NONVERBAL COMMUNICATION IN MUSICAL PERFORMANCES

CO-PERFORMER COMMUNICATION AND AUDIENCE PERCEPTION OF MUSICAL PERFORMANCES

By ANNA SIMINOSKI, B.Mus.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree Master of Science

McMaster University

© Copyright by Anna Siminoski, September 2017

McMaster University MASTER OF SCIENCE (2017) Hamilton, Ontario (Psychology, Neuroscience, and Behaviour)

TITLE: Co-performer Communication and Audience Perception of Musical Performances

AUTHOR: Anna Siminoski, B.Mus. (The University of British Columbia)

SUPERVISOR: Dr. Michael Schutz

NUMBER OF PAGES: 88

Preface

My thesis consists of two manuscripts that will be submitted independently for publication. Each draws upon different aspects of novel data collected to explore communication within a musical performance. The first paper explores communication between two performers using Granger causality analyses to see how each musician's movements affected the other under different performance conditions. The second paper explores how participants evaluated both the movements and sounds created by these performances when presented with audio-visual, audio-alone, or visual-alone stimuli. As each chapter is designed to 'stand on it's own' as a publication, there is some overlap in topics between the papers. However, they address distinct questions using unique aspects of the data set, and together help clarify the complex process of musical communication between co-performers in a duo, as well as between the duo and observers.

Acknowledgements

I would like to thank my supervisor, Mike Schutz, for giving me the opportunity to work in the MAPLE Lab. I would also like to thank my committee members, Laurel Trainor and Judy Shedden, for their excellent guidance throughout my degree. I greatly appreciate the feedback and direction I received as it helped shape my research.

I have to immensely thank Aimee Battcock for her endless support and feedback, along with Fiona Manning for guiding me in my first year. Thank you to my thesis student, Erica Huynh, for going above and beyond what she needed to do and for her continued help even after she graduated. Thank you Kimberly Germann for not only keeping the lab so organized, but also me at times. Thank you to my little army of research assistants, Maxwell Ng, Shanthika Ravindran, Nicole Divincenzo, Brannon Senger, Rachel Heo, Max Maglal-Lan, and Veronica Panchyshyn, for helping clean massive amounts of motion-capture data and running participants. I also have to thank the LIVE Lab staff, Dave Thompson, Dan Bosnyak, and J.J. Booth, for helping sort out the many technical intricacies involved in my project.

Last but definitely not least, thank you family. Thank you dad for taking the time to read my thesis and give me suggestions!

Table of Contents

List of Figures

CHAPTER 2

CHAPTER 3

List of Tables

CHAPTER 3

CHAPTER 1

GENERAL INTRODUCTION

A music performance is an exciting environment for studying joint action and nonverbal communication. This thesis utilized music ensemble performances to study nonverbal communication occurring from two different perspectives. We examined bi-directional communication between co-performers and uni-directional communication of the musicians to the participants.

The first paper (Chapter 2) used statistical measures (i.e., Granger causality analysis) to quantify head movements as a measure of joint action during a musical performance. We observed a change in direction and magnitude of information flow between co-performers based on the availability of visual and auditory cues. We wanted to extend this line of research to see how audience perception of the performances may change due to the auditory and visual manipulations of the performers. The second paper (Chapter 3) examines participant ratings of expression, cohesion, and general liking of the performance when presented with audio-visual, audio-only, and visual-only stimuli. These ratings not only show which sensory modality allows for the highest sensitivity to performer manipulations, but also provides insight into how musicians change their performances to communicate to the audience.

My aim was to create a cohesive examination of co-performer communication and observer perception of musical performances. Collectively these papers demonstrate the novel research I have completed during my Master's degree.

