University of Glasgow School of Psychology

Cross-Modal Perception of Emotion from Dynamic Point-Light Displays of Musical Performance

Holly Blyth

Supervised by: Dr. Frank Pollick March 2015

Abstract

Cross-modal interactions in the perception of emotion in musical performance were investigated in this study. Although it has been widely assumed that the expressive qualities of music are specifically auditory properties, it has also long been known that expressive body movements constitute a form of non-verbal communication. This study addresses the contribution of the visual modality, in the form of the musician's expressive body movements, to the perception of emotion in musical performance. Thirty participants were presented with either a visual (VO) auditory (AO) or audiovisual (AV) point-light representation of a violin performance and rated their perceived emotional intensity of the piece using the GEMS-45 (Zentner et al., 2008) on a Likert scale of 1-5. The intensity of ratings for 3 target words (sad, tearful & sorrowful) were compared, as well as the results of a brief MEQ (used to define participants as either musicians or non-musicians) in relation to participants' intensity ratings under each condition. Analysis of results revealed a significant contribution of being able to see a musician's expressive body movements to the perceived emotional intensity of a musical performance, providing contradictory evidence to the consensus that the expressive qualities of music are solely auditory properties. Evidence from data analysis also suggests that novices perceive a lower level of emotional intensity when they can only hear the music. It is suggested here that this is a result of novices' greater reliance on the visual modality to determine the emotional content of a musical excerpt.

Key Words: Perception Emotion Music Biological Motion Non-Verbal Communication

Contents

1. Introduction

Music has been described as possessing 'a unique ability to trigger memories, awaken emotions, and to intensify our social experiences' (Molnar-Szakacs & Overy, 2006). The emotional response to music that we experience is often spontaneous and effortless. Music constitutes a form of non-verbal communication (Malloch & Trevarthen, 2009) and recent evidence suggests it may even have preceded complete spoken language (Mithen, 2007; Hagen & Hammerstein, 2009). The majority of people report that, in their everyday lives, they primarily employ music to regulate their mood (North et al., 2004). The significance of music as an affective and communicative construct in people's lives has led to an extensive body of research addressing the nature of music and how we relate to it.

Prior to the availability of recording technology, it was not possible to hear music without being exposed to the expressive visual experience of a musician performing a piece of music. At a time when audio recordings of music are so readily available on the internet or on other media, live music performances arguably remain the quintessential experience of music. This has been identified as true across many musical traditions (Frith, 1996). It is likely that one explanation of why live performances remain of such high value to a majority of people is that they allow people to experience not just the sound of the music but also the expressive gestures of the musicians as they play. Although music is a primarily auditory construct, live music presents a multimodal experience, which is expressed through both auditory and visual modalities.

The scientific study of music perception has undergone significant growth in the last few decades (Chapados & Levitin, 2008). However, a comparatively small amount of research has been devoted to the emotionally expressive qualities of music (Vines et al., 2005). The ubiquity and availability of the radio over the past century has contributed to the neglect of research into the emotional perception of audiovisual music (Behne & Wollner, 2011). In the past, it was assumed that, concerning the emotional perception of music, it was auditory information that defined the domain. Visual cues were previously denounced as peripheral to the meaning of music (Bergeron & Lopes, 2009). Until recently, research concerning the multisensory integration of the visual and audio modalities has been limited to speech perception (McGurk & MacDonald, 1976; Vroomen & de Gelder, 2004). The expressive gestures of a musician when performing are considered analogous to the para-linguistic gestures which accompany speech (Wanderley et al., 2005). Similarly to these gestures, which have a pivotal role in enhancing speech, the expressive movements used to convey musical ideas both reinforce and clarify intentionality in music. Research into the biological motion of musicians is capable of identifying some of the fundamental processes underlying inter-personal communication, as it complements research into the relationship between speech and the gestures which accompany it (Wanderley et al., 2005). It has been suggested that the reason that expressive body movements of musicians are able to convey emotional intentionality is that the visual component of a musical performance enhances the observer's identification with the performer (Vines et al., 2006). This raises the question of whether identification is a mechanism for expression.

Although some research has begun to address the issue, the question remains of how musical emotions are perceived and through which modalities this occurs. Past research suggests that it is derived solely from the auditory component of the music and mediated by musical experience (Davies, 1994; Lerdahl, 2001). More recently, research has supported the incorporation of the visual modality in the transduction of emotion in music performance (Lopes & Bergeron, 2009; Vines et al., 2006; Vines et al., 2011; Chapados & Levitin, 2008; Krahé et al., 2015). Recent research has showed that bodily expression can affect an observer's interest in the music (Broughton & Stevens, 2009), facilitate discrimination between various emotionally expressive intentions (Dahl & Friberg, 2007) and influence an observer's perception of the duration of a note (Schultz & Lipscomb, 2007).

1.1 Ancillary Gestures

The expressive movements made by a musical performer, which are not essential to the production of musical sound, are referred to as 'ancillary gestures' (Wanderley, 2002), and are the focus of vast range of research. These gestures have an intrinsic connection with the music, constituting a link between the music itself and the expressive ideas of the musician. This component, which is common to all cultures of music (Blacking 1973), may have developed from an ancient connection with ritual, dance and ceremony. Just as in dance, the primary function of these gestures is considered to be the communication of the performer's expressive intention (Wanderley & Vines, 2006). Although typically unintended, and similar to the expressive gestures which accompany speech (Wanderley, 2002), these ancillary gestures serve to communicate information which clarifies the emotional intentions of the musician. A study by Thompson & Luck (2012) provided support for this by identifying that musicians equate expressionless playing with playing without non-essential movements. A vast body of research has addressed the involvement of expressive body language in combination with sound in the transduction of musicians' expressive musical intentions (Vines et al., 2011; Krahé et al., 2015; Vines et al., 2006; Behne & Wollner, 2011; Chapados & Levitin, 2008; Platz & Kopiez, 2012; Thompson & Luck, 2011). Some specific studies have been carried out which focus on a particular aspect of expressive body language, namely biological motion (Davidson, 1993; Dahl & Friberg, 2007; Wanderley et al., 2005). It has been identified that, even in the absence of audio stimuli, people are able to differentiate between musicians playing in a restrained manner and musicians playing in a normally expressive manner (Davidson, 1993). Being able to see the musicians' ancillary gestures has been identified to be of particular benefit to non-musicians (novices), who can more successfully identify the manner in which a musician is playing when they could only see the musician performing compared to when they could only hear the musician performing (Davidson, 1993). Truslit (1938) translated by Repp (1993) describes the biological motion of the musician not as an anatomical necessity for the purpose of controlling the instrument, but as a spontaneous expression of 'inner motion', which is the manifestation of experience and a factor involved in emotional communication. It appears that performers' ancillary gestures occur in pursuit of expressive goals. A limitation in previous research concerned with the biological motion of musical performers was the primary investigation of piano performance as a musical example. This subject matter offers a restricted range of motion cues because of the fixed nature of the piano and the seated position of the musician. As a result, ancillary gestures are limited to the hands, shoulders and head (Behne & Wollner, 2011; Davidson, 2007; Thompson & Luck, 2011). This study addresses this limitation through the investigation of a solo-violinist, a considerably more mobile instrument, allowing the musician to express a broader range of ancillary gestures.

