties of music rather than on the event of performance itself (Goehr, 1992) and it is taken for granted that aural aspects of a performance will influence its evaluation. Empirical support is lent to this idea by Juslin (2000), who demonstrated that listeners use aural cues, such as tempo and sound level, to identify the emotional character of a piece. More recently, researchers have sought to understand the role of visual cues in the evaluation of performance. In a meta-analysis, the visual component of performance was found to have a medium effect size on evaluative behavior (Platz and Kopietz, 2012) and listeners have reported perceived differences between musically identical performances as a result of changes to the visual modality (Behne and Wollner, 2011). In addition, visual features that are indirectly associated with music performance have been found to impact upon audience evaluations, for example concert dress (Griffiths, 2010), physical attractiveness and stage behavior of the performer (Wapnick, Mazza and Darrow, 1998; 2000).

There is clear evidence that audio and visual information are linked in music perception, both by the performer in their artistic interpretation and by the audience in their lived experience of the performance. For example, performers' body movements are known to be underscored by structures from the written music and as such, movement can be seen as a physical representation of a performer's interpretative decisions (MacRitchie, Buck and Bailey, 2013). With regard to the audience experience, Broughton and Stevens (2009) found that audience members gave higher overall ratings when they experienced performances in audiovisual format compared to audio alone. Furthermore, in the same study, ratings of performance were higher for a projected performance manner than for a deadpan manner when presented to participants in the audio visual condition than audio only: this suggests that movements are related to expressive intentions and that these intentions are perceived by audience members. In addition to performers' movements being linked to the music and their expressive intentions, audio and visual information are also integrated in audience members' perceptions of performance: both melodic cues and facial expressions influence observers' judgements of the emotional valance of sung intervals, which suggests that visual aspects of performance are integrated with aural content (Thompson, Russo and Quinto, 2008). This perspective is supported by the discovery that audiovisual integration is important in the successful identification of individual musicians (Mitchell and MacDonald, 2012). Furthermore, there is evidence that audience members integrate audio and visual information in judgements of note duration and those ratings of note duration vary as a function of changes in visual information but not audio information (Schutz and Lipscomb, 2007). This body of research demonstrates that visual information is synchronized with both a performer's emotional intention and the audio information they produce; these factors influence audience members' lived experiences and subsequently their perceptions of the performance.

Despite the research outlined above it remains unclear the role each modality plays in shaping audience members' evaluative judgements. Indirect evidence comes from a number of sources but as yet, no direct research has investigated the relative importance of audio and visual information on evaluations of performance quality. There is evidence that individuals make a significantly greater number of comments about musical features of a performance than they do about visual aspects (Killian, 2001), which suggests that aural information is dominant in experiences of performance. However, participants have been shown to be significantly better at identifying competition winners using visual information only, compared to using aural information only (Tsay, 2013; 2014), which indicates a dominance of visual information. In findings that support this, Davidson (1993) reported that visual information was dominant in conveying a range of performance manners. She found that observers shown performances recorded in point light (both audio-visually and vision only) were able to distinguish accurately between three intended performance manners: deadpan, projected and exaggerated. In one case it was in the vision only mode alone that observers correctly identified performance states.

Findings from research into the role of modality on emotional judgements of performance may shed some light onto the relative contribution of modality on performance quality evaluations, as there is evidence of cross-modal effects on emotional evaluation. For example, musical audio stimuli were shown to affect the emotional evaluation of non-musical visual stimuli (Van den Stock, Peretz, Grèzes and de Gelder, 2009). In an attempt to investigate the effect of each modality alone, a growing body of literature has used audiovisual isolation methods, where responses to audio only stimuli are compared with responses to visual only, and audiovisual stimuli. In several cases, it was found that visual information was a better conveyor of emotion than auditory information (Di Carlo and Guaitella, 2004; Livingstone, Thompson, Wanderley and Palmer, 2005). The importance of visual information was also illustrated in a study by Vines, Krumhansl, Wanderley, Dalca and Levitin (2011), who found that variations in performers' expressive intentions had the greatest impact on observers' ratings of emotional qualities of the performance when those performances could be seen, i.e.

in vision only and audiovisual conditions. However, visual information does not dominate judgements of music in all contexts. For example, although audio and visual modalities have been shown to convey different experiences of tension, which is linked to emotional response, they convey similar experiences of phrasing, which is related to perceived structure (Vines, Krumhansl, Wanderley and Levitin, 2006). This suggests that the role of modality may vary depending on the aspect of performance that is being evaluated. There is evidence of an emergent quality in performance when performances are both seen and heard simultaneously. This quality has been found to occur in both evaluative and physiological emotional responses to music (Vines et al., 2006 and Chapados and Levitin, 2008 respectively). Research is scarce that examines the role of modality on judgements of performance using all audiovisual stimuli, which contain that additional property. One such study investigated the relative importance of audio and visual information on perceived and felt emotions of performance using an audiovisual manipulation paradigm (Krahe, Hahn and Whitney, 2013). Clips of performance were used where musical material was either congruent with the performer's body movement, or incongruent and was taken from a different performance. They found that both audio and visual information determine which emotions are felt and perceived by observers. The emotional impact of 'sad' musical stimuli varied depending on the accompanying visual information but the authors suggest that there are limits to the impact of the visual channel depending on the nature of the audio information. To date, there has been no research into the ways in which audience members use aural and visual information to evaluate performance quality or proficiency.

