A COMPARISON OF PIANO PERFORMANCE EVALUATIONS GIVEN UNDER AUDIO-ONLY, LIMITED AND FULL AUDIOVISUAL CONDITIONS

BY

KELVIN CHAN

A MA Thesis submitted in partial fulfillment of the requirements of the degree of

Master of Arts in Music Education in the Schulich School of Music,

McGill University, Montreal, 2015

ABSTRACT

In recent years, numerous music educators and psychologists have explored the impact of the visual channel on piano performance evaluation. The majority of them found that the visual component of piano performances has a significant impact on ratings of expression, and that variables such as the amount of audiovisual information, performer body movements, performer proficiency, and expressive criteria could also be influential factors in piano performance evaluation. The main goal of the current study was to investigate whether the visual component of piano performances has a significant impact on their ratings in a performance evaluation context. The main contribution of the current study to the piano performance evaluation is its combination of elements in methodological design that is new to research in the field: isolating the hands and arms for study, utilizing a wide variety of repertoire, and employing both expert and amateur performers.

In this study, 60 pianists with at least 10 years of piano training, as well as performing or teaching experience in the last 3 years, rated 10 piano performances under one of three rating conditions: audio-only, limited audiovisual (only the hands and arms of the performer are visible), and full audiovisual (the entire upper body of the performer is visible). The performances featured pieces of diverse styles, as well as pianists of either amateur or professional proficiency. The participants were divided into three groups of 20, with each group being subjected to one rating condition. The participants rated each performance under four expressive criteria: Phrasing, Dynamics, Rubato, and Overall Quality. An independent-sample t-test (one-way ANOVA) shows that a significant difference existed between audio-only and full audiovisual, as well as limited and full audiovisual ratings at the $p < .05$ level, but not between

audio-only and limited audiovisual ones. Three two-way ANOVAs show that interactions between Rating Condition and Performance, as well as Rating Condition and Performer Proficiency at the $p < .05$ level were also found, but not between Rating Condition and Expressive Criteria.

The main finding of this study further reinforces the results of existing research, which demonstrated that the visual component of piano performances has a significant impact on their evaluation. Additional research on the various types of body movements, particularly those of the hands and arms, as well as their impact on the perception of musical expression would be valuable in furthering this topic of study.

Keywords: piano performance evaluation, audiovisual, body movements, performer proficiency, expressive criteria

RÉSUMÉ

Au cours des dernières années, de nombreux chercheurs dans les domaines d'éducation musicale et de psychologie ont exploré l'impact de différentes conditions auditives et visuelles sur l'évaluation de prestations pianistiques. La majorité des recherches a démontré que la composante visuelle des prestations pianistiques influence significativement l'évaluation de l'expression du jeu pianistique et que des paramètres, tels que la quantité d'information audiovisuelle, les gestes du pianiste, le niveau de maîtrise du jeu de l'interprète et les critères expressifs peuvent aussi influer sur l'évaluation du jeu pianistique. Le but principal de cette étude était de vérifier si la composante visuelle du jeu pianistique a un impact significatif en contexte d'évaluation. La principale contribution de cette étude au domaine de l'évaluation du jeu pianistique réside dans la méthodologie employée. En effet, la combinaison des différents éléments méthodologiques, dont la condition audio-visuelle limitée des mains et des bras, la variété de styles de répertoire utilisés, et l'emploi d'interprètes amateurs et professionnels, est originale dans le domaine.

Soixante pianistes, possédant au moins dix années de formation pianistique et ayant eu une expérience de concert et d'enseignement au cours des trois dernières années, ont évalué dix pièces du répertoire pianistique sous l'une des conditions suivantes: auditive seulement, audiovisuelle limitée (seuls les mains et bras du pianiste sont visibles), et audiovisuelle complète (l'ensemble du haut du corps de l'interprète est visible). Dix pièces de styles variés ont été interprétées par des pianistes de niveaux amateur ou professionnel. Trois groupes de 20 participants ont chacun été soumis à une seule condition d'évaluation afin d'évaluer chaque pièce selon quatre critères expressifs: le phrasé, les dynamiques, le rubato, et la qualité

expressive globale du jeu. Une analyse de variance à un facteur (ANOVA) a démontré une différence significative ($p < .05$) entre les conditions d'écoute auditive seulement et audiovisuelle complète ainsi qu'entre les conditions audiovisuelle limitée et complète mais pas entre les conditions d'écoute auditive seulement et audiovisuelle limitée. Trois analyses de variance à deux facteurs (ANOVA) ont démontré des interactions entre les conditions d'évaluation et les pièces, ainsi qu'entre les conditions d'évaluation et les niveaux de maîtrise des pianistes ($p < .05$) mais pas entre les conditions d'évaluation et les critères expressifs.

Ces résultats renforcent les conclusions d'études précédentes selon lesquelles la composante visuelle a un impact significatif sur l'évaluation du jeu pianistique. L'étude supplémentaire de différents types de gestes, en particulier ceux des mains et des bras, ainsi que leur impact sur la perception de l'expression musicale contribuerait à l'avancement des connaissances dans ce domaine.

Mots clés: évaluation du jeu pianistique, conditions auditives et visuelles, geste pianistique, compétence instrumentale, critères expressifs

ACKNOWLEDGEMENTS

I would like to express my sincerest gratitude towards the following individuals and groups:

Dr. Joel Wapnick, for sharing with me a passionate interest in piano performance research and for his enormous patience and support during the formative stage of my thesis project;

Dr. Isabelle Cossette, for guiding me through the experimental and writing stages of my thesis project, with the utmost dedication, encouragement and support;

Dr. Liza Lorenzino, for accepting me into the Music Education program in the first place, and for fostering my development as an aspiring music educator;

Schulich School of Music and CIRMMT for their generous financial and academic support, as well as for providing the state-of-the-art facilities in which my experiment was held;

Professor Michelle Mares at the University of Victoria, my alma mater, for igniting my passion for all things piano, and for nurturing me into not only a finer pianist, but a more mature musician;

Senior colleagues Audrey-Kristel Barbeau and Catherine Massie-Laberge, for their kindness and generosity in lending their expertise by offering me constructive criticism and valuable advice;

All of the pianists who participated in my experiment, without whose contribution this thesis project would not have been possible;

My parents, for their immeasurable sacrifice for, and unconditional contribution to, my academic pursuits;

And finally, my dear sister, Sophia S. Y. Chan, for being the greatest inspiration of my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction

Overview

Piano educators have traditionally neglected the role of the visual component in piano performance evaluation. In preparing their students for a successful performance, be it on the stage of an entrance audition or prestigious competition, piano pedagogues would spend innumerable hours scrutinizing every detail in a musical score, critiquing and fine-tuning every nuance in their pupils' playing. Most, however, probably neglect to encourage students to explore how different physical gestures may be suitable for expressing different musical intents, or strategies performers could use to refine their stage behavior, attire, and overall visual presentation to the audience. In the light of recent research findings in music education and psychology, this perhaps needs to be changed. As Thompson, Graham and Russo (2005) commented, a growing body of empirical evidence supports that the visual component of musical performances significantly influences our experience of music, even though music is often regarded as a "purely aural experience" (p. 220). The visual component has been found to be "more informative than the auditory component in perceptual judgment of a [musical] performer's level of expression" (Luck, Toiviainen, & Thompson, 2010, p. 47), in direct contrast with the long held misconception that sound alone is "sufficient to define musical expression" (Juslin, 2005, p. 88). In particular, many studies in the last 20 years have demonstrated that the performer's body movements play a significant role in the audience's perception of their expressive intention (Dahl & Friberg, 2007; Davidson, 1993; Thompson & Luck, 2012) and performance quality (Gillespie, 1997; Schutz, 2008). Performers' body movement characteristics are also largely affected by, and demonstrative of, their experienced emotions (Zijl & Luck,

2012).

However, despite an acknowledgement among many researchers that visual research on piano performance provides important information to pianists that could help them build strategies to improve their own evaluation (through using their musical intent as a basis), there remains a severe lack of discussion on the topic in the pedagogical literature. This has been true from the first discussions of pianoforte technique by the likes of Clementi (1803), Cramer (1812), Hummel (1828), up until the present. In-depth analyses of physical movements featured in modern method books are limited to the mechanical aspect of piano playing, such as the range of viable muscle movements and postures (Sandor, 1995), or how different parts of the finger could help produce a different tone (Lhevinne & Lhevinne, 1972). A discussion of how visual gestures could affect piano performance evaluations has always been neglected and seldom mentioned even in passing, let alone an examination of specific movements and their connection to expressive intentions. In these pedagogical resources, it is extremely rare to find practical advice on how to create an overall visual presentation that could improve artistic communication with the audience, leading to more convincing performances. This is unfortunate, since a pragmatic, acute understanding of performance evaluation, as well as the factors that influence it could be tremendously beneficial to developing pianists, especially when for the vast majority of them, success at auditions and competitions is essential for career establishment (Tsay, 2013).

In addition, the piano is a particularly suitable instrument for visual research, as the range of physical gestures and motions used in its performance is highly varied. This is due to the piano's immensely vast repertoire as well as both the breadth and depth of technique required to fully

explore its potentials. The layout of the piano on the stage also readily allows its performer's body movements and stage behavior to be observed by an audience, as the layout of the keyboard requires a greater degree of movement from the performer compared to most other instruments, and the pianist is rarely playing from behind a music stand. Researchers such as Davidson (1993, 1994) and McPherson & Schubert (2004) have noted that the physical movements of a pianist are important cues for an audience to judge his or her musicality. Schumann once famously remarked that if Liszt performed behind a screen, a great deal of magic in his playing would be lost (Bergeron & Lopes, 2009). The pianist's physical gestures are an integral part of their performance, not only because musical sound is greatly dependent on the body movements used in its production (Davidson, 2007), but also because the performer's body movements and his or her mental representations of the music being performed are intimately connected (Kochevitsky, 1967; Lehmann & Davidson, 2002). This renders the visual component of a piano performance inherently expressive, and one that is integral to a performance evaluation context in which the pianist is visible to the raters.

Through conducting research on the visual component of piano performance and discussing their impact on performance evaluation, pedagogues will be more informed and more able to promote an awareness of the topic in their students, leading to better preparation for career-making challenges such as entrance auditions and international competitions.

The Research Problem

Previous studies have either allowed the pianist's whole body to be viewed by the raters (Davidson, 1993, 2007; Shoda, et al., 2007), or the upper body only (Juchniewicz, 2008); none of

them has explored the effects of a full view, a limited view, and no view (audio only). In this study, the hands and arms were singled out for investigation under the limited audiovisual condition, as they are assumed to be the most important part of a pianist's playing mechanism in terms of visual interest. Under Delalande's (1988) Three Level Gesture Classification, which divides physical gestures used by musical performers in three levels, "from purely functional to purely symbolic", the movements of a pianist's hands and arms would be considered the "effective gesture", which is necessary for the mechanical production of sound. The "effective gesture" is also referred to as the "instrumental gesture" in recent research by Wanderley (Wanderley, Vines, Middleton, McKay & Hatch, 2005), as opposed to the "ancillary gesture", which has no direct involvement in sound production (but is often indirectly involved, such as the back leaning closer to and farther from the keyboard). The paramount role that hand gestures plays is also supported by the fact that individualistic finger kinematic patterns have been suggested to be an important contribution to the unique tone of a pianist (Bella & Palmer, 2011). A limited visuals condition was included in the current study as it allowed for the examination of the impact of hand and arm movements alone. Ratings of piano performances given under three conditions: Audio-only, limited audiovisuals, and full audiovisuals were compared, with the goal of exploring the main research question of *whether or not the rating conditions would affect ratings for musical expression*. The same set of performances were rated by experienced pianists under each rating condition: in the limited audiovisual condition, only the hands and arms of the performers were visible to the observers; in the full audiovisual condition, the entire upper bodies of the performers, as well as the performance setting around them, were visible to the observers.

The relationship between four expressive criteria (Phrasing, Dynamics, Rubato, and Overall Quality) and the pianist's body movements had been suggested by Juchniewicz (2008) as one of the prominent questions arisen from recent research, and he had also used them in his own study (Juchniewicz, 2008). Wapnick, Mazza, & Darrow (2002) had also employed essentially the same criteria for their study on the effect of attractiveness, stage behavior and dress on children's piano performance evaluation, and numerous other researchers have suggested that using these criteria could help musicians enhance their performances (Gabrielsson, 1999; Gabrielsson & Juslin, 1996; Johnson, 1996; Nakamura, 1987; Palmer, 1989). As many researchers have used these particular criteria in previous research on the topic, I asked participants under each condition to rate performances by four expressive criteria: Phrasing, Dynamics, Rubato, and Overall Quality in order to study *whether or not there was an interaction effect between rating condition and these expressive criteria.*

As performances of different repertoire, performers and performer proficiency (amateur vs. professional) were used, two other research questions of this study were *whether or not there is an interaction effect between rating condition and performance* and *whether or not there was an interaction effect between rating condition and performer proficiency.* The former question would explore the role of musical selection and performer recognition in the ratings; while the latter would help determine if different audiovisual information influenced the ratings for amateur and professional performances differently.

The Purpose and Significance of the Study

The main purpose of this study is to add to a growing body of existing research on the visual

channel in a performance evaluation context, by investigating the effect of the visual component of piano performances. Traditionally, sound has traditionally been considered by both novice and expert judges to be the most important channel, with the visuals often being overlooked (Tsay, 2013). Due to recent research indicating that visual information in musical performances is likely more influential in evaluation than most musicians realize, it is valuable for music educators to develop a fuller understanding of the phenomenon through further study and discussion. In the current study, as the performers' hands and arms become visible in the limited audiovisual condition, the assumption was that the additional information conveyed by the physical movements would allow observers to assess the performers' musicality and expressive intentions differently compared to the audio-only condition, which would result in significantly different (whether higher or lower) ratings. In the full audiovisual condition, where other visual elements such as performance setting, performer attire, stage behavior and upper body movements are included, the ratings were expected to depart from the audio-only condition to a greater extent, and in the same direction. It was possible that they would differ significantly from those of the limited audiovisual condition as well.