1.2 Motion Capture Technology

The shift in focus of the domain of emotional perception of music towards the performers' movements has led to the use of motion-capture and video-analysis techniques. The advance of these technologies has provided the possibility of measuring movements using only reflective markers attached to the body and no wires (Palmer & Deutsch, 2013). Point-light displays can be generated by motion-capture technology and can be used to convey the kinematics of the human body. Although it has been argued that point-light displays reduce the ecological validity (Bergeron & Lopes, 2009) they serve to reduce many confounding factors such as facial expressions (Bradley et al., 1997), attractiveness (Grammer et al., 2003) and gender (Hess et al., 2000) and allow kinetic information to be foregrounded. During human perception of motion, visual information retrieved from biological motion (originally termed and described by Gunnar Johansson, 1973) and the related figurative contour patterns (the shape of the body) are interconnected. McDonnell et al., (2009) identified that differing body shapes have an impact on the range of emotion perceived from biological motion. Point-light displays are able to negate the effects of the figurative contour patterns by representing the motions of the main joints with an array of points of light. The process of motion-capture provides detailed visual data that must be aligned with the acoustic information. Motion-capture technology has provided a new insight into the domain of emotion perception in musical performance.

1.3 The Present Study

This study utilised a Vicon motion-capture system (Vicon Systems, 2009), allowing recording of the accurate spatial location of each joint of the body at all times, using reflective markers attached to the body, whose position is recorded with the use of infrared cameras, with very fine temporal resolution. This study investigates the question of through which perceptual channels an observer derives the emotional content of a musical performance, by comparing the reports of perceived emotional intensity between participants who either saw, heard or both saw and heard a point-light display of a violin performance, created using the Vicon motion-capture system.

Hypothesis 1 postulates that participants exposed to the audiovisual (AV) condition will report higher ratings of perceived emotional intensity than participants exposed to either audio-only (AO) or visual-only (VO) conditions.

Vines et al (2011) offer the possibility that any ambiguity in the auditory component of music can be resolved with assistance from the musician's ancillary gestures, which elaborate on the expressive emotional intention of the music. Therefore, Vines et al. (2011) pose the question of whether the visual component of music will be more heavily relied on by the observers who are not familiar with the music, in attempt to identify the emotional intention of the music. It has been suggested that observers who are unskilled at listening rely completely on huge visual component to identify emotional content. A study by Davidson (1993) identified that, even under conditions which only allow participants to see the biological motion of the performer without any audio input, participants were still able to differentiate between musicians playing in a restrained manner and musicians playing in a normally expressive manner. Non-musician participants (novices) were identified as relying more heavily on the visual stimuli in order to discern the emotional expressive intentions of the musician. This is identified by their ability to more successfully identify the version being performed when the visual stimuli was presented, compared to when the audio stimuli was presented, whereas musician participants did not display this difference in ability. Prior to exposure to the relevant stimuli, each participant completed a Brief Musical Experience Questionnaire (Werner et al., 2006), allowing them to be categorised as either expert or novice musicians, as defined by the dimension of Innovative Musical Aptitude.

Hypothesis 2 postulates that novices will make lower ratings of emotional intensity than experts under the audio-only (AO) condition.

2. Methods

2.1 Participants

Thirty student participants (16 male, 14 female) from the University of Glasgow participated in this study as volunteers. Age of participants ranged from 18-52, with a mean age of 26.5 and standard deviation of 3.4. All participants were naïve to the purpose of the experiment. The participant group consisted of both musicians and non-musicians (as defined by the Brief MEQ (Werner et al., 1992).

2.2 Design

Participants were randomly allocated to one of three conditions: audio-only, visual-only or audiovisual. In the audio-only condition (AO), participants heard the musical performance in the absence of the visual component. In the visual-only (VO) condition, participants saw the musical performance in the absence of the auditory component. In the audiovisual (AV) condition, participants both saw and heard the musical performance.

A between-subject design was employed to control for the possibility of participants in the VO condition being influenced by their memory of the sound of the performance, or participants in the AO condition being able to visualise the musician's movements from their memory of the performance. Although this design decreased the power of the statistical analysis, it served to control for any possible carry-over effects as a result of being exposed to all 3 presentation conditions.

2.3 Materials

2.3.1 Musical Excerpt

The musical excerpt selected for this study was taken from a traditional music piece known as a slow-air. The term ‘slow-air’ refers to a melody which gives pause for thought and reflection, characterised by its ability to allow the musician to express specific emotional qualities, while influenced by their musical interpretation of the piece. This piece is entitled 'Niel Gow's Lament for the Death of his Second Wife' (1727-1807: see Figure 1 for sheet music). This excerpt was selected for its distinct emotional nature. Only the melody line of the composition was used here, although the piece is sometimes performed with an accompanying harmony on a range of instruments. The excerpt was performed on a standard 4/4 size violin (an instrument previously utilised in research on musical performance gestures, see for example Dahl & Friberg, 2007). This instrument was selected as it allows body movements to be easily observed without the obscuring effects of larger instruments.