One factor that may be important when considering the relative influence of aural and visual information is the expertise of the listener. Differences in auditory acuity between expert and novice musicians have been well documented. For example, musicians detect changes in pitch faster and more accurately than non-musicians (Tervaniemi, Just, Koelsch, Widmann and Schröger, 2005); musicians are better able to process immediate temporal information (Rammsayer and Altenmuller, 2006); and multiple studies have found that musicians have superior timbre recognition compared to non-musicians (Münzer, Berti and Peckmann, 2002; Chartrand and Belin, 2006). However, Thompson and Russo (2007) found that musical training did not affect participants' ability to identify pitch information using visual information gained from watching singers' faces.

With regard to visual information, musicians have been shown to perceive differences between manners of stage behavior in both audio and audio-visual conditions, whereas non-musicians perceived differences in an audio-visual condition only (Huang and Krumhansl, 2011). However, Juchniewicz (2008) found that although ratings of musical aspects of a performance increased as the amount of expressive movement shown increased, there was no effect of musical training on participants' ratings. Similarly, Tsay (2013), found that expert and novice musicians did not differ in their abilities to identify competition winners using audio, visual, or audio-visual information.

The above review of the literature has shown that both audio and visual information contribute to audience members' perceptions of performance. There is evidence that both modalities are integrated in music perception and represent both musical content and a performer's expressive intentions. While research exists that shows that in certain contexts visual information is dominant in audience members' perceptions e.g. selecting a performance winner (Tsay, 2013) and communicating performance manner (Davidson, 1993), there is as yet no direct research into the relative importance of modality on evaluations of performance quality or proficiency. This is an important issue on which to gain clarity, as evaluations of performance quality occur frequently in formal and informal contexts ranging from performance examinations to audience members' post-concert discussions. We know that the role of modality may vary with the aspect of performance that is being evaluated (Vines et al., 2006), and as such we cannot assume that findings from other contexts will carry over into the judgements of performance quality. Related research by Tsay (2013) found that participants were more successful at identifying competition winners using visual information than audio information; while this interesting research found a dominance of visual information for this task, the task required participants to focus on perceptions of the performer rather than to evaluate the quality of the performance. As selecting a winner may involve participants consciously or unconsciously considering factors other than performance quality, research is required in a specifically performance evaluation context. Although research findings that shed light indirectly on the relative importance of audio and visual information in performance evaluations are mixed, findings from work that investigates the same relationship between modalities in the evaluation of emotion in performance are more unidirectional. In this body of work, visual information has been shown repeatedly to convey emotion more accurately than audio information. Taking findings from these two bodies of work on balance, we hypothesize that visual information will have will have greater importance than audio information for audience members' evaluations of performance quality.

Regarding audience members' level of expertise, evidence is somewhat mixed as to whether musical training affects perceptions of performance. We know that in certain circumstances, musicians and non-musicians behave differently in response to musical stimuli and as such it is important gain the extra level of granularity that investigating musical expertise can give when looking at how audience members use audio and visual information to evaluate performance. There is clear evidence to show that musicians have greater levels of aural acuity than non-musicians and evidence to suggest that non-musicians are more reliant on visual information to make judgements about performances. Therefore, we hypothesize that for musicians the role of audio information will be greater than that of visual information, and non-musicians will rely more heavily on visual information to evaluate performance quality.

The current study

In order to investigate the two hypotheses outlined above, we independently manipulated the quality of auditory and visual components of performances. We accept from research into evaluations of emotion in performance that there is a quality that emerges when a performance is both seen and heard. Within that audiovisual condition, we are interested in the synergistic relationship between audio and visual modalities in audience members' evaluations of performance. The focus of this investigation is not on the additive benefit of multi-modal over uni-modal experience and as such using an audiovisual isolation methodology would be insufficient in this case. Therefore, the current study uses an audiovisual manipulation paradigm and for the basis of the methodology, draws on research into evaluations of emotion by Krahe, Hahn and Whitney (2013), outlined above, which used congruent and incongruent versions of performances as their stimuli. In the study reported here, two recordings of a piece were used, one by an amateur and one by a professional pianist. For the purposes of this study, we use the terms 'professional' and 'amateur' in the sense of skill level with the terms corresponding to musicians performing at an elite and sub-elite standard respectively. As well as two clips that were created using congruent audio and visual material, two incongruent clips were created where audio and visual information were juxtaposed. Participants rated the four performances on aspects of performance quality. This approach allowed comparison of the importance of each modality relative to the other in audience members' judgements of performance quality.

The specific measures by which we appraise performance quality are also important. Technical proficiency, musicality and overall performance quality were selected as evaluative measures of performance quality as they permitted assessment of theoretically distinct aspects of performance that are seen as key to a well-rounded performance: both technical mastery of the instrument and the musical understanding of the performer were measured explicitly, as well as an overall judgement. These measures are commonly used in research on performance evaluation (cf. Thompson and Williamon, 2003; Thompson, Williamon and Valentine, 2007; Griffiths, 2008, 2010). Thompson and Williamon (2003) included these measures in their work investigating methods of performance evaluation as they were taken directly from guidelines of the Associated Board of the Royal Schools of Music, a system ubiquitous in UK music education. Previous research has shown that these three concepts, although correlated, do appear to be distinct. Audience members have been shown to exhibit different patterns of ratings of technical and musical aspects of a performance over time (Thompson, Williamon and Valentine, 2007), which suggests that listeners can and do perceive differences in these aspects of performance and are able to discriminate between them. By using these three measures, we were able to gather both segmented evaluative data, i.e. on specific aspects of performance, and a holistic measure of overall performance quality, thus covering the two most common forms of evaluation. Ratings of overall performance quality have been shown to be distinct from a summative score created from segmented evaluations (Mills, 1991) and as such, we recorded a rating of this quality in addition to ratings of specific aspects of the performances.