Not only will the results of this study contribute to a growing knowledge of how the visual component affects performance evaluation, they will also have powerful implications for educators, especially piano pedagogues. If the visual channel is in fact as important as the audio channel for musical expression, then piano pedagogues will do well to incorporate more discussion of it into their teaching, whether in educational resources or in the studio classroom.

CHAPTER I
Literature Review

Music Education and Visual Research

The visual component of musical performances in general is very seldom, if at all discussed in pedagogical literature such as method books. Despite the fact that no study on or survey of these teaching resources has even pointed out this defect, prominent researchers in music education and psychology have noted the importance of visual research and how an awareness of it is sorely lacking in the field.

One of the pioneers of visual research in music perception, particularly on performer movements, Jane Davidson (1993) was among the first to suggest that "whilst music itself is an aural phenomenon, its visual components seem to be significantly and empirically neglected sources of perceptual information" (p. 103). Her study on the visual perception of a pianist and four violinists' expressive intentions, discussed further in this chapter, remains hugely influential today, inspiring researchers such as Ohgushi (2006) and Shoda et al. (2007), whose studies are also examined in a later section. More recently, Thompson, Graham, and Russo (2005) also regretted the fact that although there is already empirical evidence that indicates our experience of music is highly influenced by its visual aspects, the music pedagogy and theory fields still largely consider it exclusively an aural one.

Fortunately, since Davidson (1993), more and more researchers have explored visual research from a music education, specifically performance evaluation, perspective. Their findings have

been very telling and illustrative of the visual component's influence: through mock auditions and recital recordings, Wapnick (1997, 1998, 2000) discovered that stage behavior and attractiveness significantly affect the evaluation of various kinds of performers, whether they are vocalists, pianists, or violinists, male or female, adult or children; McPherson and Schubert (2004) concluded that body movements, in addition to attractiveness and flair, are among the main performer-related aspects of performance evaluation, the awareness of which is important to performers seeking to enhance their own assessment; Tsay (2013) showed how our "natural" and "nonconscious" reliance on the visual channel could affect piano performance evaluations given by even the most trusted experts, thereby impacting the outcomes of major competitions that so many aspiring pianists depend on in order to establish a career; Morrison, Price, Smedley and Meals (2014) demonstrated that the level of expressiveness shown in conductors' gestures significantly influence the ratings of their ensemble's expressiveness, regardless of actual performance quality.

A common thread that runs through the work of these researchers is the demonstration that the visual component of musical performances has real and direct influence on the way performers are evaluated, as well as the suggestion that this discovery has powerful implications for music education. With a much needed introduction of the topic to teaching resources, educators and students alike would be better informed and therefore better equipped to undertake their pedagogical and performance endeavors. In general, as will become apparent in the following sections of this chapter, visual research in music perception has been gathering momentum in recent years, with most studies producing congruent findings that could, and should lead to a greater awareness of the topic in performance evaluation, and by extension, the field of music

education.

Overview of Existing Research

The visual channel and its ability to communicate expressive ideas in human interactions have been studied extensively by researchers in contexts both related and unrelated to music. It has been found to be instrumental in transmitting emotional information, in mundane actions from drinking and lifting (Paterson, Pollick, & Sanford, 2001; Pollick, Paterson, Bruderlin, & Sanford, 2001) to elaborately choreographed dancing (Dittrich, Troscianko, Lea, & Morgan, 1996; MacFarlane, Kulka, & Pollick, 2004; Sörgjerd, 2000; Walk & Homan, 1984). Movements of the head, face and hands have also been demonstrated by researchers to improve the effectiveness of verbal communication (Campbell, Dodd, & Burnham, 1998; Munhall, Jones, Callan, Kurate, & Vatikiotis-Bateson, 2004).

In the venue of musical performance, the visual channel has been shown by many researchers to be a crucial component of artistic communication. It has often been demonstrated to improve the audience's perception of the performers' expressive intentions and/or performance qualities (Castellano, Villalba, & Camurri, 2007; Dahl & Friberg, 2007; Davidson, 1993; Gillespie, 1997; Thompson & Luck, 2012; Schutz, 2008), but also sometimes to detract the attention of the audience from the performance if the performer seems to be emotionally uninvolved, or display odd stage behavior (Finnäs, 2001).

Recently, a growing amount of recent research has been devoted to exploring the effects of the visual channel on musical performance evaluation. In the past 20 years or so, studies on topics

ranging from expert ratings of student violinists' performances under audiovisual versus audio-only conditions (Gillespie, 1997), the effect of performer attractiveness on performance ratings (Wapnick, Mazza, & Darrow, 2000), to the perception of expression in a conductor's gestures (Luck, Toiviainen, & Thompson, 2010) have shown that the visual component of musical performances play an influential role in their evaluations. In particular, these studies often either highlighted the effect of the visual component compared with the sound component (Davidson, 1993; Krahé, Hahn & Whitney, 2013; Platz & Kopiez, 2012; Shoda et al., 2007), or investigated the effect of visual information on performance ratings directly (Dahl & Friberg, 2007; Juchniewicz, 2008; Williamon, 1999; Vines, Wanderley, Krumhansl, Nuzzo, & Levitin, 2004). A particularly remarkable finding common to all of these studies, is that the visual component of a musical performance has a significant effect on the performance evaluation. This phenomenon has often been investigated by comparing evaluations of musical expression given under different audiovisual conditions, which is also the method utilized in the current study (see Methodology chapter, p. 49). Before a range of major studies are examined in detail, it is necessary to identify the ways musical expression has been defined by researchers who have done work on the topic.

Rating Musical Expression

In performance evaluation research, musical expression has been rated based on its different aspects, with the most frequently chosen ones being (a) performers' expressive intention, such as "deadpan" or "exaggerated" (Davidson, 1993, 1994, 2002; Ohgushi, 2006; Vines, Krumhansl, Wanderley, Dalca, & Levitin, 2005), (b) audience's emotional impression, based on a variety of moods such as "depressing", "cheerful" and "fear" (Dahl & Friberg, 2007; Shoda et al., 2007;

Vines, Krumhansl, Wanderley, Dalca, & Levitin, 2005), and (c) expressive criteria that are based on musical elements, such as “phrasing”, “dynamics”, and “rubato” (Juchniewicz, 2008; Wapnick, Mazza, & Darrow, 2002). These three aspects have all been used in research on the visual component of musical performances. Each of the studies mentioned above, as well as their approaches, will be further discussed in more detail in this chapter.

As for the current study, the design utilizes expressive criteria to measure musical expression, instead of expressive intention or emotional impression. The rationale is that although the latter two variables allow valuable insight into the transmission and reception of musical expression to be gained, respectively, they focus on emotional feelings and responses rather than critical judgements of the performances. Also, in the case of investigating the expressive intention aspect, the performer(s) must be instructed to play in a certain way (e.g. the aforementioned “deadpan”, “projected”, and “exaggerated”), which greatly limits their expressive freedom and interpretation of the given music, an issue which would be further discussed in the chapter. Since it is a goal of the current study to investigate how piano performances are evaluated in terms of their quality of expression, and specifically individual expression, it is more appropriate to keep performance instructions to a minimum. Naturally, it is also more appropriate to utilize an assessment rubric based on the ratings of expressive criteria (i.e. in this case phrasing, dynamics, rubato, and overall quality) for the judges, as is often done in musical institutions (Wapnick & Darrow, 2012).

In the following sections, major studies in visual research in the context of performance evaluation are presented and examined. According to their main findings, these studies have

been organized in four categories: audiovisual rating conditions, amount of visual information, body movements and expression, and rater and performer proficiency. A table summarizing the design and key finding of the most important studies is presented at the end of each section.

Audiovisual Rating Conditions

In a pioneering study, Davidson (1993) carried out two separate point-light experiments in which performers had pieces of reflective tape attached to various parts of their bodies (head, elbows, knees, ankles, hips and shoulders) that reflected a set of theatre lights that shone directly at them. A video camera then recorded only the reflecting beads of light of each performer. The result was that the observers in the experiments could only see the performers through their body movements as traced by points of reflected light, and no other aspects of their appearance. In the first of the two aforementioned experiments, Davidson (1993) recruited 21 undergraduate music students to view and rate four violin performances. According to the instructions given, the four violinists each played a different excerpt with three expressive manners, “deadpan”, “projected” (a moderate degree of expression), and “exaggerated”. Then, each of the excerpts was edited down to audio only, video only and audiovisual conditions. The observers were asked to rate each excerpt under each condition on a seven-point scale, from deadpan on one end, and exaggerated on the other. The results indicated that enough information was provided in the body movements of the performers alone to allow observers to identify the expressive manner (video only condition). They also showed that the video only condition conveyed the biggest differences between the three expressive manners, showing the highest difference in score between the deadpan and exaggerated performances. Davidson (1993) concluded that the visual component is the most effective indicator of performance manner, even more so than the audiovisual.

In the second experiment, Davidson (1993) asked 34 music students to rate excerpts from a performance of Mussorgsky's *Pictures at an Exhibition*, under audio-only, visual-only, and audiovisual conditions. A student pianist was asked to play the excerpts in either a "deadpan," "projected," or "exaggerated" manner. Results showed that the visual-only ratings for all three performance manners matched those of the audiovisual ones more closely than did the audio-only ratings, suggesting that the visual condition provided an amount of information comparable to the audiovisual condition to the raters. Davidson concluded that visual component could be "more informative than sound in the perceiver's understanding of the performer's expressive intentions" (Davidson, 1993, p. 112).

Despite the fact that the expressive manners used in the two studies above were specified, artificial and thus unrealistic in an actual performance setting, Davidson's (1993) results provide valuable insight into the perception of expressive intention through the visual channel. The finding that the visual channel could be as informative as the audio channel in this regard remains a remarkable one, one that continues to inspire ongoing studies in performance evaluation.

Ohgushi (2006) carried out two studies modeled after Davidson's (1993, 1994, 2002) series of experiments in order to further investigate the visual channel and its ability to convey expressive intentions. In the first, he recruited music, psychology and fine arts students to view and rate an excerpt from Chopin's Nocturne Op. 9 No.2, under various audiovisual conditions. The excerpt was played by the same performer in three expressive manners, "deadpan", "projected", and

"exaggerated", the same ones used by Davidson in many of her studies (1993, 1994, 2002). Ohgushi (2006) found that even when observers were asked to rate the "sound expressiveness of performance" (paying attention to the sound only) as opposed to the "expressiveness of performance (paying attention to both sound and visuals) in the audiovisual condition, the visual information had a strong influence on the ratings, leading Ohgushi (2006) to conclude that visual information has much impact on sound perception in audiovisual performances. Another remarkable finding was that while both music and psychology students' understanding of "sound expressiveness of performance" were significantly influenced by the addition of the visual component, psychology students were unable to identify the expressive intention of the performer based on sound when contradictory visual information was added (e.g. when visuals of the "exaggerated" intention were synchronized to audio of the "projected" intention, or vice versa); whereas music students were still able to discern the expressive intention of the performer based on sound even in the presence of contradictory visual information. This suggests that different levels of expertise in music could affect the way observers rate musical expressiveness.

It is worth pointing out, however, that the visual component Ohgushi (2006) used for his experiment only included a view of the performer's upper body, without her hands. This curious exclusion was not explained by the researchers. This alone is a cause for further research, since the hands of a pianist are capable of a rich repertoire of movements and gestures that could convey important expressive information. It would be very interesting and worthwhile to investigate how adding the hands into the visual stimulus would affect the observers' ratings.

Shoda et al. (2007) conducted a study in which 91 observers were asked to rate two recordings of

two piano pieces by Rachmaninoff, one of them a performance of his Prelude Op. 32 No. 5, the other of his Etude Tableaux Op. 39 No. 1. For each performance, the observers were asked to rate 31 emotional impressions (e.g. "cheerful", "dreamy", "depressing", "heavy", etc. [1]) on a nine-point scale. The ratings were given under visual-only, audio-only, and audiovisual conditions. Results showed that the observers' comparatively weaker emotional impressions (termed "non-specific" by the researchers) under the audio only condition were significantly strengthened by the addition of visual information, resulting in higher emotional response ratings for the audiovisual condition. Furthermore, when the participants rated how much they liked the two pieces under each condition, the audiovisual ratings were closer to the visual-only ones than to the audio-only ones, suggesting that the audiovisual ratings were more affected by visual information. The researchers concluded that the body movements of the pianist significantly affected the listener's emotional impressions of the musical selections.

Important though its findings are, this study could perhaps benefit from a more diverse selection of repertoire. The two pieces used in it were both by Rachmaninoff, and even though they feature a contrasting character, both the Prelude and the Etude-Tableau are firmly Romantic in style, with the former being intensely lyrical and the latter being sweepingly virtuosic. As the researchers admitted, they might achieve different results if they had included other genres (Shoda et al., 2007). If they had included pieces from the Classical or Baroque era, for instance, the effect of the visual channel on the observers' impressions might have been rather different,

[1] The 31 emotional impressions were: cheerful, joyous, bright, exhilarated, humorous, gorgeous, quiet, leisurely, tranquil, tender, dreamy, graceful, quick, longing, majestic, lofty, robust, passionate, emphatic, solemn, exaggerated, depressing, doleful, whimsical, restless, dark, heavy, frustrating, deadpan, metrical, and sober (Shoda et al., 2007).

given that in that repertoire one would usually not see such openly virtuosic gestures as the ones in the Etude-Tableau op. 39 no. 1 that was chosen for the experiment. The effect of dramatic body movements on the perception of expression is especially relevant in this study, since all the 91 participants were relatively inexperienced, as the researchers pointed out that there were "no highly trained musicians among them" (Shoda et al., 2007). Previous research has shown that expert observers and raters could provide more informative results than amateur ones (Gibson, 1979; Scully, 1986), who are likely to be more reliant on purely visceral indicators of musical skillfulness, such as the presence of rapid and dynamic body movements, when judging the expressive merits of a piano performance.