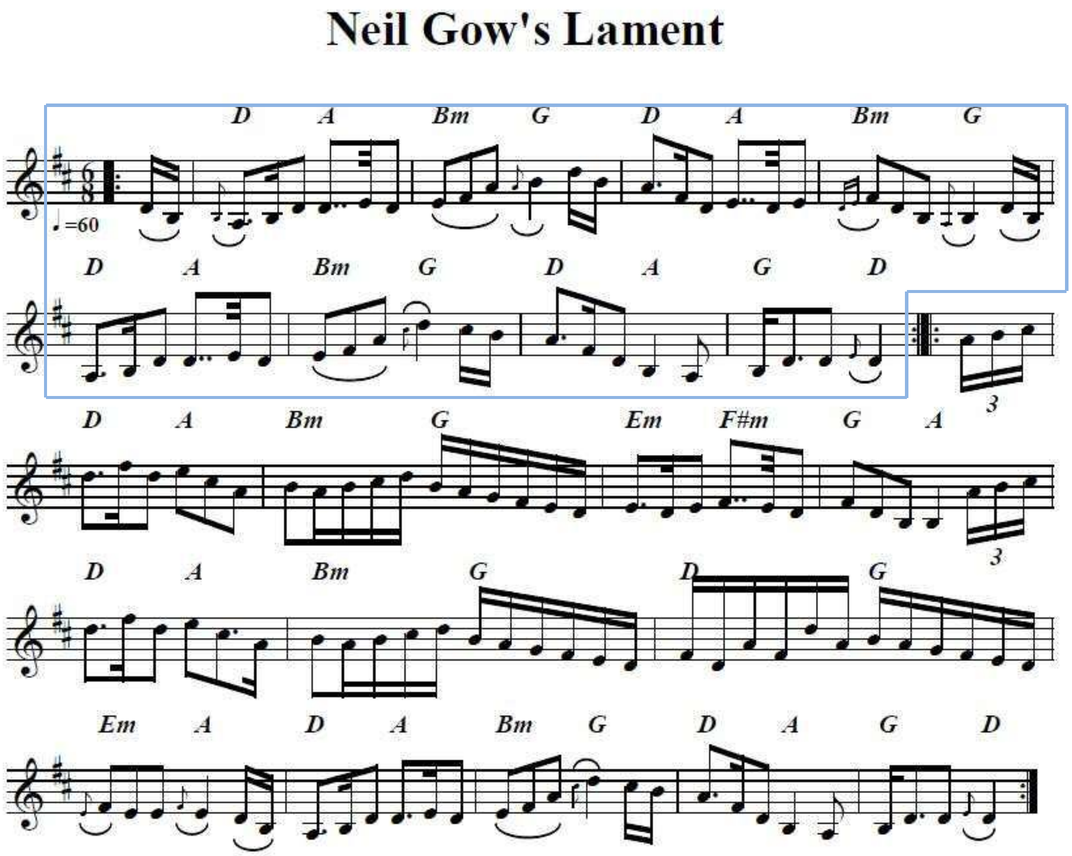


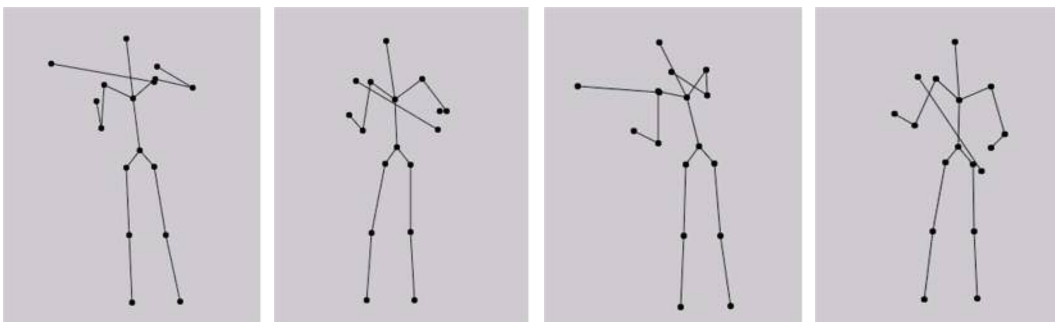
Figure 1: Sheet Music: blue box indicates excerpt selected for this study

2.3.2 Audio/Visual Stimuli

The audio/visual stimuli were recorded and created using a Vicon motion capture system (Vicon Systems, 2009) and a high-quality microphone, capturing the biological motion of a female musician while playing the musical excerpt described above. The motion capture system utilised 12 Vicon MX Series Cameras and two Vicon Ultranet units. The musician wore a dark, tight-fitting suit with 41 reflective markers attached at the location of major joints and at specific points on the violin bow, in order to include the bow movements throughout the piece in the recordings. Following practice trials, 9 audiovisual captures (each lasting around 60 seconds) were recorded. After these were processed, one capture was selected for use in the study, based on the quality of the recording. A typical issue encountered in motion capture is the failure of the system to identify the locations of certain markers at certain points and results in the impossibility of reconstructing a point-light display from this capture. The musician’s movements were processed using Vicon Nexus and the recording was then rendered as a point-light visual display showing 16 coordinates at joint centres, using MATLAB scripts (MATLAB Version 7.12.0, The MathWorks Inc, Natick, MA, 2010). The video file produced in MATLAB and the corresponding audio file from Vicon Nexus were then integrated using Adobe Premiere Pro software. This process resulted in the creation of high-quality, 60-second, video, audio and audiovisual files for use in this study (see Figure 2 for screenshots of stimuli).

The stimuli were prepared for presentation such that 2 seconds of blank screen or silence preceded the onset of the video, audio or audiovisual performance. Stimuli were presented on a 10" Apple iPad 2 screen, using Apple EarPods headphones. In the interest of consistency across conditions, participants in the AO condition were requested to look at the blank screen while listening to the performance, and participants in the VO condition were requested to wear headphones while watching the performance, even though no sound was presented.