CHAPTER 2

QUANTITATIVE ANALYSIS OF NONVERBAL CO-PERFORMER COMMUNICATION IN MUSIC PERFORMANCE

Quantitative Analysis of Nonverbal Co-performer Communication in Music Performance

Anna Siminoski[1] & Michael Schutz [2,1]

[1]Department of Psychology, Neuroscience & Behaviour, McMaster University, Hamilton, ON, Canada

[2]School of the Arts, McMaster University, Hamilton, ON, Canada

Funding:
This work was supported by the Natural Sciences and Engineering Research Council (RGPIN/386603-2010), Canadian Foundation for Innovation-Leader's Opportunity Fund (CFI-LOF 30101), Early Researcher Award (ER10-07-195), and the McMaster Arts Research Board.

Conflicts of interest:
All authors state that they have no conflicts of interest.

Contributorship:
AS researched literature, conceived the study, collected data, analysed data, and wrote the manuscript. MS was involved in the study design, gaining ethical approval, collecting data, analysing data, and editing the manuscript. Both authors reviewed the manuscript and approved the final version.

Corresponding Author:
Anna Siminoski, 424 Togo Salmon Hall, 1280 Main Street West, Hamilton, ON, Canada, L8S 4M2, Tel: 905-525-9140, ext 23159, Email: siminosa@mcmaster.ca

Abstract

Musicians constantly adapt their actions based on a number of factors, including communicative goals with other performers, making music ensemble performance a useful paradigm for studying real-time joint action. The present study uses Granger causality analyses to quantitatively examine bi-directional information flow between highly trained clarinettists and pianists performing excerpts of well-known classical music. We motion captured musicians performing under four different conditions: (1) normal performance setting, (2) no visual/full auditory feedback, (3) full visual/partial auditory feedback, and (4) no visual/partial auditory feedback. Using Granger causality we found a main effect of manipulating performers' visual feedback, suggesting that visual information is advantageous for inter-performer communication. Additionally, there was greater informational flow amongst performers during the more complex, rhythmically involved excerpt, compared to the slow flowing excerpt. Interestingly, there was a trend for the pianist to influence the clarinettist's movements more than the clarinettist influenced the pianist, especially when the clarinettist could not hear the pianist. This research presents a novel way of exploring how interactions change when different communicative channels are manipulated. Furthermore, it suggests that a complex visual dialogue is occurring between the solo clarinettist (nominally duo's 'leader') and the collaborative pianist (nominally the 'follower'), in addition to an auditory interchange. This research also has practical implications for music pedagogy as it distinguishes which sensory modalities may be important for cohesive musical performances.

Introduction

Human interaction involves an intricate exchange of multisensory information relying on many cognitive-motor skills. Interacting with another person can be as straightforward as two people unintentionally synchronizing their rocking in chairs (Richardson, Marsh, Isenhower, Goodman, & Schmidt, 2007) or as complex as an orchestra intentionally coordinating their movements when performing a complex composition (D'Ausilio et al., 2012b). Although varied in their degree of intentionality, complexity, and difficulty, such tasks involve sophisticated faculties in order to coordinate actions temporally and spatially to achieve a common goal or to perform a joint action (Richardson et al., 2007).

Joint Action in Music Ensembles

Joint action can be achieved through the use of nonverbal communication such as body movements and/or non-language sounds. In a performance, musicians interact in real-time without speaking. The movement sequence of each musician is typically both distinct and highly complex, and movements must be coordinated precisely within tight temporal constraints. Musicians are therefore trained to become highly skilled at communicating nonverbally to co-performers and audience members. Musicians must take into account co-performers' intended actions and integrate that information with their own motor repertoire to create a cohesive performance (Loehr, 2013). These inter-musician movements are also seen by audience members and may potentially influence the way audiences perceive and evaluate musical performances. Consequently, the movements used by musicians when performing in ensembles represent a rich paradigm for studying joint action in an ecologically valid setting (D'Ausilio, Novembre, Fadiga, & Keller, 2015).