The impact of the visual channel in piano performances has also been shown to be sufficient to overpower that of the audio channel. In a series of seven experiments, Tsay (2013) recruited both novice and expert participants to analyze six-second clips from past editions of music competitions that are considered among the most prestigious in the world, such as the International Tchaikovsky Competition, Van Cliburn International Competition, and Cleveland International Piano Competition. From a group of finalists, participants were asked to identify the prizewinners of the competitions based on clips that were either in audio-only, visual-only, and audiovisual mode. Within this series, experiments two and four, which only utilized audio-only and visual-only clips, even though 83.3% of novice and 96.3% of expert participants claimed that the audio-only would matter more than the visual-only condition for their evaluations, neither group was able to identify the winners better than chance under the former, with only 20.5% of expert participants being able to do so (Tsay, 2013). These poor performances by both novice and expert observers continued in experiments three and five, when

the audiovisual condition was added. Again, neither groups did better than picking the winners at random. This was in contrast to the visual-only condition in experiments two and four, under which both groups performed significantly better, with 46.4% and 46.6% of novices and experts, respectively, correctly identifying the winners of the competitions (Tsay, 2013). The researcher emphasized that although both novice and expert raters believed that they could better tell who the winners were with audio-only clips, the visual-only ones turned out to provide more useful information, allowing the raters to more accurately identify the winning performances. Tsay (2013) concluded that the influence of visual information on our perception of sound is significant, regardless of training and background. As such, she warned that visual information, despite being "deemed as peripheral in the domain of music", could actually be "overweighted" in a performance evaluation setting. This is undesirable, since extra-musical factors, rather than the musical performance itself, could end up being focused on in such contexts (Tsay, 2013).

Besides mainly piano performance, studies on other instruments such as the clarinet have also shown the impact of the visual component on creating emotional impressions and on the perception of expressive intentions. Vines, Krumhansl, Wanderley, Dalca, and Levitin (2005) asked 30 musically trained participants to rate audiovisual recordings of a Stravinsky piece for solo clarinet under "immobile", "standard" and "exaggerated" manners. These three expressive intentions are similar to the ones used by Davidson (1993, 1994, 2002) in her experiments. Observers were randomly divided into three treatment groups that rated the recordings under either the audio-only, visual-only, or audiovisual conditions. Each recording was rated on a five-point Likert scale, for how much they felt 19 emotions (e.g. "disgust", "happiness", etc.[2]), with 1

[2] The 19 emotions were: expressivity, intensity, movement, quality, surprise, interest, amusement, disgust, anxiety,

being “not at all” and 5 being “very much). Results showed that the “immobile” manner received significantly lower ratings than the other two manners overall, but especially so for the visual-only and audiovisual rating conditions. Ratings were very similar for all three expressive manners under the audio-only condition. The researchers concluded that the visual information in the other two conditions were responsible for the significantly lower ratings for the “immobile” manner, and that the visuals were “the primary channel” through which the emotional impact of the performances on the observers varied (Vines, Krumhansl, Wanderly, Dalca, & Levitin, 2005).

Not only do these results corroborate those of Davidson’s (1993) study, in that the visual channel was able to transmit accurate information about the performers’ expressive manner to the observers; they also reinforce those in the experiment by Shoda et al. (2007), where the observers’ emotional impressions of the performances were heightened by the addition of visual information.

Platz and Kopiez (2012) conducted a meta-analysis in an attempt to quantify the effect size of the visual component in music performance evaluation. They calculated the average effect size of the visual component based on a meta-analysis of 15 aggregated studies on audio-visual music perception, most of which included making the body movements visible to the observers, by deducting ratings for the audio-only condition from those for the audio-visual condition. The researchers focused on evaluations of liking, expressiveness, or overall quality of musical

---

anger, contempt, fear, contentedness, pleasantness, relief, happiness, familiarity, embarrassment, and sadness (Vines, Krumhansl, Wanderley, Dalca, & Levitin, 2005).

performances, and they found that the visual component has “an average medium effect size of 0.51 standard deviations – Cohen’s *d*; CI (0.42, 0.59)” (Platz & Kopiez, 2012, p. 75). This indicates that its effects are “visible to the naked eye” to an attentive observer (Cohen, 1988, p.26). Platz and Kopiez (2012) concluded that the visual component is not a “marginal phenomenon” in music perception but a highly important factor in aesthetic expression and the communication of meaning, not least due to the expressive information that body movements and gestures could convey to the audience.

Table 1

Summary of Designs and Key Findings, Audiovisual Rating Conditions Studies

Study	Design	Finding
Davidson (1993)	• Conducted study on piano and violin • Used three expressive intentions (“deadpan”, “projected”, “exaggerated”) • Used three audio-only, visual-only, and audiovisual conditions • Recruited only musicians as raters	• Visual information was more informative than the audio to raters as they tried to determine the expressive intention of the performers
Ohgushi (2006)	• Conducted study on piano • Used three expressive intentions (“deadpan”, “projected”, “exaggerated) as in Davidson (1993) • Recruited a mix of musicians and non-musicians as raters	• Visual information highly influenced sound perception in audiovisual performances, even when raters were asked to evaluate only the audio component
Shoda (2007)	• Conducted study on piano • Used different emotional impressions for rating • Used three audio-only, visual-only, and audiovisual conditions • Recruited a mix of musicians and non-musicians as raters	• Visual information greatly helped raters, especially non-musicians, to form an emotional impression on the music performed
Tsay (2013)	• Conducted study on piano • Used footage of major competitions for rating • Used three audio-only, visual-only, and audiovisual conditions • Recruited a mix of musicians and non-musicians as raters	• Visual information strongly helped both musician and non-musician raters to correctly identify the winners of each competition; neither group did better than chance in the audio-only condition
Vines et al. (2005)	• Conducted study on piano • Used intentions (“immobile”, “standard”, “exaggerated”) • Used three audio-only, visual-only, and audiovisual conditions • Recruited only musicians as raters	• Visual information drastically lowered ratings in the “immobile” intention, which did not drop much in the audio-only condition compared to the other two expressive intentions

Amount of Visual Information

Besides different audiovisual rating conditions, different amounts of visual information visible to the observers have also been found to affect ratings of musical expression.

Williamon (1999) recruited a cellist to perform J. S. Bach's Cello Suites I, II, and III under five different conditions. Three of them involved placing the performer behind a music stand, which partially obstructed her. The same cellist was also asked to perform the pieces from memory without the music stand, and to perform with the score, but with the music stand positioned outside of the audience's view. The researcher found that the added visual information of the cellist's body movements in the two unobstructed performances led to a significantly higher rating in the overall quality of communication and musicality compared to the obstructed performances on a scale from one to six, with six being highest. (Williamon, 1999). This result supports Davidson's (1994) finding that the enhanced visual information is a crucial factor that helps the audience understand the performer's expressive intentions. The effect of the music stand's presence in the study was dramatic likely because it obstructed the most direct part of the cellist's playing mechanism, "her movements (e.g. bowings)", the display of which was likely the crucial factor that led to a "heightened experience for audience members" in the unobstructed performance (Williamon, 1999). Furthermore, the researcher pointed out that in contrast to Davidson's (1994) study, the non-musician participants in his study were not actually more influenced by the mere presence of visual information, but instead were influenced most when visual information was enhanced by the body movements of the cellist being visible (Williamon, 1999). This suggests that a varying degree of information carried in the visual component may itself have a significant effect on ratings, which warrants research that compares the effects of differing amounts of visual information, such as the partial visuals versus full visuals of a

musical performance.

In a study on the amount of visual information provided to observers and its effect on their perception of expressive intention, Dahl and Friberg (2007) recruited a professional marimba player to perform in silent video clips that were intended to display four different expressions: anger, happiness, sadness, and fear. For each expressive intention, the video clips were edited into four viewing conditions, with varying parts of the performer being visible to the viewers: full image, only without the hands, only the torso, and only the head. The viewers were then asked to rate the emotional content in each video clip on a seven-point scale for the four expressions, from "nothing" to "very much" (Dahl & Friberg, 2007). The researchers found that three of the four emotions (anger, happiness, sadness) were successfully conveyed by the marimba player under practically all viewing conditions. Dahl and Friberg (2007) also observed that a "staircase" relation between the viewing conditions, where the achievement values of the participants become higher as more of the performer's body was shown, only existed for the anger expressive intention. They concluded that the body part(s) shown in each video clip only had a slight influence on the identification of expressive intention. (Dahl & Friberg, 2007). This finding, that enhanced visual information only has a slight influence on the perception of expressive intentions, directly contradicts that of Davidson (1994) and Williamon (1999). However, it must be noted that the researchers of this study had heavily altered the video clips that were used, so that the facial expressions of the performer were removed, and that the whole image was converted into pure black and white tones, rendering it impossible for viewers to see anything other than the performer's body movements. The main advantage of this design is that it prevents other factors such as the performer's sex, attractiveness and attire from affecting the

ratings and forced raters to focus on rating the musical aspects of the performance (in this case the emotional content). Furthermore, by isolating the body movements from the scenes, it also allows raters to evaluate the performances specifically based on the differing amounts of visual information provided. The disadvantage is that it is a stark contrast to the two aforementioned studies in its highly unrealistic nature, one that is far-removed from the conditions of a typical performance.

Table 2

Summary of Designs and Key Findings, Amount of Visual Information Studies

Study	Design	Finding
Williamon (1999)	• Conducted study on cello • Used visually obstructed (music stand in front of performer) and unobstructed (no music stand visible) performances • Recruited a mix of musicians and non-musicians as raters	• Visually unobstructed performances were rated significantly higher in overall quality of communication and musicality than obstructed ones; obstructed performances have lower ratings in general
Dahl and Friberg (2007)	• Conducted study on marimba • Used four expressive intentions (“anger”, “happiness”, “sadness”, and “fear”) for rating • Used four visual conditions (“full image”, “without the hands”, “only the torso”, and ”only the head”) • Recruited a mix of musicians and non-musicians as raters	• All emotional intentions were accurately identified regardless of the amount of visual information available; “anger” was more accurately identified when more visual information was available in the rating conditions

Body Movements and Expression

Various researchers have also investigated the characteristics of body movements shown to observers and how they influence ratings of musical expression. This has been done by varying the size and intensity of the gestures shown, connecting movements used during different structural points of the music to the perception of expressive intentions, and using conflicting audiovisual information to explore how body movement characteristics affect the perception of expressive intentions.

In a follow-up study to her aforementioned landmark experiments (Davidson, 1993), Davidson (1994) investigated the relationship between the movement characteristics of a pianist and his expressive intent. With the use of two-dimensional (vertical and horizontal) movement tracking, which recorded movement direction and size on these axes, the size of the performer's physical gestures was found to be linked to his expressive intentions. The stronger the intention, the larger the movement sizes were. For example, physical gestures used in the "exaggerated" performance manner were larger than those used in the "deadpan" manner. In a second study, Davidson (1994) explored the extent to which various regions of a pianist's body were indicative of his expressive intentions. The researcher discovered that the upper torso and head area (including facial expressions) was sufficient for the observers to make a perceptual judgment as to whether the pianist was playing in "exaggerated", "projected" or "deadpan" manner. Davidson (1994) also pointed out that the observers were most accurate in their judgment when both the upper torso/head region and the hands of the pianist were visible. The latter finding in part suggests that the hands, the most direct part of a pianist's playing mechanism, could convey important expressive information through their movements. The researcher did not detail in the study to

what extent the hands influenced the judgment of the raters, however.

Davidson (2002) carried out two further studies that were closely related to her 1993 and 1994 studies. In the first one, the researcher investigated whether or not information of a pianist's expressive intentions from his physical movements were delivered to the observers in a continuous stream, or if it was available only at specific points of a performance. Davidson (2002) found a significant connection between the identifiable expressive gestures of the pianist and the musical structure of the piece that he was performing. Based on the responses of the participants, who were asked to identify the expressive intention of the pianist at different points throughout his performance, the researcher suggested that the expressive intentions of the performer were most often identified at important structural points of the music, such as during a rest or at a cadence point at the end of a phrase. According to Davidson (2002), it is during key moments like these where the hands of the performer would usually make informative movements – such as an extra care given to the gentle lifting of the wrist when moving the hand up and away from the keyboard.

In the second study, Davidson (2002) once again focused on various regions of a pianist's body, and found that the movements of the pianist's upper torso alone transmitted information that allowed observers to much more easily determine his expressive intention compared to the information transmitted by the movements of his hands. The researcher suggested that this was due to the pianist's centre of gravity being situated at his contact point with the piano stool, which caused body movements around this point to be especially apparent to the observer, such as the swinging and swaying motions of the upper torso that Davidson (2002) claimed to

distinguish the “exaggerated” and “projected” manners from the “deadpan” manner. This result contradicts those from her 1994 study, where Davidson (1994) found that although the upper torso/head region contained enough information to reveal a pianist’s expressive intention, more accurate judgments were made when the hands were also visible to the observers.

The Davidson studies above prompt further research to be conducted, specifically to compare the impact of the movements of a pianist’s different body regions on the perception of musical expression. The hands and arms, in particular, should be singled out for study, in order to investigate whether or not they would result in significantly different ratings for musical expressiveness when viewed alone, as opposed to being viewed in combination with other body regions such as the head and torso.

Juchniewicz (2008) investigated the effect of the size and intensity of a pianist’s physical gestures on listeners’ perception of performance quality. Undergraduate and graduate music students were recruited to rate three video excerpts of Chopin’s Etude in E major op. 10 no. 3, one with “no movement” (only essential movements to play the piece allowed) in the performer, one with “head and facial movement,” and one with “full body movement”. In all of the excerpts, the performers’ entire body was visible. Each of the three video clips was synchronized to the same soundtrack. For each excerpt, participants were asked to rate the performances in terms of four expressive criteria: phrasing, dynamics, rubato, and overall grade. The researcher found that even though the audio component of all three performances remained the same, the more areas of the body shown moving to the observers, the higher the observers’ total ratings were. This finding corroborates that of Williamon (1999) in that more body

movements in the performer visible to the raters leads to higher ratings in expression and overall musicality. The primary difference is that Juchniewicz (2008) allowed the performer's whole body to be seen, regardless of amount of movement displayed, whereas Williamon (1999) varied the amount of visible movements by physically blocking a part of the performer with a music stand. As Juchniewicz (2008) also found a significant difference in ratings between the expressive criteria across the three excerpts, he also concluded that the performer's physical gestures could have an impact on criteria-specific ratings in performance evaluation.