Figure 2. Examples from stimuli of expressive body movements during musical performance



2.3.3 Brief Music Experience Questionnaire (MEQ)

Prior to observing audio/visual/audiovisual stimuli, participants were requested to complete a brief Music Experience Questionnaire (MEQ; Werner et al., 2006). The Brief MEQ is a self-report measure of individual differences in reactions to music which comprises 6 scales encompassing 53 items asking about a number of issues aimed at assessing the place of music in a person' slide. Rational development and rigorous testing across six countries (Werner et al., 2006) has identified reliability of the Brief MEQ in terms of internal consistency and test-retest reliability. As a supplement to the Brief MEQ, some more specific questions relating to length and type musical training were asked within a demographic form. One scale in particular, Innovative Musical Aptitude, was selected for use in defining expert and novice musicians for further analysis in relation to Hypothesis 2.

2.3.4 Geneva Emotional Music Scale (GEMS)

Following the presentation of audio/visual/audiovisual stimuli, participants completed a Geneva Emotional Music Scale questionnaire (GEMS-45), with edited instructions directing participants to record the emotions they perceived, as opposed to felt (Zentner et al., 2008: edited instructions can be seen in Appendix A). The GEMS-45 provides a taxonomy based on emotional concepts which are particularly relevant in the domain of music. Three items from the GEMS-45 were selected to measure the intensity of the sad quality of perceived emotions: 'sad', 'tearful' and 'sorrowful'. Importantly, the remaining 42 items of the scale were retained as filler items in order to avoid task demands and to distract participants from the purpose of the experiment.

2.4 Procedure

Participants completed the experiment in a private library meeting room, under quiet conditions to ensure they were entirely focused on the task at hand and were randomly assigned to one of three conditions, defined by the three types of stimuli: audio-only (AO), visual-only (VO) or audiovisual (AV). Participants were informed that their participation in the study would involve watching/listening to a musical performance (depending on their group condition) and completing two questionnaires. To begin, participants were provided with some information about the study to read (Appendix B), then requested to sign a standardised consent form informing them of their right to confidentiality and right to withdraw. They were then required to complete a demographic form (Appendix C) and a Brief MEQ (which cannot be attached as an appendix due to copyright). Following this, participants were presented with the appropriate stimuli for their group condition (AO, VO or AV). Subsequently they were required to complete a GEMS-45 with edited instructions (Appendix A) to record their emotional perception of the performance. Participants were thanked and debriefed. The entire process lasted approximately 20 minutes.

3. Results

3.1 **Hypothesis 1:** Participants exposed to the audiovisual (AV) condition will report higher ratings of perceived emotional intensity than participants exposed to either audio-only (AO) or visual-only (VO) conditions.

Considering the distinct negative emotionality of the musical excerpt selected for this study, three target items were identified within the GEMS-45 which were descriptive of typical negative emotion, 'sad', 'tearful' and 'sorrowful'. Also, these target items were identified by the GEMS-45 (Zentner et al., 2008) as belonging to the musical emotion category of 'sadness'.

Cronbach's alpha was conducted for the three target items (sad, tearful, and sorrowful) within each of the three conditions (AV, AO and VO), to determine the reliability of the intercorrelation of the three items within the scale. Cronbach's alpha for the three target items was good to very good with α= .89 for AV, α= .95 for VO and α= .94 for AO. Therefore, an average was taken of the intensity ratings for the three items for each condition (AV, AO and VO), with higher scores indicating higher levels of perceived emotional intensity. The average perceived emotional intensity ratings for each of the three conditions are displayed in Figure 3.

Figure 3. Effect of Group on Perceived Emotional Intensity

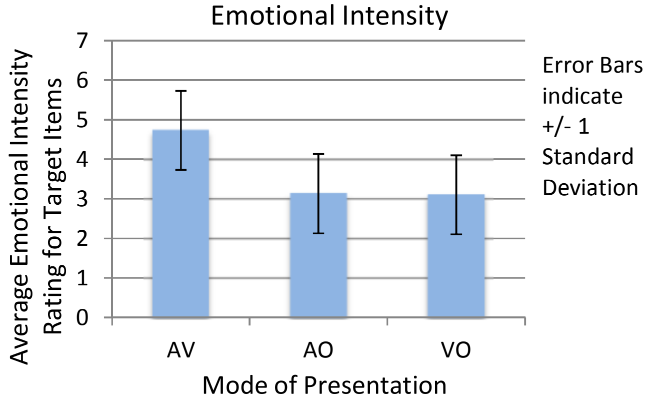


Table 1. Average Intensity Ratings and Standard Deviations by Condition

Condition	Average Emotional Intensity Rating	Standard Deviation
AO	3.13	0.76
AV	4.73	0.41
VO	3.10	0.65

To assess any statistically significant differences in average emotional intensity rating across the three modes of presentation of the stimuli, a Kruskal-Wallis H Test was conducted followed by post-hoc Mann-Whitney U Tests to identify between which presentation conditions these differences exist.

A Kruskal-Wallis H Test was used to determine if there was any statistically significant difference between the effects of the three different modes of presentation (AV, AO, VO) on the average emotional intensity rating for the three target items. A statistically significant effect of mode of presentation on intensity rating was identified ($H(2)=17.872$, $p<0.001$), with a mean rank of 24.9 for AV, 10.9 for AO and 10.7 for VO.

Separate post-hoc Mann-Whitney U tests with Bonferroni correction, used to locate the differences in intensity ratings between conditions, showed significant differences between AV and AO ($U=4$, $Z=-3.55$, $p<0.001$) and between AV and VO ($U=2$, $Z=-3.72$, $p<0.001$). However, no significant difference was identified between AO and VO ($U=50$, $Z=.00$, $p=1.000$).