Music provides a particularly useful domain for informing our understanding of joint action for two reasons. First, musicians are used to coordinating actions by following a script (i.e. the musical score), allowing for the systematic study of intricate and highly practiced movements. Second, music is traditionally considered an auditory phenomenon with visual aspects of secondary importance. This creates an opportunity for studying joint action in situations where the stated goal of the actors deals primarily with auditory production, rather than movement per se. Many studies of joint action use situations where the actors are focused on the movement goals: for example, building Lego models (Clark & Krych, 2004), stacking blocks (Flanagan & Johansson, 2003), lifting and gripping objects (Sebanz, Bekkering, & Knoblich, 2006), and rocking in chairs (Richardson et al., 2007). Our study complements previous joint action experiments by exploring movement in a scenario where the focus is on a different modality.

Movement in Music Performance

Musicians make specific movements in order to communicate nonverbally; some of these movements are used for sound generation on instruments. Examples of these *effective gestures* include properly shaping facial muscles to allow a reed to vibrate on a clarinet mouthpiece and depressing the keys and covering the holes on a clarinet to generate the appropriate pitch (Wanderley, 2002). *Ancillary gestures,* on the other hand, are movements that do not directly affect sound production of an instrument; for instance, swaying the body, bobbing the head, or dramatically lifting the arms. Effective and ancillary gestures can both be thought of as forms of nonverbal communication (Dahl & Friberg, 2007). Effective gestures during a musical performance influence the transmission of information through the auditory domain. In speech, articulation, tone, speed, and pitch can all contribute additional information to a listener, which in turn helps clarify meaning. This is equivalent to how effective gestures can influence musical performances. For example, musicians can signal shifts in musical style and mood through tempo and volume changes. Ancillary gestures, on the other hand, influence co-performers through the visual domain. Ancillary gestures used in musical performance are analogous to body movements used when speaking. During dialogue, people punctuate and emphasize certain aspects of their speech through body language such as hand movements or nodding the head. Both body language and ancillary gestures have the ability to convey additional information to the viewer (Davidson, 1993; Teixeira, Yehia, & Loureiro, 2015; Wanderley, Vines, Middleton, McKay, & Hatch, 2005).

Ancillary gestures in music serve multiple purposes: to support sound-producing gestures (e.g., moving the instrument downwards when taking a breath), to transmit information to the audience, and to communicate with other musicians in the ensemble (Lähdeoja, Wanderley, & M. Malloch, 2009; Wanderley, 2002). Here we have chosen to focus our attention on the latter: how information is communicated amongst co-performers using ancillary gestures. Gestures become ingrained in the mannerisms of highly trained musicians and are consistent between performances of the same piece (Wanderley, 2002; Wanderley et al., 2005). Therefore, ancillary gestures of professional musicians can be analysed in meaningful ways.

When studying social interactions, it is difficult to encompass entire bodies in analyses, so choosing a representative movement dimension is helpful. Past musical performance studies have measured ancillary movements using body sway and showed that body sway was used for nonverbal communication between co-performers (Badino, D'Ausilio, Glowinski, Camurri, & Fadiga, 2014; Goebl & Palmer, 2009; Keller & Appel, 2010). Moreover, Chang, Livingstone, Bosnyak, & Trainor (2017) clarified that body sway was not just a motor by-product, but was representative of inter-musician communication. We have therefore chosen to analyse body sway as quantified by kinematic head movements of performers. Head movements are representative of entire body sway in addition to movements unique to neck joint rotation, and are useful in measuring gestures during a musical performance (Teixeira, Loureiro, Wanderley, & Yehia, 2014; Verfaille, Quek, & Wanderley, 2006; Vines, Krumhansl, Wanderley, & Levitin, 2006; Wanderley, Vines, Middleton, McKay, & Hatch, 2005). Ragert, Schroeder, and Keller (2013) measured both body sway and head movements, finding head movements to show greater interpersonal coordination than body sway. Consequently, head movements might offer greater sensitivity for observing differences in inter-performer communication between our experimental manipulations. Here we examine head movements of musicians playing two different instruments. This allows us to observe how musicians coordinate when using movements unique to their individual instruments. Additionally, we examine movement of the clarinet bell to provide a useful point of comparison with previous studies focused on bell movement (Teixeira et al., 2014, 2015).