Here, it is worth pointing out that an important concern must be taken into account when performers are required to play in a certain way. Juchniewicz's (2008) study, for example, like the one by Davidson (1993), limits the expressive freedom of the performer by providing specific performance instructions. In Davidson's (1993) experiment, the performers were asked to play in a specific manner (either "deadpan", "projected", or "exaggerated"), while Juchniewicz (2008) specified the amount of movement ("no movement", "head and facial movement", "full body movement") the pianist should include in his performance. In the latter study, due to the required degrees of body movement, certain passages of the piece played would likely be either unnaturally limited in expression, such as when head and facial movement is prohibited during an *ff* climax, for instance, or exaggerated, such as when full body movement is involved in a simple scalar run. With that in mind, the concern is that the sizes and types of physical movements most viewers would find expressive, or even appropriate, would most likely vary greatly depending on the repertoire, as one would normally expect more full body movements in a virtuosic Chopin Etude than in a graceful Mozart Sonata. A study carried out by Huang and Krumhansl (2011) demonstrates this difference clearly, as it found that ratings for

expressive criteria such as Dynamics, Phrasing, and Melody/Harmony were highest for a Chopin performance when it featured "exaggerated" (as opposed to "natural" or "minimal") stage behavior; whereas it was highest for a Copland performance when it featured "natural" stage behavior. For the above reasons, this study by Juchniewicz (2008), like the one by Shoda et al. (2007), could perhaps benefit from a wider range of musical selection. When pieces of diverse styles and characters are included and are subject to the same experiment, the results would more accurately reflect how the sizes and types of physical movements affect the perception of musical expression. This is the primary reason for the current study's inclusion of a varied list of repertoire (see Table 6, p. 52).

As in research on different audiovisual rating conditions, studies on body movements have also shown that performers' gestures in clarinet performance, like in piano performance, have a strong influence on the perception of musical expression. Krahé, Hahn and Whitney (2013) explored the effect of the visual and audio components of a clarinet performance on viewer ratings by matching soundtracks with video recordings of congruent and incongruent movements. A solo arrangement of Beethoven's Bagatelle for piano in two movements (*Lustig*, "happy", and *Traurig*, "sad") was used in the experiment. The performer recorded *Lustig* and *Traurig* with congruent body movements (sad movements for the sad piece, happy movements for the happy piece), then the researchers created incongruent versions of each recording by splitting the video and soundtracks in editing. In these versions, happy movements were matched with the sad piece, and sad movements were matched with the happy piece. In the video clips, the performer's breaths were carefully matched to the soundtrack, and repeating visual material was also matched to the length of the soundtrack. Participants, who were all non-musicians, then

commented on what the performances made them feel (felt emotion), what the performances were expressive of (perceived emotion), and how strong the evoked emotions were. The results showed that both visual and audio information were important in determining the felt and perceived emotion of the raters, and that neither visual nor audio information alone could singlehandedly sway the ratings. Although the researchers found that "sad" music was perceived to be more so when paired with "sad" body movements than when paired with "happy" body movements, they suggested that visual information alone may be "insufficient to override the musical content altogether." The effect of the visual component was strongest in cases where the musical content was more emotionally vague (i.e. vague according to the participants in their aforementioned written comments), as it helped the viewers rate "sad" music as sadder and "happy" music as happier. This finding supports that of Shoda et al. (2007), which also suggests that the visual channel plays an important role in strengthening the emotional responses of the observers, although it may not be able to override audio information in the emotional perception. In addition, the outcome of the study could have been quite different if the researchers had recruited musicians as raters instead, or at least some participants with musical training. It would be interesting to see if their experience would enable them to perceive the conflicting audiovisual information differently than non-musicians.

Furthermore, Vines, Wanderley, Krumhansl, Nuzzo, and Levitin (2004) found that removing visual information from a solo clarinet performance decreased the experience of expressive tension in a section where note density, pitch and volume were all lower relative to the rest of the piece. The visual information was a full camera shot of the performer, making the body movements and gestures in the performance clearly visible. Using sliders on a Peavey 1600X

MIDI controller, participants in the study were asked to indicate their experience of expressive tension throughout the performance of Stravinsky's second piece for solo clarinet. Results of a functional data analysis showed that during the middle section of the piece where the dynamics dropped from mezzo forte to pianissimo, and where the note density decreased from 16th and 32nd notes to 8th notes accompanied by a fall in pitch, the values of the Tension Linear Model plummeted significantly (to below 0.1) when the participants were exposed only to the audio, compared to a far greater value (above 0.6) when they are exposed only to the visuals of the performance. This suggests that in addition to communicating expressive intentions (Davidson, 1993; Williamon, 1999) and transmitting emotion (Krahé, Hahn & Whitney, 2013; Shoda et al., 2007), the visual channel also plays a role in conveying expressive tension, a vital element of musical expression. This finding agrees with Davidson's (1993) set of experiments, which found that the perception of expressive intensity in both violin and piano performances was highly impacted by the visual channel. In this particular case, the body movements of the performer were likely an important factor that contributed to the heightened perception of performance intensity in the observers.

Table 3

Summary of Designs and Key Findings, Body Movement and Expression Studies

Study	Design	Finding
Davidson (1994)	• Conducted study on piano • Used three expressive intentions ("deadpan", "projected", "exaggerated") • Investigated relation between movement size, area, and expressive intention • Recruited only musicians as raters	• The stronger the expressive intention, the larger the movement sizes; performer's upper torso and head carried enough information for most raters to identify the expressive intention
Juchniewicz (2008)	• Conducted study on piano • Investigated relationship between movement degree, intensity and performance ratings • Used three degrees of movement ("no movement", "head and facial movement", and "full body movement") for rating • Recruited only musicians as raters	• The greater degree of body movement shown and the higher its intensity, the higher the ratings for musical expression were
Krahé, Hahn and Whitney (2010)	• Conducted study on clarinet • Mixed "happy" and "sad" visual and audio recordings for rating • Recruited only non-musicians as raters	• Visual and audio information were equally influential in forming a perceived and felt emotional impression
Vines, Wanderley, Krumhansl, Nuzzo, and Levitin (2004)	• Conducted study on clarinet • Compared ratings for an audio-only vs. visual-only performance • Recruited only musicians as raters	• Visual information was strongly linked to the degree to which raters experienced musical tension; perception of tension fell drastically when visuals were removed from performance

Rater and Performer Proficiency

There has also been research that shows that musically trained and musically inexperienced raters are affected by the audiovisual stimuli differently in a performance evaluation context.

Wapnick, Ryan, Lacaille, and Darrow (2004) found that compared to pianists, raters who were non-musicians were much more positively influenced by the addition of the visual component in an experiment that involved evaluating piano performances given at the 2001 Van Cliburn Competition. The pianists' ratings were not found to be significantly different across the audiovisual and audio-only conditions. Huang and Krumhansl (2011) also reported a similar finding, as they pointed out that musician raters in their experiment were able to perceive differences in stage behavior, from minimal to exaggerated, in both the audiovisual and audio-only conditions, whereas non-musicians were able to do so only in the audiovisual condition, with the help of the visual cues. Both studies reinforce the findings of Ohgushi (2006), who found that music students did not rate "sound expressiveness of performances" significantly differently when contradicting visual information was added to audio recordings; as opposed to psychology students, who were unable to grasp the expressive manner intended by the pianist once contradicting visuals were introduced.

In a previously mentioned study, however, Tsay (2013) found that both novices and experts were mostly able to identify the winners of some of the world's top music competitions, based on short, video-only clips. Just under half of each group of participants were able to identify the winners and do so at better than chance. When presented with sound-only versions of the same clips, however, both novices and experts did no better than picking the winners at random, with barely a fifth of experts being able to successfully identify the winners (Tsay, 2013). The

researcher then concluded that visual information significantly affected both novice and expert raters in their evaluations of piano performances, even though novices may rely more on visual cues alone in audiovisual recordings, due to their lack of musical training. These results contradict those of other studies mentioned above by suggesting that the visual component influences both experienced and inexperienced observers similarly by providing observers with information useful for determining performance quality. However, it is worth keeping in mind that the study asked participants to identify the winners among the performers in the recordings, and not rate the performances directly. Many factors could have been influencing the observers in guessing the winners, including stage behavior, attire, attractiveness, and other external factors. It is possible that the novice and expert observers would have responded to the performances more differently under the different audiovisual conditions, if they were asked to rate each performance for its level of expression by criteria such as phrasing, dynamics, rubato, and overall quality. As Tsay (2013) herself pointed out in her article, novice raters were more likely to rely solely on visual cues when evaluating the performances, which may have led to their highly mixed ratings under the audio-only condition.

Performer proficiency has also been demonstrated to influence the way performance evaluations are given. The attractiveness bias in judges, for instance, was found to be significant when the performers being rated were novice and college-level musicians, whereas it was found to be insignificant when the performers were top-level players, such as candidates at the Eleventh Van Cliburn International Piano Competition (Ryan, Wapnick, Lacaille, & Darrow, 2006). In addition, the researchers found that the undergraduate judges were positively influenced by high attractiveness only on two out of six measures (“adherence to style” and “overall impression”),

while graduate and faculty judges were not positively influenced on any measure. Overall, the main effect of attractiveness on the ratings was found to be insignificant (Ryan, Wapnick, Lacaille, & Darrow, 2006). The researchers concluded by pointing out that these results were especially interesting when compared to those of previous studies, which showed a significant attractiveness bias in judges when the performers were amateurs, such as college-level singers (Wapnick, Darrow, Kovacs, & Dalrymple, 1997), and violinists (Wapnick, Mazza, & Darrow, 1998). This suggests that performer proficiency has a direct influence on the existence of an attractiveness bias in judges.

Finally, a curious interaction found in numerous previous studies in performance evaluation is that for expert performances, usually longer test excerpts that were as long as 10 minutes yielded higher ratings than did shorter excerpts that were as short as less than a minute, while the reverse is true for amateur performances (Bergee, 1993; 1997; Fiske, 1975, 1979; Hewitt, 2007; Hewitt & Smith, 2004; Thompson, Williamon, & Valentine, 2007; Wapnick et al., 2005). This suggests that a parallel could perhaps be drawn between the duration of excerpts and the degree of visual information shown. For expert performers, for example, a higher degree of visual information in their performances, such as under the full audiovisual condition, may improve their ratings, while for amateur performances it may not do the same, and perhaps even worsen their ratings. McClaren (1985), for example, found that non-musicians rated solo marimba performances more highly under the audiovisual condition for the performers who had also been rated highly by experienced judges. Ratings on scales such as “Ineffective-Effective”, and “Sensitive-Insensitive” tended to be more positive only for the more proficient performers, who had been rated more highly for both “visual effect” and “aural effect” by a panel of experts (McClaren,

1985). In the cases mentioned above, the total amount of audiovisual information and the expertise of the performers highly influenced the results. The current study therefore utilized both expert and amateur performances across the audio-only, audio with limited visuals and audio with full visuals conditions, in order to investigate this potential interaction.

Table 4

Summary of Designs and Key Findings, Rater and Performer Proficiency Studies

Study	Design	Finding
Wapnick, Ryan, Lacaille, and Darrow (2004)	• Conducted study on piano • Compared amateur with expert ratings • Used real competition footage • Used audio-only and audiovisual conditions for rating	• Amateur ratings were significantly more highly influenced by the addition of visual information, compared to musician ratings
Huang and Krumhansl (2011)	• Conducted study on piano • Compared amateur with expert ratings • Used three expressive intentions for rating (“minimal”, “natural”, and “exaggerated”) • Used audio-only and audiovisual conditions for rating	• Experts were able to accurately identify the expressive intentions by audio-only, whereas amateurs were only able to accurately do so with the help of visuals in the audiovisual condition
Tsay (2013)	• Conducted study on piano • Compared amateur with expert ratings • Used real competition footage • Used audio-only and visual-only conditions for rating	• Neither experts and amateurs were able to accurately identify the true winners of the competitions by audio only; but both were able to do so by visual only
Ryan, Wapnick, Lacaille, and Darrow (2006)	• Conducted study on piano • Compared amateur with expert ratings • Compared attractiveness bias for amateur vs. expert performers, as well as among amateur vs. expert raters • Used real competition footage	• Attractiveness bias was significant when performers were amateurs and insignificant when they were professionals; it was also insignificant when the raters were professionals, and only sometimes significant when they were amateurs

Implications for the Current Study

With the findings of previous research as overviewed above in mind, the current study sought to further explore the effect of the visual component of piano performances on performance evaluation, while addressing some of the outstanding issues of existing studies, the most important of which being the hands and arms having never been singled out for investigation. In addition, having learnt from the designs of previous research, a wide variety of repertoire, performers and performer proficiency were featured in order to increase the ecological validity of the results. To reiterate, the research questions of the current study were:

- *Whether or not the rating conditions would affect ratings for musical expression*;
- *Whether or not there would be an interaction effect between rating condition and these expressive criteria;*
- *Whether or not there would be an interaction effect between rating condition and performance;* and
- *Whether or not there would be an interaction effect between rating condition and performer proficiency.*

CHAPTER II

Methodology and Procedures

In order to address the four research questions, a quantitative approach similar to the ones used for existing studies mentioned in the Literature Review (p. 18-48) was employed: An experiment was conducted to collect data for statistical analysis, the results of which were then examined.