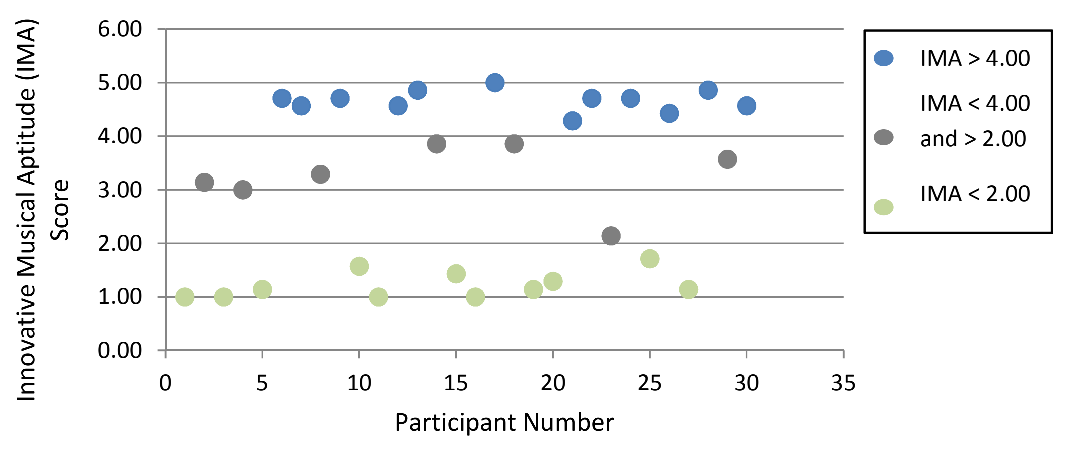
These results support the hypothesis that participants exposed to the audiovisual (AV) condition will report higher ratings of perceived emotional intensity than participants exposed to either audio-only (AO) or visual-only (VO) conditions, as average intensity ratings for the AO condition are significantly higher than for either AO or VO to the extent of statistical significance.

3.2 **Hypothesis 2**: Novices will make lower ratings of emotional intensity than experts under the audio-only (AO) condition.

The Brief-MEQ (Werner et al., 2006) consists of 53 items which assess the place of music in the participant's life and are scored from 1 to 5 on a Likert scale. All participant scores for the 53-items were analysed in SPSS using a Brief-MEQ scoring syntax created by Werner and colleagues (2013).

Cronbach's alpha for the seven items of the Brief MEQ which are attributed to the scale of Innovative Musical Aptitude (IMA: a self-report measure of musical performance ability, including the ability to generate musical themes and works) was very good with α= .98. The thirty participants involved in this study ranged on the scale of Innovative Musical Aptitude from a score of 1.00 to 5.00

Figure 2. Innovative Musical Aptitude of Participants



Using the scale of Innovative Musical Aptitude, 11 participants were defined as novices (IMA score<2.00, identified in green on Figure 2.) and 12 were defined as experts (IMA score>4.00, identified in blue on Figure 2.). Within the range of scores, these upper and lower quartile definitions were chosen to identify a number of participants who would be representative of experts and novices, excluding those in the centre of the range (identified in grey on Figure 2.) of scores who are less clearly definable as expert or novice musicians. Of the 11 participants defined as novices, 4 completed the experiment under the AO condition and of the 12 participants defined as experts, 5 completed the experiment under the AO condition.

Addressing Hypothesis 2 (novices will make lower ratings of perceived emotional intensity than experts under the audio-only (AO) condition), the intensity ratings of expert and novice participants in the AO group were analysed and compared to determine if there was a statistically significant difference. Within the AO condition, 4 participants were defined as novices and 5 participants were defined as experts. The Innovative Musical Aptitude (IMA) score, resultant experience status (expert/novice) and average intensity rating of each participant in the audio-only (AO) condition are shown in Table 2. The average intensity rating for each group within the audio-only (AO) condition (experts/novices) was calculated to provide data for further statistical analysis.

Participant No.	Innovative Musical Aptitude (IMA) Score	Expert/Novice Status	Participant's Average Intensity Rating	Expert/Novice Group Average
5	1.14	Novice	2.00	
11	1.00	Novice	2.33	
19	1.14	Novice	2.67	
27	1.14	Novice	3.00	
				2.50
6	4.71	Expert	4.00	
7	4.57	Expert	4.00	
12	4.57	Expert	4.33	
26	4.43	Expert	3.00	
28	4.86	Expert	3.00	
				3.67

Using a Mann-Whitney U Test, average intensity ratings of the two groups within the audio-only (AO) condition were compared to determine if there was a significant difference present. The Mann-Whitney U Test conducted showed a significant difference between the intensity ratings of experts and novices ($U=1$, $Z=-2.25$, $p<0.05$).

This result provides evidence which supports the hypothesis that novices will make lower ratings of emotional intensity than experts under the audio-only (AO) condition, as intensity ratings of novices were statistically significantly lower than intensity ratings of experts.

4. Discussion

This study aimed to provide support for previous research identifying the visual component as a contributing factor in the communication of the emotional intention of a musical performance (Vines et al., 2011; Krahé et al., 2015; Vines et al., 2006; Behne & Wollner, 2011; Chapados & Levitin, 2008; Platz & Kopiez, 2012; Thompson & Luck, 2011) using motion capture technology and point-light displays to identify the contribution of pure biological motion in isolation from confounding factors such as body shape and facial expressions. A further aim of this study was to investigate in what ways expert and novice musicians vary in their perception of the intensity of emotions from musical performance.

This study confirmed that the target items 'sad', 'tearful' and 'sorrowful' defined in GEMS-45 (Zentner et al., 2008) as belonging to the musical emotion category of 'sadness' were internally consistent for participants' perceptions. Results identify that participants who were presented with the audiovisual (AV) stimuli perceived a significantly higher emotional intensity than those who were presented with either audio-only (AO) or visual-only (VO) stimuli. This finding provides support for previous research and for Hypothesis 1. Expert and novice participants were compared on their intensity ratings when presented with the audio-only (AO) stimuli and it was identified that experts perceived a significantly higher level of emotional intensity than novices when they were only able to hear the performance, providing support for Hypothesis 2, albeit with a limited sample size. The results gained from this study are therefore a testament to the influence of the visual component, which was previously considered extraneous, on the emotional perception of live music performance. Further results of this study highlight the importance of the nature of the observer. Here it is identified that level of expertise also has a significant impact on the emotional perception of live music performance across modalities.