Method

Design

A mixed measures design was employed. The between-subjects factor was musical training (musician vs non-musician). 'Musician' was defined as an active music-maker with at least Grade 8 ABRSM instrumental or vocal practical examination or equivalent experience. A 'non-musician' was defined as an individual who is not an active music-maker and who has no formal musical training. Ambiguous cases, for example where an individual had received formal training to a reasonably high level during school years but was no longer musically active as an adult, were excluded. The within-subjects factor was performance viewed and had four levels: $\text{Professional}_{\text{audio}}$ + $\text{Professional}_{\text{video}}$; $\text{Professional}_{\text{audio}}$ +$\text{Amateur}_{\text{video}}$; $\text{Amateur}_{\text{audio}}$ + $\text{Professional}_{\text{video}}$; $\text{Amateur}_{\text{audio}}$ + $\text{Amateur}_{\text{video}}$.

The present study used two recordings of Debussy's *Clare de lune*, one by an amateur pianist and one by a professional pianist. The skill level of the performer was used as a way to ensure that there were real differences between the two congruent conditions. The independent variable was performance viewed, as described above. Therefore in the incongruent conditions, which contained both higher and lower quality aspects of performance, it is possible to identify which modality was most important in participants' evaluative judgements based on which clip was rated most highly. Congruent and incongruent pairings of audio and visual information were made and participants were asked to evaluate each performance for technical proficiency, musicality and overall performance quality on a 7 point Likert scale, where 1 was 'low' and 7 was 'high'. We showed in the Introduction above, that a performer's movements are synchronised with and related to their auditory output; as such, in order to avoid a confound of synchronicity, it was necessary for both congruent and incongruent clips to be asynchronous. We used the established practice (cf. Thompson, Russo and Quinto, 2008) of pairing different instances of each modality from the same performance in the congruent clips. Details of how this was achieved appear in the Materials section.

Participants

Thirty-four musicians (8 male and 26 female, M_{age} = 26.4 years, *SD* = 7.04, age range 20-51 years) and 26 non-musicians (6 male and 20 female, M_{age} = 44.0 years, *SD* = 14.48, age range 18-71 years) completed the study. Participants were recruited using an opportunity sample and consisted of staff and students of Teesside University, Trinity Laban Conservatoire of Music and Dance, and the University of Sheffield; members of City of Birmingham Choir and Hartlepool Music Society. No incentives of either cash or course credit were given for participation.

Materials

Musical performances

Two video performances (one professional and one amateur) of Debussy's *Clare de lune* were obtained from www.YouTube.com in line with the website's Fair Use provision. *Clare de lune* was used in the present study as it is a piece for solo piano, which prevents characteristics of any accompanying musicians from influencing participants' ratings of performance quality. Furthermore, this piece is written in ternary form with three distinct sections, ABA, and so it was possible to isolate the A sections for use as test material (see Appendix 1). The two video performances were selected as both performers played the piece at around ♩. = 64.

There were intentional inherent differences between the performances: we aimed for there to be a difference in performance quality so that when comparing ratings of the two incongruent clips, if either one were rated more highly than the other, this would suggest that the participants were placing greater weight on the modality of the performance that corresponded with the professional performance. Both performers were female and of white ethnicity; the pianists' facial expressions were clearly visible but their hands were not in sight. Only the performance space and not the wider auditorium was shown in both performances. The professional performance was given on a black grand piano and the camera position changed once from a head and shoulders view of the performer around five feet from the camera, to a full view of the pianist and piano around fifteen feet from the camera. The camera was positioned end on to the piano, facing the performer. The performance took place on a wooden floor and was lit from overhead. The amateur performance was also given on a black grand piano from a fixed camera point, which gave a full view of the pianist and piano around ten feet from the camera. The camera was positioned side on to the piano giving a view of the performer's right hand side. This performance took place in an educational setting, with carpet on the floor and a display on the wall, and was lit from overhead. Both clips were assessed by two musicians, who took no further part in the study, to verify that the audio and visual quality of the recordings (rather than the quality of performances) was matched.

Video clips

Four separate 90 second performances were created using Final Cut Pro 7 software: $\text{Professional}_{\text{audio}}$ + $\text{Professional}_{\text{video}}$; $\text{Professional}_{\text{audio}}$ +$\text{Amateur}_{\text{video}}$; $\text{Amateur}_{\text{audio}}$ + $\text{Professional}_{\text{video}}$; $\text{Amateur}_{\text{audio}}$ + $\text{Amateur}_{\text{video}}$. To create the congruent clips in which either both professional audio and video, or amateur audio and video were present, the audio from the A1 section of each performance was combined with video of the A2 section from the same performance. For the incongruent clips, audio and video from A2 sections of the performances were combined.

Audio levels were equalized between performances to ensure that sound level was consistent. Audio and video tracks were adjusted on a frame by frame basis to ensure the best fit for the two incongruous clips. A fade in and fade out, each of 2.5 seconds, was used to begin and end each clip. The four performance videos were uploaded to the lead researcher's YouTube account, which is used only for research purposes. Under 'Advanced settings', comments and ratings were disabled so that no comments or 'likes' could be left to influence participants' views. The videos were unlisted, which allowed the clips to be embedded in the online survey tool but prevented the clips from being discovered by a general search.

Questionnaire

Participants engaged with this study via online questionnaires created using Bristol Online Survey. The survey collected information about the participants' age, sex, and musical training. The second section of the questionnaire contained a video of a performance and three, seven-point Likert-type scales upon which participants rated the performance for (a) technical proficiency, (b) musicality and (c) overall performance quality, where 0 was low and 7 high. Participants received the instructions:

> "Please watch the performance and rate it on the scales below.
>
> **Technical proficiency** is defined as the level of instrumental competence shown by the performer, e.g. technical security, rhythmic accuracy.
>
> **Musicality** is defined as the level of musical understanding communicated by the performer, e.g. expressiveness, interpretative imagination.
>
> **Overall performance quality** is defined as your general impression of the quality of the performance."