Musical Instrumentation

Many studies examining the role of ancillary gestures during musical performance focus on pianists. These have included the velocity of pianists' head and body movements (Castellano, Mortillaro, Camurri, & Volpe, 2008); leader/follower relationships in piano duets (Goebl & Palmer, 2009); and finger height in relation to synchonization (Goebl & Palmer, 2009). Additionally, topics such as the structural components of music in relation to timing and physical demands (Palmer & Loehr, 2013); effects of auditory feedback on synchonrization (Zamm, Pfordresher, & Palmer, 2014); the role of action simulation improving synchroncy (Keller et al., 2007); and cross-modal interactions in a piano performance have been studied (Vuoskoski, Thompson, Clarke, & Spence, 2014). There is also considerable work studying the ancillary gestures of clarinettists, such as examining consistency of gestures (Teixeira, Loureiro, Wanderley, & Yehia, 2014; Wanderley, 2002); types of body movement used for gesturing (Verfaille, Quek, & Wanderley, 2006); gestures and their relation to expressive content (Davidson, 2012; Palmer, Koopmans, Carter, Loehr, & Wanderley, 2009); the relationship between the structure of the music and gestures (Wanderley, Vines, Middleton, McKay, & Hatch, 2005); and cross-modal interactions during a clarinet performance (Vines et al., 2006).

Many previous studies of clarinetists' gestures look at the movements of the clarinetist alone. In some cases this captured the entirety of the musical score – i.e., Stravinsky's *Three Pieces for Solo Clarinet* (Vines et al., 2006; Wanderley et al., 2005) requires one musician. But in other cases, movement analyses focused on only a subset of the gestures involved. For example studies of Mozart's *Clarinet Concerto* in A major (Palmer et al., 2009), Mozart's *Quintet for Clarinet and Strings* in A major (Teixeira et al., 2014), and Brahms' *Clarinet Sonata No. 1* in F minor (Wanderley, 2002; Teixeira, Yehia, & Loureiro, 2015) examine only clarinetists' movements within pieces written for multiple musicians. Although these studies provide useful insight into the musical role of ancillary gestures, overlooking the potential communicative role of these gestures could lead to challenges with their interpretation.

Fully understanding the role of ancillary gestures is hard given the complex, multi-layered nature of musical structure and performance. For example, Teixeira et al. (2015) suggest that the ancillary gestures used by clarinettists playing Brahms' *Clarinet Sonata No. 1* are for musical realization of the written part. The authors propose that the musical intentions of the performers are the reason that recurrent gestures are strong at phrase boundaries, harmonic transitions, and dynamic changes. Although these reasons undoubtedly play a role in the performer's ancillary gestures, we suspect certain aspects of such movements represent a history of communication between musicians in previous performances. Gestures could be employed not only to enhance expression, but also to synchronize parts between the clarinettist and pianist. As an example, if there is a tricky transition between musical passages, the clarinettist may move the bell of the clarinet more rigorously to clearly demarcate his/her intended melodic path. Here we quantitatively investigate aspects of clarinettists' movements previously overlooked when studying only one part of a musical duet.

We chose a clarinet and piano duet in order to observe communication between performers in a setting of asymmetrical roles determined by preconceived ideas of which positions the instruments should take. The clarinet is usually thought of as the *solo* instrument, while the piano is meant to *accompany* the clarinettist. This intuitively implies that the clarinettist is the leader, while the pianist is the follower. Additionally, different effective gestures are involved with playing the piano versus playing the clarinet, so it is of interest to examine coordination of head movements between co-performers despite the movement constraints uniquely imposed by each instrument. To explore coordination between two instrumentalists simultaneously, we used a statistical tool – Granger causality analyses – to examine the magnitude and direction of nonverbal information flow between clarinettists and pianists (Granger, 1969).