A secondary outcome identified in the course of the analysis of the data obtained from the Brief MEQ (Werner et al., 1995) was participants' responses to items on one of the 6 scales of the Brief MEQ termed 'Affective Reactions' (AR). All participants' responses to items on this scale were significantly higher than the other 5 scales of the Brief MEQ – on a possible scale of 1-5, participants' scores on the scale of AR ranged from 3.70-4.90 with a significantly higher average score of 4.42. Affective Reactions (AR) is calculated by the average score of the items on the Brief MEQ which relate to spiritual and emotional responses to music. This result seems to support the idea that music is of central importance in the majority of people's lives and the conclusion of North et al., (2004) that people primarily employ music in their everyday lives to regulate their mood. This finding has the potential to provide an alternative platform to other measurement methods for future research in the domain of music and affect

The findings of this study fall in line with the underlying theory of music as a form of non-verbal communication. Previous research has identified, using motion-capture technology and point-light displays that expressive body movements are used as a form of non-verbal communication by musicians within group performances as a way of conveying expressive intentions (Glowinski et al., 2013). Juslin (2003) presented the theory that it is the expressive quality which makes musical performance so valuable to us, not solely the 'deadpan' acoustic component. Both musicians and music teachers consider musical expression to be the essential dimension of a performer's skills (Lindstrom et al., 2003). The phenomenon identified here of the effect of expressive body movements on emotional perception of music performance reflect the basic human abilities which underlie non-verbal communication and are not restricted to the music domain. The results of the present research also constitute support for the assumption that musicians are successful in communicating emotion. Previous research suggests that communication is one of the primary functional intentions of music performance, as evidenced by a questionnaire study in which 83 percent of music students state that they attempt to express specific emotions in their performance 'always' or 'often' (Lindström et al., 2003). The research presented here complements research in the domain of speech and the gestures which accompany it, which also constitute a form of non-verbal communication, acting to clarify and further express the verbal component (Graham & Argyle, 1975; Cassell et al., 2001; Goldin-Meadow, 1999). A meta-analysis by Juslin & Laukka (2003) identifies that musicians are able to communicate basic emotions through music performance at least as successfully as through facial expressions and verbal expressions.

The results of this study replicate those identified by previous research (Vines et al., 2006; Vines et al., 2011; Davidson, 1993) and provide an effective platform for further research in this domain. This research contributes to the domain of multimodal emotional perception of music by specifically investigating the impact of one aspect of the visual component (biological motion), while excluding other variables such as facial expressions and body shape, on perceived emotion from a musical performance.

4.1 Realistic Spatial Audio

A potential methodological issue which can be raised with this study is the spatial validity of the audio stimuli when it is conveyed through headphones. It has been argued that the spatial realism of a musical recording can be affected when it is perceived via headphones as opposed to when it is perceived in the same environment in which the music was originally performed during recording (Algazi & Duda, 2011). In order to achieve realistic playback, research suggests that headphone technology is required to advance in its ability to respond dynamically to the movements of the listener through the use of very small, low-power motion sensors (Algazi & Duda, 2011). It is suggested that development of this technology will be able to provide an extremely realistic representation of being immersed in a live music performance. However, the headphone technology currently available can offer various advantages over 3D speaker placement in the context in which it was employed in this study. Firstly, headphones are able to isolate the listener from any extraneous noise or reverberations in the room environment (Gardner, 1998). This ensures that the acoustics of the room or the participants' seated position within the room do not have an impact on their auditory perception (Huopaniemi, 2007). According to Kyriakakis (1998), one of the major limitations of headphone use is 'Inside-the-head Localisation' (IHL) which refers to the issue that sounds are not perceived by the participant as if emanating from outwith themselves but rather as if they originate from inside the head, and are biased towards the rear of the head (Kendall, 1995). IHL occurs as a result of a number of interconnected factors, including a lack of the correct environmental reverberation. Begault (1994) suggests that, other than this limitation of headphone use in perception research, headphones are more effective and even preferable for communicating 3D audio (the sound which occurs in the three-dimensional space surrounding the listener).

4.2 Perceived and Felt Emotions

Previous research has highlighted differences between felt and perceived emotions (Zentner et al., 2008; Gabrielsson, 2002; Kallinen & Ravaja, 2006). The seminal research by Gabrielsson (2002) addressed the question of how the musician's expression of music relates to the listener's emotional response to the music. As a result, a distinction between emotion perceived in a piece of music and emotion induced by a piece of music (perceived vs. felt) was formed. Gabrielsson (2002) describes these two alternatives not as separate concepts but rather as two extremes on a continuum from simple emotionless perception at one end to the experience of an intense emotional reaction at the other. It is suggested that, in reality, as opposed to experiencing one extreme or the other, the majority of observers of music performance experience the emotion of the music somewhere along this continuum. Research in response to Gabrielsson's (2002) paper by Kallinen &Ravaja (2006) suggests that a musical performance can arouse emotions which are very similar to the perceived emotional content of the performance, but levels of perceived and felt emotions differed on various dimensions such as valence and arousal. For example, fearful music was perceived as negative but felt as positive. As a result there is some disparity of opinion on the difference between perceived and felt emotion. A more recent development following Gabrielsson's (2002) paper was documented by Schubert (2007), who identified that felt emotions (internal locus of emotion) fluctuated significantly more across conditions than perceived emotions (external locus of emotion). People appear able to make distinction between these two perspectives on emotions, accepting that emotion can be experienced or it can be observed (Juslin & Laukka, 2004). One theory proposed to explain the quantitative difference of perceived and felt emotions was proposed by Konecni (2008), who suggested that ambiguity in the wording of instructions provided to participants can lead to an inability to identify the target locus (felt or perceived emotion). It was identified that, when there is ambiguity in the wording, and therefore the requirements, of the study, the majority of people tend to carry out the task with reference to their own felt emotion, emphasising the importance of clear and effective communication of instructions in research in this domain.

The nature of musically-induced emotions is a topic still under considerable debate. Some research has questioned the adequacy of non-music-specific emotion measurement tools for application in musical contexts (Vuoskoski & Eerola, 2011). This is a result of the argument that emotions related to music appear to differ from emotions experienced in everyday life (Lykartsis et al., 2013). The GEMS-45 (Zentner et al., 2008) was designed specifically to address this issue. This study employed the Geneva Emotional Music Scale (GEMS-45; Zentner et al., 2008), a domain-specific measurement tool designed specifically for musically induced emotions. The scale was originally designed to measure emotional reactions to music (i.e. emotions felt in response to a musical piece). However, the instructions provided to the participant before completing the GEMS-45 were altered for this study in order to direct participants to record the emotions they *perceived* as opposed to *felt.* Similar adaptations to the instructions provided with the GEMS-45 have been made in previous research (Krahé et al., 2015). However, at present, the validity of the GEMS-45 has only been assessed through research in relation to *felt* emotions. Considering past research suggesting differences between felt and perceived emotions, a dedicated assessment of the validity of the GEMS-45 for the measurement of perceived emotions would be beneficial for the advancement of research in this area and to infer more valid conclusions.