Participants were only able to move onto the next performance after completing all ratings scales for the preceding performance.

The final section of the questionnaire allowed participants to make open comments about the rationale for their ratings and thanked them for taking part in the study. To control for order effects, performance order was counterbalanced across participants. Furthermore, participants were unable to return to previous screens on the questionnaire to alter their responses in light of subsequent items.

Procedure

Performances were randomized into two orders of presentation and participants were randomly allocated to one of these presentation orders. A link to the relevant online questionnaire was emailed to potential participants. Participants completed the questionnaire using their own computer equipment and were asked to complete the experiment in one sitting with no interruptions. Participants followed on-screen instructions and viewed each performance in turn (order was dependent upon random allocation to counterbalanced presentation order). After each performance, participants rated each performance on the three criteria listed above. Results were automatically stored in a Bristol Online Survey database. Ethical approval was granted by the School of Social Sciences and Law Ethics Committee.

Results

To investigate the relative importance of audio and visual information on audience members' evaluations of performance, a MANOVA was carried out. Results of this analysis are reported first, followed by results of the univariate ANOVAs from the MANOVA, and finally the results of pairwise comparisons are reported to explore the direction of effects.

Ratings of technical proficiency, musicality, and overall performance quality were analyzed across the four performances ($Professional_{audio}$ + $Professional_{video}$; $Professional_{audio}$ +$Amateur_{video}$; $Amateur_{audio}$ + $Professional_{video}$; $Amateur_{audio}$ + $Amateur_{video}$) and across two levels of musical training (musicians; non-musicians). To carry out the analysis IBM SPSS Statistics v23 was used. MANOVA analyses confirmed that there was a significant multivariate effect of performance: λ = .267, F (9, 50) = 15.274, p = < .001, $\eta_p{}^2$ = .733. There was no significant multivariate effect of musical training, λ = .979, F (3, 56) = 0.409, p = .747, $\eta_p{}^2$ = .021; and no significant multivariate interaction of performance and musical training, λ = .765, F (9, 50) = 1.711, p = .111, $\eta_p{}^2$ = .235. Table 1 contains details of means and standard deviations of audience members' rating of performance quality, where 0 = low, 7 = high.

"Insert Table 1 here"

Univariate mixed-measures ANOVAs showed a significant main effect of performance on all three dependent variables: technical proficiency ($Professional_{audio}$ + $Professional_{video}$ *M* = 5.68, *SD* = 1.08; $Professional_{audio}$ +$Amateur_{video}$ *M* = 4.68, *SD* = 1.19; $Amateur_{audio}$ + $Professional_{video}$ *M* = 5.15, *SD* = 1.34; $Amateur_{audio}$ + $Amateur_{video}$ *M* = 3.52, *SD* = 1.23), F (2.44, 141) = 51.3, p = <.001, $\eta_p{}^2$ = .469; musicality ($Professional_{audio}$ + $Professional_{video}$ *M* = 5.82, *SD* = 1.10; $Professional_{audio}$ +$Amateur_{video}$ *M* = 4.45, *SD* = 1.43; $Amateur_{audio}$ + $Professional_{video}$ *M* = 5.20, *SD* = 1.31; $Amateur_{audio}$ + $Amateur_{video}$ *M* = 3.55, *SD* = 1.32), F (2.30, 134) = 43.5, p = <.001, $\eta_p{}^2$ = .429; overall performance quality ($Professional_{audio}$ + $Professional_{video}$ *M* = 5.73, *SD* = 1.12; $Professional_{audio}$ +$Amateur_{video}$ *M* = 4.33, *SD* = 1.22; $Amateur_{audio}$ + $Professional_{video}$ *M* = 5.13, *SD* = 1.40; $Amateur_{audio}$ + $Amateur_{video}$ *M* = 3.33, *SD*

= 1.20), F (2.09, 121) = 54.4, p = <.001, η_p^2 = .484. There was no significant effect of musical training on any of the dependent variables: technical proficiency, F (1, 58) = 1.25, p = .268, η_p^2 = .021; musicality, F (1, 58) = 0.894, p = .348, η_p^2 = .015; overall performance quality, F (1, 58) = 1.083, p = .302, η_p^2 = .018. There was no significant interaction between performance and musical training on any of the dependent variables: technical proficiency, F (2.44, 141) = 1.32, p = .272, η_p^2 = .022; musicality, F (2.30, 134) = 1.21, p = .305, η_p^2 = .020; overall performance quality, F (2.09, 121) = 2.32, p = .100, η_p^2 = .038.

In order to explore the relative importance of audio and visual information on ratings of technical proficiency, musicality, and overall performance quality pairwise comparisons were carried out, see Figure 1.

"Insert Figure 1 here"

The congruent professional performance was rated significantly higher than the congruent amateur performance and both incongruent performances. Across the three dependent variables, $\text{Professional}_{\text{audio}}$ + $\text{Professional}_{\text{video}}$ (technical proficiency *M* = 5.68, *SD* = 1.08; musicality *M* = 5.82, *SD* = 1.10; overall performance quality *M* = 5.73, *SD* = 1.12), was rated significantly higher than $\text{Professional}_{\text{audio}}$ +$\text{Amateur}_{\text{video}}$ (technical proficiency *M* = 4.68, *SD* = 1.19; musicality *M* = 4.45, *SD* = 1.43; overall performance quality *M* = 4.33, *SD* = 1.22), (technical proficiency: p = <.001, 95% CI, lower bound = 0.64, upper bound = 1.31; musicality: p = <.001, 95% CI, lower bound = 0.95, upper bound = 1.78; performance quality: p = <.001, 95% CI, lower bound = 0.96, upper bound = 1.75), $\text{Amateur}_{\text{audio}}$ + $\text{Professional}_{\text{video}}$ (technical proficiency *M* = 5.15, *SD* = 1.34; musicality *M* = 5.20, *SD* = 1.31; overall performance quality