Granger Causality

Granger causality is a statistical method used to quantify directional information flow between two or more time series (see Data Analysis section for more detail). There is growing literature in the field of psychology and neuroscience that uses Granger causality to analyse many forms of time series data (Seth, Barrett, & Barrett, 2015). Recently, Granger causality has been applied to quantify actions in musical ensemble performance (D'Ausilio et al., 2012a, 2012b; Glowinski & Badino, 2012; Papiotis, Marchini, Perez-Carrillo, & Maestre, 2014). For example, Badino et al., (2014) used this approach to show that more complex music increased communication among performers in a string quartet. Furthermore, Chang et al. (2017) studied joint action in string quartets by assigning different members to be 'secret leaders' in different trials. Granger causality analysis showed assigned leaders to be more influential of followers' movements than vice versa. They also manipulated visual feedback by having musicians face each other or turn away from one another. Although the leader influenced followers' movements in both conditions, this influence was more pronounced in the presence of visual information. The authors concluded that both auditory and visual cues are used for nonverbal communication in a music performance. In the current study, we apply Granger causality analysis to examine the bi-directional information flow between clarinettists and pianists when performing well-known classical music excerpts.

The Present Study

We investigated the use of movement by clarinettists and pianists performing as a duo in relation to solo-collaborator roles, auditory feedback, and visual feedback. Musicians performed Brahms' *Clarinet Sonata No. 1* for clarinet and piano, either by themselves (solo condition) or together (duet condition, clarinet and piano). Acoustically, musicians received either full (both clarinet and piano parts) or partial (the clarinettist could not hear the pianist, but the pianist could hear both parts) auditory feedback. Visually, musicians either saw (full visual) or did not see (no visual) each other's movements. This allowed us to explore whether ancillary gestures are so ingrained in musicians' mannerisms and performance styles that they remain consistent regardless of these manipulations, or if there is some flexibility and dynamic interchange amongst performers. Moreover, we wanted to see how these variables influence the usage of ancillary gestures for co-performer communication.

The current study focuses on head movements – which we believe serve as the most direct point of comparison when examining musicians playing instruments requiring very different types of movements. Our objectives were: (1) to explore how co-performers influence each other's movements, (2) to determine whether gestures change when auditory feedback is diminished, and (3) to assess changes in gestural usage when collaborating musicians could not see each other and could only communicate through sound. We predicted bi-directional information flow between performers, consistent with previous findings showing that performers use gestures to communicate and influence one another's playing (Badino et al., 2014; Chang et al., 2017, D'Ausilio et al., 2012b). Specifically, Keller & Appel (2010) observed that body sway communicated high-level time scale coordination – i.e. longer musical phrasing instead of note-to-note synchronization – among co-performers in a piano duo. We also predicted that reduced auditory feedback would increase the use of ancillary gestures for communication purposes between the performers, consistent with Goebl and Palmer's (2009) findings. Goebl and Palmer (2009) reduced auditory feedback for pianists playing duets and noticed that pianists' head movements became more synchronized as the auditory information was reduced. The authors suggested that visual cues become more important and may serve as an alternative form of communication as the readability of auditory information decreases. Chang et al. (2017) found leaders to influence followers more when visual communication was intact. Therefore, we predicted that when performers could not see each other, gesture use would not be as prevalent.

Method

Participants

Six professional musicians from the Greater Toronto Area participated in the experiment for monetary compensation. Three clarinettists (1 female) with 22, 45, and 50 years of playing (years of training = 14, 10, and 12) and three pianists (2 female) with 34, 35, and 36 years of playing (years of training = 30, 35, and 28) were motion-captured and audio-video recorded. Professionals were chosen as subjects because it has been shown that highly experienced musicians demonstrate more consistent gestures than less experienced performers (Wanderley, 2002). We wanted highly consistent gestures so we could reliably study them in our experimental setup resembling a professional musical performance. Each clarinettist performed with each pianist, resulting in nine pairings. Only one pairing had previously performed together, but with minimal experience as a duo and never on the piece used for this experiment. We tried to recruit novel pairs to control for musician’s familiarity with each other’s performance styles. All subjects reported normal hearing and right-handedness. This study met the criteria set by the McMaster University Research Ethics Board.