Participants

A group of pianists ($N = 60$) who had at least 10 years of piano training, and who had either performed or taught regularly in the past three years were recruited from university music departments in the Greater Montreal area (see Table 5, p.50). The majority of them ($n = 37$) were also at least one of the following at the time of the experiment: Associate of The Royal Conservatory (ARCT) Piano Diploma holder, Bachelor, Master, or Doctor of Music student in piano performance. The other participants ($n = 23$) were all students at McGill university who have or had been taking private piano lessons and performing and/or teaching in their leisure. Only such highly trained pianists were recruited, as previous research suggested that experienced observers in perceptual experiments could produce more informed results than naïve ones (Scully, 1986; Gibson, 1979), results that would be more reflective of piano performance evaluations given in a competition or audition context. Participants were recruited using mouth-to-mouth and social media advertising, as well as by a digital poster circulated on the electronic student mailing lists at McGill University (see Appendix C, p. 91). The participants were asked to rate 10 piano performances of diverse styles and proficiency (see Table 6, p.51) under one of the following conditions: audio-only, limited audiovisual, and full audiovisual. All participants completed the experiment independently, as they scheduled appointments individually with the

researcher.

Table 5

Demographic Information of Experiment Participants (*N* = 60)

Sex	***n***	***%***
Female	37	67.33
Male	23	38.33
Age		
18-20	30	50
21-25	21	35
26-30	5	8.33
30+	4	6.67
Piano training received		
10-15 years	39	65
16-20 years	14	23.33
20+ years	7	11.67
Current level of expertise		
ARCT	12	20
B. Mus	13	21.67
M. Mus	10	16.67
D. Mus	2	3.33

Note. Data compiled from participant questionnaire (see Appendix B, p. 90). Participants who were not accounted for in the "Current level of expertise" section had achieved none of the four credentials listed, but each had acquired at least 10 years of private instruction.

Preparation of Materials

Ten audiovisual recordings, each of a different piano piece, were extracted from publicly available videos on www.youtube.com. These videos featured five amateur and five professional performers (see Table 6).

Table 6

List of Featured Musical Selections, Performers and Their Level of Proficiency

Performance	Performer and piece	Expertise
A	Hall, *Invention No. 12 in A Major, Bach*	Amateur
B	Lisitsa, *Presto Agitato, Piano Sonata op. 27 no. 2*, Beethoven	Professional
C	Sing, *Andante Grazioso, Piano Sonata K. 331*, Mozart	Amateur
D	Barton, *Etude Op. 10 No. 1,* Chopin	Amateur
E	Horowitz, *Carmen-Fantasie*, Bizet-Horowitz	Professional
F	Moiseiwitsch, *Tannhäuser Overture*, Wagner-Liszt	Professional
G	Varisco, *Allegro Barbaro*, Bartok	Amateur
H	Astanova, *Transcendental Etude No. 10,* Liszt	Professional
I	Kim, *Etude Op. 40 No. 2,* Kapustin	Professional
J	Stürtzer, *La fille aux cheveux de lin*, Debussy	Amateur

Note. All audio and visual clips were extracted from www.youtube.com. Performances were played in alphabetical order (A-J) for all participants. URLs are available in the References list, p. xx.

Proficiency was defined by whether or not each performer's occupation was a concert pianist. In order to determine this, each pianist was contacted via private communication through his or her YouTube account, except for deceased or well-known performers such as Horowitz, Moiseiwitsch, and Lisitsa. Each performance featured a different pianist, and there were seven men and three women in total.

Unlike some of the studies mentioned in the Literature Review chapter, such as Juchniewicz (2008) and Shoda et al. (2007), the musical selections of this study featured stylistically diverse repertoire from the Baroque era to the early 20th century. The reason for this was so that a wide variety of expression and physical gestures required to play the pieces would be displayed. Another reason for the diversity of musical selection was to reflect better an actual audition or competition setting, in which the judges would most likely encounter a wide range of contrasting repertoire. Finally, it is the main objective of this study to investigate how different rating conditions impact the evaluation of piano performances in general, and not only performances of repertoire from a specific stylistic period.

Unlike those used in perceptual experiments such as Davidson's (1993), the performances chosen for this study were not produced in a specific manner ("deadpan", "projected" or "exaggerated") – the performers had not been instructed to play in any particular way. This design results in a listening environment closer to a typical live performance for the observers and is therefore more ecologically valid.

For the sake of consistency, only videos featuring comparable camera angles, as well as audio

and video quality were selected. Remastered versions of historical recordings with improved A/V quality were used whenever possible, such as the Horowitz performance (Sissco, 2006). The full audiovisual version of the videos featured the performer's upper body, hands and arms without any editing. Each video was post-processed in Apple iMovie so that they were the same in length (1 minute and a half), and that they were all scaled to the same high-definition resolution (see Figure 1). The limited audiovisual version of the videos were created by inserting black boxes that covered up almost the entire frame, except for the keyboard and the performers' hands and arms. After editing, only a view of the performers' hands and arms was observable. The videos in the limited audiovisual condition were then converted from colour to black and white, so as to further prevent external factors, such as the colour of the performer's attire, from influencing the performance ratings (see Figure 2, p. 54). The same was not done for the full audiovisual condition, as its purpose was to give observers a view that more closely resembles that of an actual performance, without drawing special attention to the hands and arms at the same time.

Figure 1. Example view of the full audiovisual condition.

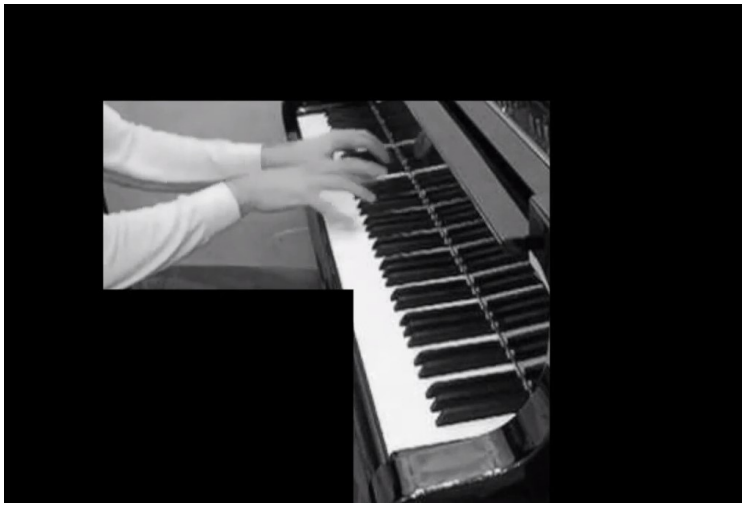

Figure 2. Example view of the limited audiovisual condition (same frame).

In addition, one of the videos in the limited audiovisual condition, Kapustin's *Etude Op. 40 No. 2* performed by Kim (Kim, 2014) needed cropping to fill the frame, so that raters would have the same view of all ten performers' hands and arms. No other video required such compensative editing. Audio-only condition videos were created using black boxes to completely obscure the frame. Next, the ten recordings used for each rating condition were combined into a 17:30 minute long video for each of the three rating conditions. 15 seconds of silence were inserted between each performance, so that participants would have sufficient time to evaluate them. Rating forms and a questionnaire was also prepared for each participant (see Procedure section, p. 55). All participants used the same Mac Pro workstation equipped with reference-grade Sennheiser HD 650 headphones for the experiment.

Procedures

The study took place at the Perceptual Testing Lab at CIRMMT (Centre of Interdisciplinary

Research in Music, Media and Technology), McGill University. The setting allowed each participant's experience to be meticulously controlled, so as to ensure uniformity across each experimental group.

The 60 participants were randomly divided into three groups of twenty, with each group rating the 10 performances under one of the three conditions: audio-only, limited audiovisual, and full audiovisual. Each participant only gave ratings under one condition, in order to prevent participants from rating from memory of how they had previously rated the performances. Participants were scheduled for the experiment individually, and were not aware of the identity of any other participant. A random process was used to determine which of the three rating conditions each participant would partake in: this process took place once for every three participants. The first of every three participants in was asked to choose a number between 1 and 3, with each number representing a rating condition. The second participant chose between the two remaining numbers, and the last participant had no choice but to take the only rating condition left. This random process ensured that there were exactly 20 participants for each rating condition, and that one condition would not be selected much more frequently than the others.

At the beginning of the experiment, before any recordings were played to each participant, a short questionnaire on the musical background and demographic information of the participants was completed by the participant (see Table 1 and Appendix B, p. 50 and p. 90). Once beginning to evaluate the performances, each participant was asked to think of himself or herself as a judge for a professional piano audition or competition, and to rate the 10 performances on

four expressive criteria: phrasing, dynamics, rubato, and overall quality. The 10 performances were played in the same order for all participants, regardless of rating condition. The four expressive criteria were rated on a scale from 0 to 100, which is a system commonly used in music schools for adjudication according to Wapnick and Darrow (2012), who have conducted numerous studies on performance evaluation. Before the 10 recordings were played, two sample recordings were played for each participant, in order for them to have an idea of what the actual experiment recordings would sound, or look like. The two sample recordings featured pianists whose proficiency fell in the middle of the spectrum with regards to the 10 pianists in the actual experiment (Babic, 2014; Zhou, 2011). This effectively established a moderate standard by which participants could judge the 10 recordings, which would have been especially useful for rating the first recording, performance A. In the actual experiment, participants were allowed to pause the video during the 15 seconds of silence between every performance, if they needed more time to evaluate a recording. The entire rating process for each participant took approximately 20 minutes. At the end of the experiment, each participant was compensated by either entering a draw for a $20 gift card or being paid with $10 cash, as approved by the McGill University Research Ethics Board.

Data Analysis

In order to address each of the four research questions of this study, the results were analyzed to explore the effects of the four independent variables: rating conditions, expressive criteria, performances, and performer proficiency on the dependent variable, the ratings. To do this, a one-way ANOVA or an independent-sample t-test was carried out for each of these four independent variables. Then, three two-way ANOVAs were conducted to investigate the

potential interaction effects between rating condition and each of the other three independent variables. This was to determine if the visual channel influenced the ratings across their levels. The two-way ANOVAs were only used for the interaction effect, since rating condition is compared against each of the other three independent variables. It would be redundant if two-way ANOVAs were used to calculate the main effects of each independent variable, as data for rating condition would be unnecessarily repeated. The software IBM SPSS Statistics 21 was used to carry out all of the above procedures.

CHAPTER III

Results

Overview

In accordance to the research questions of the study, the Results chapter is organized in three subsections: rating condition, expressive criterion, and performance and performer proficiency. The last two independent variables are combined because of their close relation to one another, and also because performer proficiency only has two levels (amateur and professional).

Rating Condition

In order to compare ratings given under the three rating conditions, the means and standard deviations of the total scores for each condition were calculated (see Table 7, p. 59). The ratings for the three conditions were as follows: limited audiovisuals ($M = 76.96$, $SD = 19.06$), audio-only ($M = 75.83$, $SD = 18.92$), and full audiovisuals, ($M = 73.40$, $SD = 21.00$). When comparing these data, significance ($p > .05$) found in *Levene's F* test showed that the homogeneity of variance assumption was not met, therefore *Welch's F* test was used. A one-way ANOVA revealed that there was a significant effect of rating condition on the rating scores at the $p < .05$ level, *Welch's* $F(2,1594.64) = 6.45$, $p = .002$. Post-hoc comparisons using the Tukey HSD test showed that mean ratings for the full audiovisual conditions were significantly lower than both those for the limited audiovisual and audio-only conditions. No significance in mean ratings between the audio-only and limited audiovisual conditions was found (see Table 7, p. 59).

Table 7

Post-hoc Multiple Comparisons of Mean Ratings between the Rating Conditions

Rating Conditions	Rating Conditions	Mean Difference	*p*
AO	LAV	-1.13	.48
	FAV	2.43	**.04**
LAV	AO	1.13	.48
	FAV	3.56	**.00**
FAV	AO	-2.43	**.04**
	LAV	-3.56	**.00**

Note: AO = Audio-only, LAV = Limited Audiovisual, FAV = Full Audiovisual. Significant *p* values ($p < .05$) are in bold.

Expressive Criterion

The means and standard deviations of the total scores of each expressive criterion was then calculated. The ratings for the four expressive criteria were as follows: Phrasing ($M = 75.33$, $SD = 19.78$), Dynamics ($M = 74.80$, $SD = 20.44$), Rubato, ($M = 75.14$, $SD = 20.150$, and Overall Quality ($M = 76.30$, $SD = 18.54$). However, a one-way ANOVA indicated that there was no significant effect of expressive criteria on the rating scores, $F(3, 2396) = 0.64$, $p > .05$.

A two-way ANOVA that compared the rating conditions with the expressive criteria showed that the interaction effect was insignificant, $F(6, 2388) = 0.05$, $p > .05$, which suggests that the different rating conditions did not significantly impact how the expressive criteria were judged.

Performance and Performer Proficiency

In terms of total scores (i.e. including all rating conditions and expressive criteria), Performance A was found to have the lowest mean rating, while Performance E had the highest. Except for Performance F, all professional performances received a total mean rating above 80, while no amateur performance received a total mean rating above 75 (see Table 8).

Table 8

Mean Total Ratings and Standard Deviations for each Performance

Performance	***N***	***M***	***SD***	***Proficiency***
A, Bach Invention	240	50.23	24.25	Amateur
B, Beethoven Sonata	240	82.22	14.23	Professional
C, Mozart Sonata	240	65.68	18.12	Amateur
D, Chopin Etude	240	74.98	17.21	Amateur
E, Horowitz-Bizet Fantasie	240	87.38	11.23	Professional
F, Wagner-Liszt Overture	240	78.48	16.20	Professional
G, Bartok Allegro Barbaro	240	74.79	17.28	Amateur
H, Liszt Etude	240	84.44	13.34	Professional
I, Kapustin Etude	240	84.99	11.91	Professional
J, Debussy Prelude	240	70.75	18.93	Amateur

A one-way ANOVA showed that there was an effect of performance on the ratings scores, *Welch's F*(9, 971.47) = 86.85, $p < .05$. Post-hoc comparisons again using the Tukey HSD test showed many differences in mean ratings between the performances (see Table 9, p. 62). For example, the two performances with the lowest mean ratings, Performance A (M = 50.4, SD = 24.20) and Performance C (M = 65.68, SD = 18.08), both had significantly lower mean ratings than the other performances. The most pronounced difference was found between Performance

A (M = 50.4, SD = 24.20) and Performance E (M = 87.38, SD = 11.23).