M = 5.13, *SD* = 1.40), (technical proficiency: p = .001, 95% CI, lower bound = 0.21, upper bound = 0.78; musicality: p = <.001, 95% CI, lower bound = 0.28, upper bound = 0.90; performance quality: p = .001, 95% CI, lower bound = 0.25, upper bound = 0.86) and $\text{Amateur}_{\text{audio}}$ + $\text{Amateur}_{\text{video}}$ (technical proficiency *M* = 3.52, *SD* = 1.23; musicality *M* = 3.55, *SD* = 1.32; overall performance quality *M* = 3.33, *SD* = 1.20), (technical proficiency: p = <.001, 95% CI, lower bound = 1.72, upper bound = 2.52; musicality: p = <.001, 95% CI, lower bound = 1.79, upper bound = 2.66; performance quality: p = <.001, 95% CI, lower bound = 1.92, upper bound = 2.75).

The congruent amateur performance was rated significantly lower than the congruent professional performance and both incongruent performances. Across the three dependent variables $\text{Amateur}_{\text{audio}}$ + $\text{Amateur}_{\text{video}}$ was rated significantly lower than $\text{Professional}_{\text{audio}}$ + $\text{Professional}_{\text{video}}$ (technical proficiency: p = <.001, 95% CI, lower bound = 1.72, upper bound = 2.52; musicality: p = <.001, 95% CI, lower bound = 1.79, upper bound = 2.66; performance quality: p = <.001, 95% CI, lower bound = 1.92, upper bound = 2.75), $\text{Professional}_{\text{audio}}$ +$\text{Amateur}_{\text{video}}$ (technical proficiency: p = <.001, 95% CI, lower bound = 0.83, upper bound = 1.47; musicality: p = <.001, 95% CI, lower bound = 0.48, upper bound = 1.24; performance quality: p = <.001, 95% CI, lower bound = 0.69, upper bound = 1.28) and $\text{Amateur}_{\text{audio}}$ + $\text{Professional}_{\text{video}}$, (technical proficiency: p = <.001, 95% CI, lower bound = 1.24, upper bound = 2.02; musicality: p = <.001, 95% CI, lower bound = 1.22, upper bound = 2.04; performance quality: p = <.001, 95% CI, lower bound = 1.38, upper bound = 2.18).

To determine the relative importance of audio and visual information in evaluations of performance we looked to differences in ratings of the two incongruent performances, in which the clips were made of incongruent audio and visual information. The results suggest that more importance is placed on the visual element of performance than the audio element when evaluating performances on these three measures. Across all three evaluative measures of performance, $\text{Amateur}_{\text{audio}}$ + $\text{Professional}_{\text{video}}$, was rated significantly higher than $\text{Professional}_{\text{audio}}$ +$\text{Amateur}_{\text{video}}$ (technical proficiency: p =.021, 95% CI, lower bound = 0.89, upper bound = 0.07; musicality: p =.004, 95% CI, lower bound = 1.29, upper bound = 0.26; performance quality: p =.002, 95% CI, lower bound = 1.29, upper bound = 0.31).

Discussion

The findings of the current study confirm that both audio and visual information can affect audience members' evaluations of performance quality. As hypothesized, findings revealed that visual information has a greater impact than audio information on observers' judgements of technical proficiency, musicality and overall performance quality. In this study the effect of musical training on evaluations of performance was examined alongside the relative importance of audio and visual information on performance evaluation for the first time. It was hypothesized that musically trained individuals would put greater emphasis on auditory information and non-musically trained would place more emphasis of visual features: this hypothesis was not supported. Musicians and non-musicians were found to use audio and visual information similarly when evaluating musical performance.

Both audio and visual information were shown to affect participants' evaluations of performance quality. In congruent clips, where audio and visual material came from the same performance, the professional performance was rated significantly higher than was the amateur performance across all three measures. Changes in the quality of the audio or visual information, i.e. in the incongruent clips, led to ratings that fell between those given for the congruent clips. So, the performance that combined both professional audio and visual information was rated the highest of the four clips for performance quality; however, changing either audio or visual information to amateur (in the incongruent clips) moderated ratings of performance quality, which were significantly lower than when both modalities were professional. Conversely, the performance in which both amateur audio and visual information were combined was rated the lowest of the four performances and substituting either the audio or visual with professional information (in the incongruent clips) enhanced ratings of performance quality across all three measures. From this pattern of results we can see that both audio and visual information affect participants' evaluations of performance quality.

We have shown what is implicit in the discourse surrounding music performance from the Western cannon: that changes in the quality of audio information do affect perceptions of performance quality. Previous research into perceptions of emotion in music has shown that audio information is used by people to identify the character and emotional valance of a piece (Juslin, 2000; Thompson, Russo and Quinto, 2008 respectively). We now provide evidence that the quality of audio information affects evaluations of performance quality, both in terms of overall performance quality and in the subsidiary measures of technical proficiency and musicality.

The finding from this study that visual information affects evaluations of performance quality confirms the effect of visual information on evaluations of performance found in previous research, such as Behne and Wollner (2011), Wapnick et al. (1998, 2000), and Griffiths (2010). Findings from the current study extend those of Wapnick et al. (1998; 2000) and Griffiths (2010), who examined the effect of specific aspects of visual information, such as attractiveness, stage behavior and concert dress, on evaluations of performance: the study reported here has shown that when taken holistically, visual information also affects evaluations of performance quality. Behne and Wollner (2011) found that visual information in general can affect ratings of performance quality, however the method of the current study extends our knowledge of this. In Behne and Wollner's (2011) investigation, participants were asked to judge the performance quality of one video in relation to another; as such the effect of visual information on evaluations of performance was studied comparatively. The current investigation obtained a separate rating for each audio visual combination so that the effect of information from each modality could be examined absolutely. We have provided further evidence of an effect of visual information on evaluations of performance quality.