Stimuli

Musicians performed two excerpts from Brahms' *Clarinet Sonata No. 1* in F minor, considered standard repertoire for professional clarinettists. The excerpts were selected for their contrasting rhythmic meters, tempi, and emotional/ expressive material. The first excerpt (excerpt A) consisted of measures 1 to 28 of the first movement (about 45 seconds), which has been used in several previous studies (Teixeira et al., 2014, 2015; Wanderley, 2002). In 3/4 time and marked *allegro appassionato*, this movement is intended to be played fast and with passion. It provides an opportunity for performers to manipulate many musical factors – dynamics, timbre, phrasing, timing, and accents – providing great expressive content. In contrast, the second excerpt (excerpt B; measures 1 to 17 of the second movement; about 50 seconds) is marked *andante un poco adagio* – moderately slow. It is in 2/4 time, with the clarinet playing a lyrical melody that is supported by sustained notes in the piano part.

Materials

A Qualisys motion-capture system[1] recorded the participants' movements when performing in the LIVE Lab at McMaster University. Clarinet players wore 18 reflective markers to allow full body movements to be captured. The clarinet had two markers, one on the bell and another on the barrel. Piano players wore 14 reflective markers to capture the movements of the upper half of the body. The

---

[1] http://www.qualisys.com/

[2] The other markers were used to generate point-light display musicians for another

piano had two markers on either side of the keyboard. Here we analysed only the forehead marker on each musician, and the marker on the bell of the clarinet[2]. 18 Qualisys cameras recorded the infrared signals reflected off the markers using the Qualisys Track Manager (QTM) software.[3]

The clarinettists brought their own professional model clarinets. An AKG C414 XLS directional microphone, placed within a microphone shield to minimize the sound of keyboard noise from the piano, recorded the audio of the clarinet to a computer running Reaper software[4]. The pianists used a Roland FP-80 digital piano that recorded the MIDI output to a computer running Reaper. The clarinettists wore earplugs (32 dB noise reduction rating) to eliminate the sound of piano key clicks. This ensured no rhythmic information was communicated through sound in the partial auditory feedback condition. The clarinettist also wore Sennheiser HAD 200 closed-back headphones and the pianist wore NADA QH560 open-back headphones. The piano was played through both headphones. The clarinet was played through the clarinettist's headphones to imitate a natural sound level through the earplugs. The clarinet was not played through the pianist's headphones as they could hear it naturally in the room. The volumes were adjusted until musicians indicated they could hear each other reasonably well despite the setup. The instruments were heard through the headphones according to the various feedback conditions (see below for details).

Design and Task

---

[2] The other markers were used to generate point-light display musicians for another experiment.

[3] http://www. qualisys.com/software/qualisys-track-manager/

[4] http://www.reaper.fm/

We independently manipulated both visual and auditory feedback, creating four performance conditions (see fig. 1). Musician pairs could either see each other (Fv: full visual feedback) or had an acoustically transparent screen placed between them (Nv: no visual feedback). They could either both hear each other (Fa: full auditory feedback) or the clarinettist could only hear themselves, while the pianist could still hear both parts (Pa: partial auditory feedback). The clarinet part was never silenced, as sound exits the instrument from its many holes making muting it impractical. All participants completed all conditions in this 2 (visual feedback) x 2 (auditory feedback) within-subjects design.

	Full Visual Feedback	No Visual Feedback
Full Auditory Feedback	Condition 1 FvFa	Condition 2 NvFa
Partial Auditory Feedback	Condition 3 FvPa	Condition 4 NvPa

Fig. 1. Summary of the four performance conditions. In conditions involving partial auditory feedback, the clarinettist could not hear the pianist.