4.3 Choice of musical excerpt

The vast majority of research conducted in this domain has used Western Classical music (for example, Vines et al., 2006; Chapados & Levitin, 2008; Krahé et al., 2015). In particular, music from the Romantic period is often selected because of its ability to effectively evoke and express emotions (Romantic(ism), 2013). However, some studies employed specific musical excerpts for the express reason that participants were unfamiliar with it (Dibben, 2004; Hunter et al., 2010). This study employed a traditional music excerpt (as described above), selected specifically for its 'sad' emotional quality. Participants were asked a number of demographic questions including 'Are you familiar/trained in traditional/folk music?' to which 21 participants responded 'No' and 9 participants responded 'Yes'. In light of past research, such as those studies by Blood & Zatore (2001), Panksepp (1995) and Rickard (2004), another effective method of employing musical excerpts in a study with the aim of evoking intense emotions is to allow the participant, rather than the experiment, to select the musical excerpt. A study by Salimpoor et al., (2009) identifies greater levels of emotional arousal of participants in response to self-selected as opposed to experimenter-selected musical excerpts. Such research has implications for this study in terms of a future methodological revision, as a result of the significant majority of participants who were unfamiliar with the style/genre of the musical excerpt selected. If self-selected excerpts elicit more intense emotional perception and responses, it is possible that more valid conclusions could be drawn from a revision of this study in which participants are able to self-select the musical excerpt they see, hear or both see and hear.

5. Conclusions

As indicated in the above, this study hosted some methodological and theoretical limitations which are debated in the literature and have been discussed. Future research will acknowledge and address these limitations, specifically in terms of clarifying the distinction and effect of felt and perceived emotions, and the method of choosing the musical excerpt observed by participants. Contradicting the broad consensus that the auditory component is at the core of the musical domain, the outcome of this study contributes to an increasing body of research which provides evidence for the equality, if not primacy of the visual component.

6. Acknowledgments

I would first like to thank Dr. Frank Pollick, for his academic guidance, practical expertise and unwavering enthusiasm for this project.

Also, thanks to Dr. Paul Werner and colleagues, for allowing the use of the Brief Music Experience Questionnaire and for their help and technical support in using it.

I would like to express my appreciation to the University of Glasgow Psychology Department and all the staff of the department, for the organisation and support behind this project.

7. References

Algazi, V. R., & Duda, R. O. (2011). Headphone-based spatial sound. Signal Processing Magazine, IEEE, 28(1), 33-42.

Begault, D. (1994). 3-D Sound for Virtual Reality and Multimedia Academic Press. Boston, MA.

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

Bergeron, V., & Lopes, D. M. (2009). Hearing and seeing musical expression. Philosophy and Phenomenological Research, 78(1), 1-16.

Blacking, J. (1973). How musical is man?. University of Washington Press, Washington

Blood, A. J., & Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. Proceedings of the National Academy of Sciences, 98(20), 11818-11823.

Broughton, M., & Stevens, C. (2009). Physical movement and imagery in professional and undergraduate student solo marimba practice. In ISPS 2009: 2nd International Symposium on Performance Science (pp. 531-536). Association Europeenne des Conservatories.

Bradley, B. P., Mogg, K., Millar, N., Bonham-Carter, C., Fergusson, E., Jenkins, J., & Parr, M. (1997). Attentional biases for emotional faces.Cognition & Emotion, 11(1), 25-42.

Cassell, J., Nakano, Y. I., Bickmore, T. W., Sidner, C. L., & Rich, C. (2001, July). Non-verbal cues for discourse structure. In Proceedings of the 39th Annual Meeting on Association for Computational Linguistics (pp. 114-123). Association for Computational Linguistics.

Chapados, C., & Levitin, D. J. (2008). Cross-modal interactions in the experience of musical performances: Physiological correlates. Cognition,108(3), 639-651.

Dahl, S., & Friberg, A. (2007). Visual perception of expressiveness in musicians' body movements. Music Perception, 24(5), 433-454

Davidson, J. W. (2007). Qualitative insights into the use of expressive body movement in solo piano performance: a case study approach. Psychology of Music, 35(3), 381-401.

Davies, S. (1994). The evaluation of music. What is music, 303-27.

Davidson, J. W. (1993). Visual perception of performance manner in the movements of solo musicians. Psychology of music, 21(2), 103-113.

Dibben, N. (2004). The role of peripheral feedback in emotional experience with music. Music Perception, 22(1), 79-115.

Frith, S. (1996). Music and identity. Questions of cultural identity, 108-27.

Gabrielsson, A. (2002). Emotion perceived and emotion felt: Same or different?. Musicae Scientiae, 5(1 suppl), 123-147.

Glowinski, D., Gnecco, G., Piana, S., & Camurri, A. (2013, September). Expressive non-verbal interaction in string quartet. In Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on (pp. 233-238). IEEE.

Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. Trends in cognitive sciences, 3(11), 419-429.

Graham, J. A., & Argyle, M. (1975). A cross-cultural study of the communication of extra-verbal meaning by gestures (1). International Journal of Psychology,10(1), 57-67.

Grammer, K., Keki, V., Striebel, B., Atzmüller, M., & Fink, B. (2003). Bodies in motion: A window to the soul. In Evolutionary aesthetics (pp. 295-323). Springer Berlin Heidelberg.

Hagen, E. H., & Hammerstein, P. (2009). Did Neanderthals and other early humans sing? Seeking the biological roots of music in the territorial advertisements of primates, lions, hyenas, and wolves. Musicae Scientiae,13(2 suppl), 291-320.

Hunter, P. G., Schellenberg, E. G., & Schimmack, U. (2010). Feelings and perceptions of happiness and sadness induced by music: Similarities, differences, and mixed emotions. Psychology of Aesthetics, Creativity, and the Arts, 4(1), 47.

Huopaniemi, J. (2007, March). Future of personal audio: Smart applications and immersive communication. In Audio Engineering Society Conference: 30th International Conference: Intelligent Audio Environments. Audio Engineering Society.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. Perception & psychophysics, 14(2), 201-211.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code?.Psychological bulletin, 129(5), 770.