With the research reported here we take the first step to investigate directly the relative importance of modality in the evaluation of performance quality. We hypothesized that visual information would be of greater importance than audio information in audience members evaluations of technical proficiency, musicality and overall performance quality. In the circumstances described in this study, this hypothesis was supported. If audio and visual cues were utilized equally by participants in their evaluative responses then ratings for the incongruent performances, which each consisted of one half amateur and one half professional information, would not differ significantly; however, if we found a significant difference in ratings of these two performances, this would suggest that the professional element of the performance that was rated more highly is the modality that was given greater weight by participants in their evaluations of performance quality. We found that $\text{Amateur}_{\text{audio}}$ + $\text{Professional}_{\text{video}}$ was rated significantly higher than $\text{Professional}_{\text{audio}}$ +$\text{Amateur}_{\text{video}}$ and as the professional element of the higher rated clip was visual information, this suggests that the visual modality was given greater weight by audience members in evaluating performance quality. While the findings reported here provide evidence that audience members rely more heavily on visual than audio information to evaluate aspects of performance quality, it is not yet clear whether the dominance of visual information shown here would exist in other performance situations where the nature of the audio and visual information was different. For example, the type of music being performed may affect audience members' balance in using audio and visual information to evaluate performance. As audio and visual modes of expression differ with genre (Thompson et al., 2005), the importance and attention placed on each modality within a musical tradition may differ. Audience members' familiarity with the musical content of the performance may also affect their relative reliance on different modalities. In performances of novel or unfamiliar music, visual information may become more important if audience members have difficulty integrating acoustic cues into an existing schema and as such look to other sources of input, i.e. visual information, to make sense of the musical experience. Musical training, discussed below, may also play a role in audience members' use of audio and visual information in different performance contexts. More research is needed in these areas to build up a more nuanced picture of the ways in which audience members make evaluative judgements of performance quality.

The finding that in the current study visual information was of greater importance than audio information in audience members' evaluations supports findings from research into judgements of emotion in performance, which show that visual information in music is a better conveyor of emotion than auditory information (Di Carlo & Guaitella, 2004; Livingstone et al., 2005; Vines et al., 2011). Our finding also supports those from research into the role of modality when selecting a competition winner (Tsay, 2013) or conveying performance manner (Broughton and Stevens, 2009); these studies both showed visual information to be more important than audio information for these tasks. The finding of this study extends the domains in which visual information has been shown to be relatively more important that audio information to include music performance evaluation.

The naturalistic approach taken in this investigation means that differences exist between the recordings used as stimuli. Although we verified that the recordings were of equivalent quality, in future work researchers may wish to control for these differences. In addition, in the study reported here we investigated categorical rather than magnitudinal changes in audio and visual information between professional and amateur performances; although we did not measure the extent of the change in each modality, from the data we are certain that these categorical changes occurred. However, we are unable to say how important an equivalent change in modality is and in future, researchers may want to address this. The procedure of this study involved randomly assigning participants to one of two orders of stimuli presentation to control for order effects; future studies could consider randomizing participants to a greater number of performance orders.

Our finding that visual information is given greater weight than audio information in audience members' evaluations of performance quality adds strength to the argument that visual information is the dominant modality used to make judgements and choices about music performance from within the Western cannon. This is the first time that this has been shown in a performance evaluation context. Furthermore, we have shown this within an audio-visual manipulation paradigm, which allows for the "emergent quality" that Vines et al. (2006) and Chapados and Levitin (2008) have shown to arise when performances are both seen and heard, to form part of audience members' experience of the performances that they evaluated. Although the apparent dominance of visual information in evaluations of performance corresponds with findings from previous research in related areas, it remains somewhat surprising given that sonic qualities are at the core of what musical audiences value about performance (Goehr, 1992; Tsay, 2013). This may be best understood from a cognitive perspective. Kahneman (2011) describes the two-system approach to judgement and choice, whereby System 1 operates automatically and quickly using knowledge stored in memory with little or no sense of effort, and System 2 is associated with effortful mental activities. System

1 provides intuition, impressions and feelings, which System 2 accepts unless an event is detected that does not tessellate with these. Decisions and choices using System 1 are largely accurate but are often associated with certain biases and heuristics; as such, operations by System 2 all require attention, which makes its use more effortful than System 1. Kahneman (2011) describes how we make judgements and choices based on the law of least effort and states that if there are several ways to achieve an end, then we take the least demanding course of action. There is evidence that cognitive load is greater in aural than visual tasks (Klingner, Tversky and Hanrahan, 2011) and visual encoding has been shown to be a relatively automatic process, whereas verbal (aural) encoding is a relatively controlled, and therefore more effortful, process (Lang, Potter and Bolls, 1999). It is likely that when evaluating musical performances, audience members place more weight on visual information than aural information because visual information is used as a heuristic to minimize cognitive effort. Further research is required to ascertain whether visual information is indeed used as a heuristic in this way.