Table 9

Post-hoc Multiple Comparisons of Total Mean Ratings between the Performances

	Bach	**Beethoven**	**Mozart**	**Chopin**	**Horowitz-Bizet**	**Wagner-Liszt**	**Bartok**	**Liszt**	**Kapustin**	**Debussy**
Bach										
Beethoven	*M*= -31.99									
Mozart	*M*= -15.45	*M*= 16.54								
Chopin	*M*= -24.75	*M*= 7.24	*M*= -9.30							
Horowitz-Bizet	*M*= **-37.14**	*M*= -5.15	*M*= -21.69	*M*= -12.39						
Wagner-Liszt	*M*= -28.25	*NS*	*M*= -12.80	*NS*	*M*= 8.89					
Bartok	*M*= -24.56	*M*= 7.43	*M*= -9.11	*NS*	*M*= 12.58	*NS*				
Liszt	*M*= **-34.20**	*NS*	*M*= -18.76	*M*= -9.46	*NS*	*M*= -5.96	*M*= -9.65			
Kapustin	*M*= **-34.76**	*NS*	*M*= -19.31	*M*= -10.01	*NS*	*M*= -6.51	*M*= -10.20	*NS*		
Debussy	*M*= -20.52	*M*= 11.47	*M*= -5.08	*NS*	*M*= 16.62	*M*= 7.73	*M*= 4.04	*M*= 13.68	*M*= 14.24	

Note: The table is read vertically, rather than horizontally (e.g. the Bach performance is 31.99 points lower than the Beethoven in total mean rating, hence the "-" sign. Only significant p values ($p < .05$) are reported. *NS* = not significant. The 3 largest differences in total mean rating are in bold.

The results of a two-way ANOVA that compared the rating conditions with the performances showed that there was a significant interaction effect, $F(18, 2370) = 4.65$, $p < .05$, suggesting that effect of rating condition was greater for some performances than others.

For example, the rating conditions had a bigger impact on the ratings for Performance A than for Performance C, as the former's mean rating has a ten-point difference between limited and full audiovisual conditions, while the latter's mean rating barely changed across the three conditions (see Table 10).

Table 10

Mean Ratings and Standard Deviations for Each Performance under Each Condition, *N = 80*

Performance	**Rating Condition**	***M***	***SD***
A, Bach	AO	51.85	24.00
	LAV	54.38	25.05
	FAV	44.48	22.84
B, Beethoven	AO	84.94	10.25
	LAV	81.00	17.29
	FAV	80.73	14.06
C, Mozart	AO	64.36	17.03
	LAV	66.93	19.39
	FAV	65.75	18.01
D, Chopin	AO	78.98	16.22
	LAV	76.20	17.28
	FAV	69.76	17.00
E, Horowitz-Bizet	AO	82.78	14.96
	LAV	89.60	7.16
	FAV	89.75	8.64

Table 10 continued

Performance	Rating Condition	*M*	*SD*
F, Wagner-Liszt	AO	71.99	17.49
	LAV	80.94	17.59
	FAV	82.51	10.64
G, Bartok	AO	79.53	12.84
	LAV	75.94	13.59
	FAV	68.91	22.24
H, Liszt	AO	84.13	15.04
	LAV	83.91	12.09
	FAV	85.28	12.84
I, Kapustin	AO	84.83	12.53
	LAV	87.33	11.93
	FAV	82.83	10.93
J, Debussy	AO	74.89	16.34
	LAV	73.36	16.02
	FAV	64.01	22.15

Note: Professional performances are in bold.

A particularly notable trend was observed in the full audiovisual ratings. The condition had a stronger *negative* affect on amateur performances, such as Performances A, G, and J, than on professional ones, such as Performances B and I. The former three performances all had almost 10-point drops in mean rating from the limited to full audiovisual conditions; whereas the latter two both had less than 5-point drops. In fact, under the full audiovisual condition, two professional performances, E and F, actually benefitted from an increase in mean ratings from

the limited to full audiovisual conditions (though only very slightly for Performance E); whereas this was not the case for any amateur performance.

In addition, an independent-sample t-test revealed that there was a significant difference in ratings between amateur ($M = 67.29$, $SD = 21.38$) and professional ($M = 83.50$, $SD = 13.81$) performances, $t(2398) = -22.07$, $p < .05$. A two-way ANOVA was then conducted to investigate the interaction effect between rating condition and performer proficiency, which was found to be significant, $F(2, 2394) = 15.67$, $p < .05$, which suggests that the effect of rating condition on performer proficiency was significant.

CHAPTER IV
Discussion

Overview of Results

The results of the study indicated that the visual channel had an impact on the ratings of musical expression: the main effect of rating condition on the ratings was found to be significant. In particular, the ratings between the audio-only and full audiovisual, as well as limited and full audiovisual conditions were found to be significantly different. However, this was not true for the ratings between the audio-only and limited audiovisual conditions. Interestingly, the difference in mean rating between the two audiovisual conditions was larger than that between the audio-only and either of the audiovisual conditions. Therefore, there was no gradual increase in ratings from the audio-only through the limited audiovisual, to the full audiovisual conditions. This finding was in contrast to previous studies, which showed that ratings generally increased as more performer gestures and body movements become visible to the raters. A more detailed discussion of each research question of this study is included below.

Rating Condition

The main research question, *whether or not the rating conditions would affect the ratings for musical expression* was answered when ratings between the audio-only and full audiovisual conditions were found to be significantly different (see Table 7, p. 59). This finding, that the visual component of piano performances significantly affects their evaluations, is in line with previous research (Davidson, 1993, 1994, 2002; Juchniewicz, 2008; Ohgushi, 2006; Shoda et al., 2007; Tsay, 2013). Remarkably, ratings between the two audiovisual conditions were also found

to be significantly different. No difference for ratings between the audio-only and limited audiovisual conditions was found, however. This suggests that viewing the hand and arm movements of the performers in addition to listening to the performances was probably not enough to significantly impact the raters' perceived expressiveness. An explanation for this could be that the musical excerpts did not feature enough key structural points such as climaxes or cadence points, which have been found to influence performance factors such as the amount of body movement and expressive timing used (Davidson, 2002; Thompson & Luck, 2012).

In the full audiovisual condition, when hand and arm movements were visible in conjunction with other body movements such as those in the torso, face and head, they presumably created a stronger, albeit often negative impression that contributed to the significantly different ratings. For instance, in some amateur performances, rigid facial expressions could confirm the lack of emotional involvement that the raters might have already sensed from the stiff hand movements. Two other important factors that could also have led to the significantly lower ratings in the full audiovisual condition are performance setting and performer attire. It was only in the full audiovisual condition that a camera shot of the full scene was used, and in most of the amateur performances, the pianists were seen playing at a non-concert venue such as in a practice room, or at home studio. Due to these casual settings, most of them also wore attire that would usually be considered out of place in a performance setting. It was possible that the unprofessional performance settings and performer attire had a negative impression on the raters, as demonstrated by Wapnick, Mazza & Darrow (1998; 2000); which in this case also helps explain why the full audiovisual condition had a stronger negative impact on amateur performances compared to professional ones, both in the number of performances affected and the degree of

mean total rating decreased (see Table 10, pp. 63-64). Finally, an explanation for the ratings not having decreased or increased in the same direction across the three rating conditions could be the fact that a different set of participants were used for each condition. This design is contrary to some studies that found that ratings gradually increased as more visual information became available to the raters, such as the one by Dahl & Friberg (2007), which employed the same group of 20 participants across the rating conditions. Compared to that, the design of the current study has both advantages and disadvantages, which will be elaborated in the Limitations of the Study section of this chapter (p. 72-74).

Expressive Criteria

It was unsurprising that ratings between the four expressive criteria did not vary significantly in general, given that participants under all three rating conditions evaluated all of them. Interestingly, however, since no interaction effect between rating condition and expressive criteria was found, the results suggest that the rating conditions did not have an impact on the expressive criteria. For criteria such as dynamics and rubato, this was perhaps more easily understandable, as visual information was not expected to greatly influence their perception. For the phrasing criterion, however, visual gestures that ranged from the expressive hand lifting at phrase-ends by professionals, to the comparatively stiff finger work by amateurs, were visible to viewers in the two audiovisual conditions. It was therefore curious to find that these gestures alone did not enhance or detract from the perception of phrasing quality enough to result in significantly different ratings. The same applied to ratings for overall quality, where participants evaluated the performances based on the performers' general level of musical expression. Here, the visual component did not significantly impact the ratings either, as participants evaluated the

criterion remarkably similarly across all three rating conditions. Although these results at first seem to be in contrast with those of Juchniewicz (2008), which showed that the amount of body movement visible had a significant impact on the ratings of phrasing, dynamics, and rubato, it must be noted that said study (Juchniewicz, 2008) differs from the current study in that it did not utilize different rating conditions. The degree of body movement visible to the raters depended instead on which body parts ("no movement", "head and facial movement", and "full body movement") the performer was asked to use in his recordings, rather than having stimuli being presented under different audiovisual conditions.

Performance and Performer Proficiency

As predicted, the ten performances varied tremendously in ratings, as a result of the diverse range of repertoire, performers, and performer proficiency featured in them. Regarding the musical selections used, it is worth noting that the two highest rated performances, Horowitz-Bizet's *Carmen Fantasie* and Kapustin's *Etude op. 40 no. 2*, are both virtuosic pieces that demand very rapid tempi; while the two lowest rated performances, Bach's *Invention in A major* and Mozart's *Andante Grazioso* from *Sonata in A major*, are technically much less challenging works in slower tempi. This suggests that the performance difficulty of the music and the rapid hand movements in the former pair may have created a strong, positive impression on the raters that contributed to the high scores. The same pieces would have to be played by different performers in order to determine whether the ratings had been more influenced by the musical selection or the performers themselves.

It is also worth pointing out that since the lowest rated performance, Bach's *Invention in A*

major, was the first performance to be played for the raters, it was possible that an order effect influenced its ratings. Specifically, it was possible that the raters' expectations for the first performance depended on whether they had a positive or negative impression on the two sample recordings that were played to them before the start of the experiment. It must be noted, however, that an effort was taken to control for this order effect on the first performance: the two sample recordings featured pieces of moderate difficulty and performers of moderate proficiency, in comparison to the ten experimental performances. Thus, the low ratings were more likely a result of the pianist, Hall's comparatively lower performance ability.

The interaction effect found between rating condition and performance suggests that the performers in the recordings may also have influenced the ratings. For example, the two highest rated performances, Horowitz-Bizet's *Carmen Fantasie* was played by the legendary Horowitz himself, while Wagner-Liszt's *Tannhäuser Overture* was played by another highly revered pianist, Moiseiwitsch. This might have been a major factor that led these two performances to be also the only two that received a slightly higher mean rating under the full audiovisual condition than the limited audiovisual. It was possible that the raters under the former condition became biased, as they were the only group who could see the performers' faces. Although some raters may have been able to identify some of the pianists from their hands alone in the limited audiovisual condition, the possibility is much lower compared to the full audiovisual condition.

An attractiveness bias was highly unlikely to have influenced the full audiovisual ratings, at least not positively so, as the two pianists in the two highest rated performances were both at an advanced age (above 60 years old) at the time of the recordings and were not conventionally

attractive. All other performances, some of which featured more conventionally attractive and young pianists, received lower ratings in the full audiovisual condition compared to the limited audiovisual, which is in contrast to the findings by Wapnick, Darrow, Kovacs, & Dalrymple (1997), as well as Wapnick, Mazza, and Darrow (1998, 2000), which showed that performers who were considered attractive received significantly higher performance ratings than those who were not. However, it must be noted that the design of the current study deviates from that of the three studies above in two important ways: (a) in the current study, participants did not rate the performers on attractiveness, therefore whether or not the performers were attractive is only the personal opinion of this author; (b) the group of participants who rated the performers' attractiveness in the three studies did not rate the musical performances either, as those were rated by a different group of participants.

As expected, the professional performances received higher ratings than the amateur ones. The interaction effect found between rating condition and performer proficiency was particularly interesting– the full audiovisual condition negatively impacted amateur performances more than it did professional performances. This result supports the findings of previous research, which demonstrated that more proficient performers received higher ratings when more visual information became available in the recordings, which was not the case for the less proficient ones (McClaren, 1985; Wapnick et al., 2005). It suggests that compared to the audio-only condition, viewers were more critical of the amateur performers' musical expression when visual information such as their appearances and full body movements were visible; whereas the viewers were more impressed by the professional performers' expression when provided with the same visual information. Another reason that professional performances received higher scores

in the full audiovisual condition could be the fact that many professional performances featured recognizable pianists. In addition to Horowitz, the identities of other relatively well-known pianists that were made explicit in the full audiovisual condition could also have had biased the ratings in either direction, depending on the rater's opinion on the particular pianist(s). In contrast, the amateur performances by their nature featured largely unknown pianists; thus the same bias most likely could not have been a factor that influenced their ratings.

Limitations of the Study

Three limitations became apparent throughout the course of this study. Firstly, each of the three rating conditions (audio-only, limited audiovisual, and full audiovisual) was rated by a different set of participants in this experiment. While this design prevented the participants' memories of the performances from influencing the ratings across the conditions, the resulting difference in ratings across the conditions would also have to be partly attributed to the fact that different raters were used. If the same group of participants had been employed for all of the rating conditions, the difference in ratings across the conditions would have been even more likely to be a result of different treatment. However, there are disadvantages to the latter design as well. If the same participants had been used, the dates for the experiments under each rating condition would have had to be separated by a considerable time frame, so as to prevent the participants' memories from influencing the ratings. The fact that this study utilized performers of varying proficiency would have exacerbated the issue of memory, as performances on either end of the proficiency spectrum would be very likely to have a strong impression on the participants, making it especially difficult for them to forget how they had previously evaluated those performances.