Juslin, P. N. (2003). Five facets of musical expression: A psychologist's perspective on music performance. Psychology of Music, 31(3), 273-302.

Juslin, P. N., & Laukka, P. (2004). Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening.Journal of New Music Research, 33(3), 217-238.

Kallinen, K., & Ravaja, N. (2006). Emotion perceived and emotion felt: Same and different. Musicae Scientiae, 10(2), 191-213.

Kendall, G. S. (1995). A 3-D sound primer: directional hearing and stereo reproduction. Computer music journal, 23-46.

Konečni, V. J. (2008). Does music induce emotion? A theoretical and methodological analysis. Psychology of Aesthetics, Creativity, and the Arts,2(2), 115.

Krahé, C., Hahn, U., & Whitney, K. (2015). Is seeing (musical) believing? The eye versus the ear in emotional responses to music. Psychology of Music,43(1), 140-148.

Kyriakakis, C. (1998). Fundamental and technological limitations of immersive audio systems. Proceedings of the IEEE, 86(5), 941-951.

Lerdahl, F. (2001). The sounds of poetry viewed as music. Annals of the New York Academy of Sciences, 930(1), 337-354.

Lindström, E., Juslin, P. N., Bresin, R., & Williamon, A. (2003). "Expressivity comes from within your soul": A questionnaire study of music students' perspectives on expressivity. Research Studies in Music Education, 20(1), 23-47.

Lykartsis, A., Pysiewicz, A., von Coler, H., & Lepa, S. (2013). The Emotionality of Sonic Events: Testing the Geneva Emotional Music Scale (GEMS) for Popular and Electroacoustic Music. In Proceedings of the 3rd International Conference on Music & Emotion (ICME3), Jyväskylä, Finland, 11th-15th June 2013. Geoff Luck & Olivier Brabant (Eds.). ISBN 978-951-39-5250-1. University of Jyväskylä, Department of Music.

Malloch, S., & Trevarthen, C. (2009). Musicality: Communicating the vitality and interests of life. Communicative musicality: Exploring the basis of human companionship, 1, 1-10.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. Nature, 264, 746-748

McDonnell, R., Jörg, S., McHugh, J., Newell, F. N., & O'Sullivan, C. (2009). Investigating the role of body shape on the perception of emotion. ACM Transactions on Applied Perception (TAP), 6(3), 14.

Mithen, S. (2007) The Singing Neanderthal. Harvard University Press, Cambridge, MA.

Molnar-Szakacs, I., & Overy, K. (2006). Music and mirror neurons: from motion to'e'motion. Social cognitive and affective neuroscience, 1(3), 235-241.

North, A. C., Tarrant, M., & Hargreaves, D. J. (2004). The Effects of Music on Helping Behavior A Field Study. Environment and Behavior, 36(2), 266-275.

Palmer, C., & Deutsch, D. (2013). Music performance: Movement and coordination. The psychology of music, 405-422.

Panksepp, J. (1995). The emotional sources of" chills" induced by music. Music perception, 171-207.

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: An Interdisciplinary Journal, 30(1), 71-83.

Repp, B. H. (1993). Music as motion: A synopsis of Alexander Truslit's (1938) Gestaltung und Bewegung in der Musik. Psychology of Music, 21(1), 48-72.

Rickard, N. S. (2004). Intense emotional responses to music: a test of the physiological arousal hypothesis. Psychology of Music, 32(4), 371-388.

Romantic(ism). (2013). The Oxford Dictionary of Music, 2nd Ed. Rev. Oxford Music Online. Oxford Universty Press. Available online at: http://www.oxfordmusiconline.com/subscriber/article/opr/t237/e8648

Salimpoor, V. N., Benovoy, M., Longo, G., Cooperstock, J. R., & Zatorre, R. J. (2009). The rewarding aspects of music listening are related to degree of emotional arousal. PloS one, 4(10), e7487.

Schultz, M., & Lipscomb, S. (2007). Hearing gestures, seeing music: Vision influences perceived tone duration. Perception, 36, 888–897

Thompson, M. R., & Luck, G. (2012). Exploring relationships between pianists' body movements, their expressive intentions, and structural elements of the music. Musicae Scientiae, 16(1), 19-40.

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., Dalca, I. M., & Levitin, D. J. (2005). Dimensions of emotion in expressive musical performance. Annals of the New York Academy of Sciences, 1060(1), 462-466.

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., & Levitin, D. J. (2006). Cross-modal interactions in the perception of musical performance. Cognition,101(1), 80-113.

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., Dalca, I. M., & Levitin, D. J. (2011). Music to my eyes: Cross-modal interactions in the perception of emotions in musical performance. Cognition, 118(2), 157-170.

Vroomen, J., & De Gelder, B. (2004). Perceptual effects of cross-modal stimulation: Ventriloquism and the freezing phenomenon. The handbook of multisensory processes, 3(4), 1-23.

Vuoskoski, J. K., & Eerola, T. (2011). Measuring music-induced emotion A comparison of emotion models, personality biases, and intensity of experiences. Musicae Scientiae, 15(2), 159-173.

Wanderley, M. M. (2002). Quantitative analysis of non-obvious performer gestures. In Gesture and sign language in human-computer interaction (pp. 241-253). Springer Berlin Heidelberg.

Wanderley, M. M., Vines, B. W., Middleton, N., McKay, C., & Hatch, W. (2005). The musical significance of clarinetists' ancillary gestures: an exploration of the field. Journal of New Music Research, 34(1), 97-113.

Wanderley, M. M., & Vines, B. (2006). Origins and functions of clarinettist's Ancillary Gestures. Music and gesture, 167.

Werner, P. D., Swope, A. J., & Heide, F. J. (2006). The music experience questionnaire: Development and correlates. The Journal of Psychology,140(4), 329-345.

Zentner, M., Grandjean, D., & Scherer (2008). Emotions evoked by the sound of music: Characterization, classification, and measurement, Emotion, 8, 494-521.

Zentner, M., Grandjean, D., & Scherer, K. R. (2008). Emotions evoked by the sound of music: characterization, classification, and measurement. Emotion,8(4), 494.