Our second hypothesis stated that musicians would rely more heavily on audio than visual information when evaluating performance quality and non-musicians would place greater weight on visual than audio information. In fact, no effect of musical training was found on ratings of technical proficiency, musicality or overall performance quality, which shows that musicians and non-musicians do not differ significantly in the way that they use audio and visual information to evaluate performances. Previous research showed clearly that differences between musicians and non-musicians exist in terms of their level of aural acuity but findings were mixed as to whether audience members perceived visual aspects of performance differently based on their level of musical training. The finding from the current study supports the perspective that level of musical training does not significantly affect an audience member’s judgements about performance. There is no clear evidence to suggest that musicians and non-musicians differ in their ability to process visual information relating to music performance; therefore, as we have shown that audience members place greater weight on visual information to evaluate performance quality, it is likely that the more well developed aural skills of the musicians are largely irrelevant in this context. The two groups may have a similar skill level to make judgements based on visual information. Further research is required to investigate how musicians and non-musicians evaluate visual musical stimuli and could include formal assessment of their understanding of evaluative measures.

To investigate audience members' evaluations of performance quality, we used three measures that are commonly used in music performance evaluation research: technical proficiency, musicality and overall performance quality. Research has shown that audience members exhibit different patterns of rating technical and musical aspects of a performance over time, which suggests that despite their being correlated, these measures are distinct (Thompson, Williamon and Valentine, 2007). In the current study, ratings of the three measures followed the same pattern across the four performances and while it was beyond the scope of this study to test the distinctiveness of the measures used, it can't be confirmed with absolute certainty that participants weren't making a single evaluative judgment of performance. Further research is required to ascertain the nature of the relationship between these different evaluative measures of performance.

In summary the study reported here has provided evidence that audience members may rely more heavily on visual information than aural information when evaluating music performance quality. Further work is required to investigate the cognitive processes behind this, specifically whether visual information is being used as a heuristic to reduce cognitive load when processing audio visual musical stimuli.

References

Behne, K-E. & Wollner, C. (2011). Seeing or hearing the pianists? A synopsis of an early audiovisual perception experiment and a replication. *Musicae Scientiae.* 15(3), 324-342.

Brase, G. L. & Richmond, J. (2004). The white-coat effect: Physician attire and perceived authority, friendliness, and attractiveness. *Journal of Applied Social Psychology*, 34(12), 2469-2481.

Chartrand, J. P., & Belin, P. (2006). Superior voice timbre processing in musicians. *Neuroscience Letters*, 405(3), 164-167.

Chartrand, J. P., Peretz, I., & Belin, P. (2008). Auditory recognition expertise and domain specificity. *Brain Research*, 1220, 191-198.

Cook, N. (1998a). The domestic gesamtkunstwerk, or record sleeves and reception. In W. Thomas, (Ed.), *Composition, performance, reception: Studies in the creative process in music* (pp. 105-117). Aldershot: Ashgate.

Cook, N. (1998b). *A very short introduction to music*. Oxford: OUP.

Davidson, J. W. (1993). Visual perception of performance manner in the movement of solo musicians. *Psychology of Music*, 21, 103-113.

Elliott, C. A. (1995). Race and gender as factors in judgements of musical performance, *Bulletin for the Council for Research in Music Education*, 127, 50-56.

Frith, S. (1996). *Performing rites: Evaluating popular music*. Oxford: OUP.

Goehr, L. (1992). *The imaginary museum of musical works: An essay in the philosophy of music*. Oxford: OUP.

Griffiths, N. K. (2010). 'Posh music should equal posh dresses: An investigation into the concert dress and physical appearance of female soloists. *Psychology of Music*, 38(2) 159-177.

Huang, J., & Krumhansl, C. L. (2011). What does seeing the performer add? It depends on musical style, amount of stage behavior, and audience expertise. *Musicae Scientiae*, 15(3), 343-364.

Juchniewicz, J. (2008). The influence of physical movement on the perception of musical performance, *Psychology of Music*, 36(4), 417-427.

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology: Human Perception and Performance*, 26(6), 1797.

Kahneman, D. (2011). *Thinking, fast and slow*. London: Penguin.

Killian, J. (2001). The effect of audio, visual and audio-visual performance on perception of musical content. *Bulletin of the Council for Research in Music Education*, 148, 77-85.

Klingner, J., Tversky, B., & Hanrahan, P. (2011). Effects of visual and verbal presentation on cognitive load in vigilance, memory, and arithmetic tasks. *Psychophysiology*, 48(3), 323-332.

Lang, A., Potter, R. F., & Bolls, P. D. (1999). Something for nothing: Is visual encoding automatic? *Media Psychology*, 1(2), 145-163.

Lawson, C. (2002). Performing through history. In J. Rink (Ed.), *Musical performance: A Guide to Understanding* (pp. 3-16). Cambridge: Cambridge University Press.

Madsen, C. K., Geringer, J. M., & Wagner, M. J. (2007). Context specificity in music perception of musicians. *Psychology of Music*, 35(3), 441-451.

Mitchell, H. F., & MacDonald, R. A. (2014). Listeners as spectators? Audio-visual integration improves music performer identification. *Psychology of Music*, 42(1), 112-127.

Münzer, S., Berti, S., & Pechmann, T. (2002). Encoding of timbre, speech and tones: Musicians vs. non-musicians. *Psychologische Beitrage*, 44, 187–202.

Platz, F., & Kopiez, R. (2012). When the eye listens: A meta-analysis of how audio-visual presentation enhances the appreciation of music performance. *Music Perception*, 30(1), 71-83.

Rammsayer, T., & Altenmuller, E. (2006). Temporal information processing in musicians and non-musicians. *Music Perception*, 24, 37–48.

Said, E. W. (1991). *Musical elaborations*. New York: Columbia University Press.

Thompson, W. F., Graham, P., & Russo, F. A. (2005). Seeing music performance: Visual influences on perception and experience. *Semiotica*, 156(1), 203-227.

Thompson, W. F., & Russo, F. A. (2007). Facing the music. *Psychological Science*, 18(9), 756-757.