Secondly, another factor that could have influenced the experimental results was the fact that some of the performers featured in the recordings used in the study were highly well known concert pianists, such as Vladimir Horowitz. As discussed in the previous section, the preference of the participants might have created a bias on their ratings if they recognized the performers, which was possible for the Limited Audiovisual group and highly likely for the full audiovisual group. If the participants had presumptions about a certain pianist and his or her playing, they might have affected the ratings. The use of existing recordings made by famous pianists, however, allowed a very wide range of performer proficiency to be featured in the experiment, as it would otherwise have been difficult to find a pianist who could record a virtuosic performance similar to Horowitz's, for instance.

Lastly, the study could have benefited from employing participants with even higher levels of expertise, such as professional critics. This study specifically employed highly experienced, domain-specialists as judges in order to investigate whether the visual channel had an impact in performance evaluation. The reason for this was to develop a better understanding of how developing pianists are judged in professional auditions and competitions. Therefore, it would only have been more appropriate if top experts with experience serving on the adjudicating panels of numerous competitions, for example, were recruited for the experiments. While that had unfortunately proved to be impossible for a study of this scale, an active effort was made in the recruitment process to find the highest qualified participants possible. This included contacting the piano departments at the Schulich School of Music and University of Montreal Faculty of Music to advertise the study to the most advanced students in the above schools, such

as those at the D.Mus and M.Mus levels. This study greatly benefited from the input of those students, whose extensive experience in not only giving but also critiquing and evaluating performances at an extremely high level, was instrumental in upholding the rigorousness of the findings.

Implications for Music Education

The results of this study reinforce the discoveries of recent researchers who have conducted visual research from a music education perspective. These researchers ranged from the pioneer Davidson (1993), to the performance evaluation experts Wapnick (1997, 1998, 2000) and McPherson and Schubert (2004), to the professional selection and advancement specialist Tsay (2013). These results have clearly demonstrated the visual component's great influence on performance evaluation. This finding, and the fact that it has been frequently replicated in music psychology and education studies, forms a powerful implication for music education: that piano pedagogues and students alike must pay greater attention to the visual component of piano performance. Awareness could, and should be raised through discussion in the teaching studio, method books, and further research conducted on the topic. Discussion of piano technique and expression should no longer be confined to biomechanics and interpretation, as it largely has been throughout the history of piano pedagogy, from Clementi (1803) to Berman (2002).

Although the current study found that seeing the performer's hand and arm gestures in isolation did not have a significant impact on performance evaluation, it discovered that viewing their full body movements in conjunction with their attire and stage behavior did. As seen in the Literature Review chapter, the growing body of research that has employed a broad variety of designs

across wide range of contexts has repeatedly verified this latter finding. It is becoming increasingly clear that the visual component contributes to the communication between performer and audience, such as the "projection of expressive, interpretative, and structural features of the composition performed", one of the most typically used performance assessment criteria by music institutions (McPherson & Schubert, 2004, p. 64). Therefore, it is only logical to suggest that developing pianists would do well to be more particular and mindful about their overall visual presentation on stage, whether at a professional audition or competition.

Suggestions for Future Research

To further investigate the topic of this study, researchers could conduct experiments that target a more specific part of the visuals in piano performance evaluation. In particular, researchers should do more to focus on the most direct part of the playing mechanism. One way to do this would be to run a study that only displays the hand and arms of each performer, which would be seen playing the same excerpts using movements of various qualities (such as smooth vs. angular, heavy vs. light, etc.) Perhaps an even more interesting topic that has not been explored in previous studies, to this author's knowledge, is the amount of contact that the hands and fingers make with the keyboard. This topic is especially interesting because the visuals of the touch a pianist has could potentially affect the perception of phrasing quality, such as when a performer is seen using the damper pedal to sustain a chord rather than using his or her fingers, for instance.

Furthermore, an experimental design that could be valuable is to hold an experiment in a performance setting, such as in a recital hall, with the performer dressed professionally. This

would allow observers to evaluate the performances in a context that is as close to that of an actual concert or audition as possible, though it would not afford the researchers the meticulous control that a testing lab would. Nonetheless, it would be valuable to learn about the degree to which a pianist's visual gestures would impact his or her ratings when the observers are viewing the performer from a farther, more realistic distance, as opposed to the telephoto and close-up shots that are often used in experimental recordings.

Lastly, a qualitative analysis of the experimental results, such as a discussion about specific aspect(s) of the performers' gestures that were rated more highly or lowly, as well as possible explanations for them, could be highly interesting and informative to pedagogues, and could be included in future studies on the topic.

Conclusion

The purpose of this study was to investigate the impact of the visual component of piano performances on their evaluations. This was done by comparing ratings of the same performances that were given under three conditions. The performance setting and an entire shot of the performer was featured in the full audiovisual condition. The hands and arms were singled out for investigation under the limited audiovisual condition, due to the assumption that they were the most important parts of the pianist's playing mechanism. Although it was found that viewing their gestures alone in the limited audiovisual condition did not result in significantly different ratings of expression compared to the Audi-only condition, it was found that viewing these gestures in addition to the rest of the performers' body movements and overall appearance, did. Significant difference in ratings was also found between the two audiovisual conditions. A particularly remarkable finding was that a larger difference existed between the limited and full audiovisual mean rating than between either of the audiovisual and the audio-only ratings. This suggests that the overall visual presentation of a piano performance is a crucial factor in performance evaluation, perhaps more so than hand and arm movements alone. Furthermore, no interaction was found between rating condition and expressive criteria, while one between rating condition and performance as well as rating condition and performer proficiency were found. In summary, the results of this study reinforce the finding of previous research that the visual component of piano performances has a significant influence on performance evaluation. The methodological design that produced these results features a combination of elements new to current research on the topic (isolating of the hands and arms for study, in addition to utilizing a wide selection of repertoire, in addition to employing both expert and amateur performances).

The current study is therefore a small, though not insignificant contribution to piano performance evaluation research. Further research is needed in order to investigate whether more specific hand and arm gestures would have an impact on ratings of expression, which is especially important since they are produced by the primary part of the pianist's playing mechanism. Finally, in order to raise the awareness of the visual component of piano performances, pedagogues should discuss the visual component of piano performances more with both their students and colleagues.

REFERENCES

Astanova, L. (2010, July 5). *Lola Astanova plays transcendental etude no. 10 in f-minor by Liszt* [Video file]. Retrieved from https://www.youtube.com/watch?v=iwVp7S59fO0

Babic, U. (2014, December 14). *Chopin: etude op. 10 no. 5 (black keys) Urska Babic* [Video file]. Retrieved from https://www.youtube.com/watch?v=F3ytMEJ3qjM

Barton, P. (2013, February 14). *Chopin etude op. 10 no. 1 – p. Barton Feurich 218 grand piano* [Video file]. Retrieved from https://www.youtube.com/watch?v=L8w_0Hr13J0

Bella, S.D., & Palmer, C. (2011). Rate effects on timing, key velocity, and finger kinematics in piano performance. *PLoS One, 6*(6), e25018. doi:10.1371/journal.pone.0020518

Bergee, M. J. (1993). A comparison of faculty, peer, and self-evaluation of applied brass jury performances. *Journal of Research in Music Education, 41,* 19-27. doi:10.2307/3345476

Bergee, M. J. (1997). Relationships among faculty, peer, and self-evaluations of applied performances. *Journal of Research in Music Education, 45,* 605-612. doi: 10.2307/3345425

Bergeron, V., & Lopes, D. M. (2009). Hearing and seeing musical expression. *Philosophy and Phenomenology Research, 78*(1), 1-16. doi:10.1111/j.1933-1592.2008.00230.x

Berman, B. (2002). *Notes from the pianist's bench.* New Haven, CT: Yale University Press.

Campbell, R., Dodd, B., & Burnham, D. (1998). *Hearing by eye: Advances in the psychology of speechreading and audio-visual speech*. Hove: Psychology Press.

Castellano, G., Villalba, S. D., & Camurri, A. (2007). Recognizing human emotions from body movement and gesture dynamics. *Affective Computing and Intelligent Interaction, 4738*, 71-82. doi:10.1007/978-3-540-74889-2_7

Clementi, M. (1803). *Introduction to the art of playing the pianoforte.* (Reprint: N.Y.: Da Capo

Press 1973 ed.). London: Clementi, Banger, Hyde, Colland and Davis.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Lawrence Erlbaum.

Cramer, J. (1812). *Instructions for the pianoforte in which the first rudiments of music are clearly explained and the principal rules on the art of fingering illustrated with numerous and appropriate exercises.* London: S. Chappell, n.d.

Dahl, S., & Friberg, A. (2007). Visual perception of expressiveness in musician's body movements. *Music Perception: An Interdisciplinary Journal, 24*(5), 433-454. doi:10.1.1.118.4880

Davidson, J. W. (1993). Visual perception of performance manner in the movement of solo musicians. *Psychology of Music, 21,* 103-113. doi:10.1177/030573569302100201

Davidson, J. W. (1994). What type of information is conveyed in the body movements of solo musicians? *Journal of Human Movement Studies, 6,* 279-301.

Davidson, J. W. (2002). Understanding the expressive movements of a solo pianist. *Musikpsychologie, 16*, 9-31.

Davidson, J. W. (2007). Qualitative insights into the use of expressive body movement in solo piano performance: A case study. *Psychology of Music, 35*(3), 381-401. doi:10.1177/0305735607072652

Delalande, F. (1988). "La gestique de Gould: éléments pour une sémiologie du geste musical." In G. Guertin, ed. *Glenn Gould, Pluriel*. Louise Courteau Éditrice Inc., 83-111.

Dittrich, W. H., Troscianko, T., Lea, S. E., & Morgan, D. (1996). Perception of emotion from dynamic point-light displays represented in dance. *Perception*, *6*, 727-738. doi:10.1068/p250727

Dvorsky, M. (2011, February 28). *Benno Moiseiwitsch plays Wagner-Liszt 'Overture to Tannhäuser'* [Video file]. Retrieved from https://www.youtube.com/watch?v=XKDYla5C5cA

Finnäs, L. (2001). Presenting music live, audio-visually or aurally – does it affect listeners' experiences differently? *British Journal of Music Education, 18*(1), 55-78. http://dx.doi.org/10.1017/S0265051701000146

Fiske, H. E. (1975). Judge-group differences in the rating of secondary school trumpet performances. *Journal of Research in Music Education, 23,* 186-196. doi:10.2307/3344643

Fiske, H. E. (1977). Relationship of selected factors in trumpet performance adjudication reliability. *Journal of Research in Music Education, 25,* 256-263. doi:10.2307/3345266

Gabrielsson, A. (1999). Studying emotional expression in music performance. *Bulletin for the Council of Research in Music Education, 141*, 47-53. Retrieved from: http://www.jstor.org/stable/40318983

Gabrielsson, A., & Juslin, P. N. (1996). Emotional expression in music performance: Between the performer's intention and the listener's experience. *Psychology of Music, 24*, 68-91. doi:10.1177/0305735696241007

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Gillespie, R. (1997). Ratings of violin and viola vibrato performance in audio-only and audiovisual presentations. *Journal of Research in Music Education, 45*(2)*,* 212-220. doi:10.2307/3345581

Hall, C. (2008, September 21). *Bach: Invention 12 in a major (older version) | Cory Hall, pianist-composer* [Video file]. Retrieved from

https://www.youtube.com/watch?v=HYvz8QAxUUw

Hewitt, M. P. (2007). Influence of primary performance instrument and education level on music performance evaluation. *Journal of Research in Music Education, 55,* 18-30. doi:10.1177/002242940705500103

Hewitt, M. P., & Smith, B. P. (2004). The influence of teaching-career level and primary performance instrument on the assessment of music performance. *Journal of Research in Music Education, 52,* 314-327. doi:10.1177/002242940405200404

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. doi:10.1177/1029864911414172

Hummel, J. N. (1828). *Ausfürlich theoretische-practische Anweisung zum Piano-forte Spiel*. Vienna.

Johnson, C. M. (1996). Musicians' and nonmusicians' assessment of perceived rubato in musical performance. *Journal of Research in Music Education, 44,* 84-96. doi:10.2307/3345415

Juchniewicz, J. (2008). The influence of physical movement on the perception of musical performance. *Psychology of Music, 36*(4), 417-427. doi:10.1177/0305735607086046

Juslin, P. N. (2005). From mimesis to catharsis: Expression, perception, and induction of emotion in music. In D. J. Hargreaves, R. MacDonald & D. Miell (Eds.), *Musical Communication*, 85-115. New York, NY: Oxford University Press.

Kim, S. (2014, September 18). *Sukyeon Kim plays Kapustin etude op. 40 no. 2 "Reverie"* [Video file]. Retrieved from https://www.youtube.com/watch?v=Cii-KfNadAc

Kochevitsky, G. (1967). *The art of piano playing: A scientific approach.* Seacaucus, NJ: Sunny-Brichard.

Krahé, C., Hahn, U., & Whitney, K. (2013). Is seeing (musical) believing? The eye versus the ear in emotional responses to music. *Psychology of Music*, *0*(0), 1-9. doi: 10.1177/0305735613498920

Lehmann, A.C., & Davidson, J. W. (2002). Taking an acquired skills perspective on music performance. In R. Colwell & C. Richardson (Eds), *The New Handbook of Research on Music Teaching and Learning* (pp. 542-560). Oxford: Oxford University Press.

Lhevinne, J., & Lhevinne, R. (1972). *Basic principles of pianoforte playing.* New York: Dover Publications, Inc.