Thompson, W. F., Russo, F. A., & Quinto, L. (2008). Audio-visual integration of emotional cues in song. *Cognition and Emotion*, 22(8), 1457-1470.

Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., & Schröger, E. (2005). Pitch discrimination accuracy in musicians vs non-musicians: An event-related potential and behavioral study. *Experimental Brain Research*, 161(1), 1-10.

Tsay, C. J. (2013). Sight over sound in the judgment of music performance. *Proceedings of the National Academy of Sciences*, 110(36), 14580-14585.

Tsay, C. J. (2014). The vision heuristic: Judging music ensembles by sight alone. *Organizational Behavior and Human Decision Processes*, 124(1), 24-33.

Walsh, K.C. (1993). Projecting your best professional image. *Imprint*, 40(5), 46-9.

Wapnick, J., Mazza, J. K. and Darrow, A. A. (1998). Effects of performer attractiveness, stage behaviour and dress on violin performance evaluation. *Journal of Research in Music Education*, 46(4), 510-521.

Wapnick, J., Mazza, J. K. and Darrow, A. A. (2000). Effects of performer attractiveness, stage behaviour and dress on evaluations of children's piano performances. *Journal of Research in Music Education*, 48(4), 323-336.

Appendix 1. Music performed in the test material.


2
Tempo rubato
pp
m.d.
peu à peu cresc. et animé
8va
dim. molto

Author Note

Correspondence concerning this article should be addressed to: Dr. Noola Griffiths, School of Social Sciences, Humanities and Law, Teesside University, Borough Road, Middlesbrough, TS1 3BX, UK. E-mail: n.griffiths@tees.ac.uk

Table 1. *Means (SD) of audience members' rating of performance quality, 0 = low, 7 = high.*

	Technical Proficiency	**Musicality**	**Overall Performance Quality**
Overall			
Musicians	4.87 (0.77)	4.85 (0.73)	4.72 (0.66)
Non-musicians	4.62 (0.98)	4.63 (1.00)	4.51 (0.97)
Total	4.76 (0.87)	4.75 (0.86)	4.63 (0.81)
Professional audio + Professional video			
Musicians	5.97 (0.83)	6.03 (0.87)	6.09 (0.79)
Non-musicians	5.31 (1.26)	5.54 (1.30)	5.27(1.31)
Total	5.68 (1.08)	5.82 (1.10)	5.73 (1.12)
Professional audio + Amateur video			
Musicians	4.79 (1.10)	4.65 (1.39)	4.38 (1.23)
Non-musicians	4.54 (1.30)	4.19 (1.47)	4.27 (1.22)
Total	4.68 (1.19)	4.45 (1.43)	4.33 (1.22)
Amateur audio + Professional video			
Musicians	5.18 (1.40)	5.24 (1.30)	5.18 (1.34)
Non-musicians	5.12 (1.28)	5.15 (1.35)	5.08 (1.41)
Total	5.15 (1.34)	5.20 (1.31)	5.13 (1.40)
Amateur audio + Amateur video			
Musicians	3.53 (1.08)	3.47 (1.24)	3.26 (1.14)
Non-musicians	3.50 (1.42)	3.65 (1.44)	3.42 (1.30)
Total	3.52 (1.23)	3.55 (1.32)	3.33 (1.20)

Figure captions

Figure 1. Mean ratings of technical proficiency, musicality and overall performance quality. Error bars represent 95% confidence intervals. *$p < 0.05$. **$p < 0.01$. ***$p \leq 0.001$.

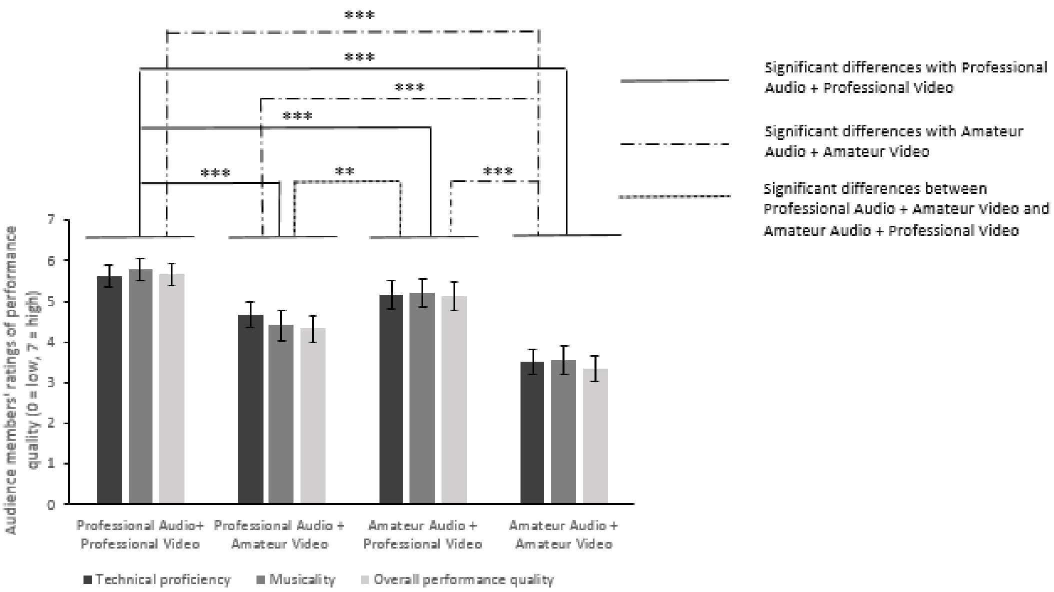

Figure 1. Mean ratings of technical proficiency, musicality and overall performance quality. Error bars represent 95% confidence intervals. *p < 0.05. **p <0.01. ***p ≤0.001.