Lisitsa, V. (2010, January 18). *Beethoven "Moonlight" Sonata op 27 #2 mov 3 Valentina Lisitsa* [Video file]. Retrieved from https://www.youtube.com/watch?v=zucBfXpCA6s

Luck, G., Toiviainen, P., & Thompson, M. R. (2010). Perception of expression in conductors' gestures: A continuous response study. *Music perception: An Interdisciplinary Journal, 28*(1), 47-57. doi:10.1525/mp.2010.28.1.47

MacFarlane, L., Kulka, I., & Pollick, F. E. (2004). The representation of affect revealed by butoh dance. *Psychologia*, *47*, 96-103. doi:10.2117/psysoc.2004.96

McClaren, C. A. (1985). *The influence of visual attributes of solo marimbists on perceived qualitative response of listeners* (Unpublished doctoral dissertation). University of Oklahoma, Norman, OK. Retrieved from: https://shareok.org/handle/11244/5357

McPherson, G.E., & Schubert, E. (2004). *Musical excellence: Strategies and techniques to enhance performance.* A. Williamon (Ed.). Oxford: Oxford University Press.

Morrison, S. J., Price, H. E., Smedley, E. M., & Meals, C. D. (2014). Conductor gestures influence evaluations of ensemble performance. *Frontiers in Psychology, 5*(806). doi: 10.3389/fpsyg.2014.00806

Munhall, K., Jones, J., Callan, D., Kurate, T., & Vatikiotis-Bateson, E. (2004). Visual prosody and speech intelligibility. *Psychological Science, 15*, 133-134. doi:10.1111/j.0963-7214.2004.01502010.x

Nakamura, T. (1987). The communication of dynamics between musicians and listeners through musical performance. *Perception & Psychophysics, 41,* 525-533. doi:10.3758/BF03210487

Ohgushi, K. (2006). Interaction between auditory and visual information in conveyance of players' intentions. *Acoustical Science and Technology, 27*(6), 336-339. doi:10.1250/ast.27.336

Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 331-346. doi:10.1037/0096-1523.15.2.331

Paterson, H. M., Pollick, F. E., & Sanford, A. J. (2001). The role of velocity in affect discrimination. In J.D.Moore & K. Stenning (Eds.), *Proceedings of the Twenty-third Annual Conference of the Cognitive Science Society*, Edinburgh (pp. 756-761). Mahwah, N.J.: Lawrence Erlbaum Associates. Retrieved from: http://conferences.inf.ed.ac.uk/cogsci2001/pdf-files/0756.pdf

Platz, F., & Kopiez, R. (2012). When the eye listens: A meta-analysis of how audio-visual presentation enhances the presentation of music performance. *Music Perception: An Interdisciplinary Journal, 30*(1), 71-83. doi:10.1525/mp.2012.30.1.71

Pollick, F. E., Paterson, H. M., Bruderlin, A., & Sanford, A. J. (2001). Perceiving affect from arm movement. *Cognition*, *82*, B51-B61. doi:10.1016/S0010-0277(01)00147-0

Ryan, C., Wapnick, J., Lacaille, N., & Darrow, A.-A. (2006). The effects of various physical

characteristics of high-level performers on adjudicators' performance ratings. *Psychology of Music, 34*(4), 559-572. doi: 10.1177/0305735606068106

Sandor, G. (1995). *On piano playing: Motion, sound, and expression.* New York: Cengage Learning.

Schutz, M. (2008). Seeing music? What musicians need to know about vision. *Empirical Musicology Review, 3*(3), 83-108. Retrieved from: https://kb.osu.edu/dspace/bitstream/handle/1811/34098/EMR000053a_revised.pdf

Scully, D. M. (1986). Visual perception of technical execution and aesthetic quality in biological motion. *Human Movement Science, 5,* 185-206. doi:10.1016/0167-9457(86)90024-2

Shoda, H., Nakamura, T., Draguna, M. R., Kawase, S., Katahira, K., & Yasuda, S. Effects of a pianist's body movements on listeners' impressions. (2007). Proceedings from ICoMCS '07: *The Inaugural International Conference on Music Communication Science*. Sydney, Australia: University of Western Sydney. Retrieved from: http://marcs.uws.edu.au/links/ICoMusic/Full_Paper_PDF/Shoda_Nakamura_Draguna_Kawase_Katahira_Yasuda.pdf

Sing, T. (2013, April 16). *Wolfgang Amadeus Mozart: Piano sonata no. 11 in A major, k.331 (1/3): I. Andante grazioso* [Video file]. Retrieved from https://www.youtube.com/watch?v=k_iDHHLgtAg

Sissco. (2006, August 21). *Horowitz plays Carmen Fantasie* [Video file]. Retrieved from https://www.youtube.com/watch?v=Qnla_5zrHAE

Sörgjerd, M. (2000). *Auditory and visual recognition of emotional expression in performances of music.* Unpublished master's thesis, Uppsala University, Uppsala, Sweden.

Stürtzer, M. (2014, December 21). *Claude Debussy – La fille aux cheveux de lin* [Video link].

Retrieved from https://www.youtube.com/watch?v=NVHM19Qb6JM

Thompson, S., Williamon, A., & Valentine, E. (2007). Time-dependent characteristics of performance evaluation. *Music Perception, 25,* 13-29. doi:10.1525/mp.2007.25.1.13

Thompson, W. F., Graham, P., & Russo, F. A. (2005). Seeing music performance: Visual influences on perception and experience. *Semiotica, 156*, 203-227. Retrieved from: http://eprints.qut.edu.au/26430/1/26430.pdf

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. doi:10.1177/1029864911423457

Tsay, C-J. (2013). Sight over sound in the judgment of musical performance. *Proceedings of the National Academy of Sciences, 110*(36), 14580-14585. doi:10.1073/pnas.1221454110

Varisco, E. (2008, May 23). *Enrico Varisco – Bartok – Allegro Barbaro* [Video file]. Retrieved from https://www.youtube.com/watch?v=fLpW-yDrGoo

Vines, B.W., Wanderley, M.M., Krumhansl, C.L., Nuzzo, R.L., & Levitin, D. J. (2004). Performance gestures of musicians: What structural and emotional information do they convey? *Lecture Notes in Computer Science, 2915,* 468-478. doi:10.1007/978-3-540-24598-8_43

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., Dalca, I., & Levitin, D. J., (2005). Dimensions of emotion in expressive musical performance. *Annals of the New York Academy of Sciences, 1060*, 462-466. doi:10.1196/annals.1360.052

Walk, R. D., & Homan, C. P. (1984). Emotion and dance in dynamic light displays. *Bulletin of Psychonomic Society*, *22*, 437-440. doi:10.3758/BF03333870

Wanderley, M. M., Vines, B. W., Middleton, M., 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. doi:10.1080/09298210500124208

Wapnick, J., & Darrow, A.-A. (2012). Sectional versus intact evaluations of four versions of a Chopin etude. *Journal of Research in Music Education, 60*(4), 462-474. doi:10.1177/0022429412465318

Wapnick, J., Darrow, A.-A., Kovacs, J., & Dalrymple, L. (1997). Effects of physical attractiveness on evaluation of vocal performance. *Journal of Research in Music Education, 45*(3), 470-479. doi:10.2307/3345540

Wapnick, J., Mazza, J. K., & Darrow, A.-A. (1998). Effects of performer attractiveness, stage behavior, and dress on violin performance evaluation. *Journal of Research in Music Education, 46*(4), 510-521. doi:10.2307/3345347

Wapnick, J., Mazza, J. K., & Darrow, A.-A. (2000). Effects of performers' attractiveness, stage behavior, and dress on evaluation of children's piano performances. *Journal of Research in Music Education, 48*(4), 323-335. doi:10.2307/3345367

Wapnick, J., Ryan, C., Campbell, L., Deek, P., Lemire, R., & Darrow, A.-A. (2005). Effects of excerpt tempo and duration on musicians' ratings of high-level piano performances. *Journal of Research in Music Education, 53,* 162-176. doi:10.1177/002242940505300206

Wapnick, J., Ryan, C., Lacaille, N., & Darrow, A.-A. (2004). Effects of selected variables on musicians' ratings of high-level performances. *International Journal of Music Education, 22,* 7-20. doi:10.1177/0255761404042371

Williamon, A. (1999). The value of performing from memory. *Psychology of Music, 27*, 84-95. doi:10.1177/0305735699271008

Zhou, Y. (2011, December 21). *Rachmaninoff Prelude in G minor op. 23 no. 5* [Video file]. Retrieved from https://www.youtube.com/watch?v=u9JmOKGVliI

Zijl, G. W., & Luck, G. (2012). Moved through music: The effect of experienced emotions on performers' movement characteristics. *Psychology of Music, 41*(2), 175-197. doi:10.1177/0305735612458334

APPENDICES

Appendix A Informed Consent Form

Project title: A Comparison of Piano Performance Evaluations Given under Audio-only, Limited and Full Audiovisual Conditions, MA in Music Education thesis

Principal Investigator: Kelvin Chan, MA in Music Education, Department of Music Research, Schulich School of Music, McGill University, kelvin.m.chan@mail.mcgill.ca

Supervisor: Professor Isabelle Cossette, isabelle.cossette1@mcgill.ca

Date: December 17, 2014

You are invited to participate in the study described below.

Description of the study:

This study investigates piano performance evaluations done under three rating conditions: audio-only, limited audiovisuals, and full audiovisuals. The participants will be divided into three groups, and each group will view and/or listen to and rate ten piano recordings under one of the three conditions.

The results of the study will be published in a Music Education MA thesis, and possibly in other publications such as journal articles, or presentations at academic conferences.

Reason for the invitation:

You are invited to evaluate these performances because of your presumed familiarity with the piano repertoire and experience in critical listening, having had at least 10 years of training on the instrument and being actively involved in teaching and/or performing.

Procedures:

The experiment will take place at the Perceptual Testing Lab at CIRMMT (Centre for Interdisciplinary research in Music, Media and Technology), McGill University, where you will view and/or listen to and rate ten piano recordings.

After viewing and/or listening to each recording, you will be asked to rate each of them out of a score from 0 (as terrible as it gets) to 100 (as good as it gets), in regards to these expressive criteria: phrasing, dynamics, rubato, and overall performance quality. There are 15 seconds of silence between each video for you to write down the ratings, but you are welcome to take as little or as much time as you would like. The ten recordings will be of ten different pieces of various styles. Please try your best to ignore the differences in sound and/or video quality among the recordings, and focus on the performance quality alone. The duration of the experiment will be approximately 20 minutes.

Confidentiality of the study:

Your identity will be kept confidential throughout this study. You will never be identified by name whenever the results of the study are published. The data from the experiment, like the participants' personal information, (e.g. e-mail address, telephone numbers) will be securely stored on digitally-encrypted USB drives and a password-protected personal computer owned by the researcher. Only the researcher himself will have access to these data.

Potential benefits:

There are no foreseeable benefits for participating in this study. Your participation is valuable and very much appreciated!

Potential harms and risks:

There are NO foreseeable harms and risks for the participants in this study.

Conflict of interests:

Some potential participants recruited for this study, including yourself, may be either friends or acquaintances of the researcher. However, your participation is completely voluntary, and you are under no obligation to participate. Please do not feel pressured to do so. Your decision not to participate will not incur any negative consequences on the relationship between you and the researcher.

Right to withdraw:

You have the rights to withdraw from this study at any point, without having to provide any reason(s), and without incurring any negative consequences. You are under no pressure to participate at any time during this study.

Compensation:

You will be given $10 cash for your time and for your willingness to participate in this experiment. This compensation will only be given to you if you complete the experiment.

Consent:

Please sign the below only if you agree to participate in this study.

Name, printed: Name: signed:

E-mail address: Telephone: Date:

If you have any questions or concerns regarding your rights or welfare as a participant in this research study, please contact the McGill Ethics Officer at 514-398-6831 or lynda.mcneil@mcgill.ca.

Appendix B

Questionnaire for Experimental Participants

- Full name: _____ Age:_____ Sex: (M / F / Other)
- How many years have you been playing the piano? _____
- What level have you achieved (RCM Grade 4, ARCT, B.Mus, M.Mus, etc.)?_____
- Do you play instruments other than the piano? (Yes / No) If so, # of years of training received? _____years
- Are you currently involved in piano performance (do you regularly perform?): (Yes / No)
    - For how long have you been involved? _____
    - On average, how many performances have you given per year in the past 3 years? _____
- Are you currently involved in piano teaching (do you regularly teach?): (Yes / No)
    - For how long have you been involved? _____
    - On average, how many students have you taught per year in the past 3 years? _____
- On average, how many hours of practice did you do per week in the past year? _____

THANK YOU!

Appendix C

Study Advertisement

Are you currently studying or teaching piano performance?

Have you had at least 10 years of piano training?

If your answer to the above questions is **YES**,

you are invited to participate in a piano performance evaluation study!

you would be rating recordings of piano performances.

This experiment would only take approximately 20 minutes.

Compensation: *Receive $10 CASH when you complete the experiment!*

Please contact Kelvin at kelvin.m.chan@mail.mcgill.ca for details!

This study is under the supervision of Prof. Isabelle Cossette (isabelle.cossette1@mcgill.ca).

Appendix D **Rating form for piano performance evaluation study**

Name: _____ E-mail address: _____ Date: _____

Instructions: Please think of yourself as a judge for a professional audition or competition, and give each performance a rating from 0 (as terrible as it gets) to 100 (as good as it gets) for each of the following expressive criteria: phrasing, dynamics, rubato, and overall quality. **Do not go back and change ratings for previous videos once you've started**. **Please also do not use decimal points, or leave any fields blank**. Thank you so much for your co-operation!

	Phrasing	**Dynamics**	**Rubato**	**Overall Quality**
Practice 1	/100	/100	/100	/100
Practice 2	/100	/100	/100	/100
Performance A	/100	/100	/100	/100
Performance B	/100	/100	/100	/100
Performance C	/100	/100	/100	/100
Performance D	/100	/100	/100	/100
Performance E	/100	/100	/100	/100
Performance F	/100	/100	/100	/100
Performance G	/100	/100	/100	/100
Performance H	/100	/100	/100	/100
Performance I	/100	/100	/100	/100
Performance J	/100	/100	/100	/100

-End-