The musicians received the musical excerpts several weeks in advance with instructions to learn the music to performance quality. Memorization was not required and printouts of the excerpts were provided on the music stands. The experiment took place in the LIVE Lab at McMaster University, which contains a

106-seat concert hall. It occurred over three days, with a different pianist on each day performing with all three clarinettists. On the day of the experiment, musicians first performed each of the excerpts solo three times and then recorded the 24 duo trials. The presentation order of the duet conditions (condition 1 – FvFa: full visual feedback, full auditory feedback; condition 2 – NvFa: no visual feedback, full auditory feedback; condition 3 – FvPa: full visual feedback, partial auditory feedback; condition 4 – NvPa: no visual feedback, partial auditory feedback) was randomly assigned to the musician pairings. There were 4 duet blocks with 6 trials (3 repetitions of 2 excerpts) and the solo condition where each excerpt was played 3 times, totaling 30 (6 solo, 24 duo) trials per participant. Each trial began with an experimenter announcing the trial's performance parameters (e.g. "excerpt A, no piano feedback") and then letting the participants begin playing on their own. Excerpt A contains a 4 bar piano introduction without any clarinet, so for condition 4 (NvPa: no visual/partial auditory) both musicians heard the first 4 bars then the piano was silenced for the clarinettist just as the clarinet was to enter. Each trial was motion-captured, audio-recorded, and video-recorded. The musicians were instructed to perform the pieces as if they were playing for an audience, regardless of the experimental manipulations. Short breaks were taken between blocks to set up or take down the screen and whenever needed by the musicians. The experiment lasted approximately 60 minutes for each pairing of musicians.

Data Analysis

Data Processing

The present study examines head movement kinematics and clarinet bell movement. The clarinet bell marker was analysed to see how accurately the head movements of the clarinettist encompass the movement of the clarinet itself.

In the Qualisys Track Manager (QTM) software, data were exported after verifying that the forehead marker on each musician and the marker on the clarinet bell were fully captured for the entire duration of each trial (100% of the epoch length) before we exported the data. In trials where data were missing (largest gap was 2% of the epoch length), most likely due to brief occlusions of markers or recording noise, the spline interpolation feature in QTM was used to gap fill. The motion capture data was then exported to MATLAB (Mathworks Inc.) where the following calculations were performed.

First, the original data acquisition of 100 Hz was down-sampled to 10 Hz by averaging every frame within a non-overlapped 10-frame window. This creates a more accurate 10 Hz representation then taking very 10th frame (Chang et al., 2017), as it avoids missing information should the movements fluctuate evenly up and down within the 10-frame period. Next, velocity was calculated for each dimension (x, y, z) of the 3 dimensional trajectories and times series of the 3D Euclidean distances were derived from the individual velocities of x, y, and z. The times series were z-score normalized, controlling for individual differences in movement magnitude. We created normalized time series for the head movement velocities of each performer for each trial and for the clarinet bell movement of each clarinettist for each trial.

Granger Causality Analysis

Granger causality (G-causality) uses predictive histories of multiple time series to determine if one time series is influencing the other. The basic principle is that time series A *Granger-causes* time series B if the past information of A is better able to predict B than the past information of B itself (Granger, 1969). In terms of our experiment, if the pianists' movements are better predicted by past movements of the clarinettist than by their own past movements, we have observed G-causality. We examined the direction and magnitude of information flow from the clarinettist to pianist and pianist to clarinettist (as well as averaged the individual directions to examine bi-directional communication) to explore how movement patterns changed as a function of performance condition.

All Granger causality calculations were performed using the Multivariate Granger Causality (MVGC) toolbox[5] (Barnett & Seth, 2014). We followed the steps outlined in the MVGC toolbox in order to compute G-causality values with our z-normalized time-series data (Barrett, Barnett, & Seth, 2010; Seth, 2010). The magnitude of G-causality was calculated within each musician pair using their head movements, and then the pianist head movement to the clarinettist bell movement. The first two trials for each pair in each condition were treated as practice trials, and the third trial was used for subsequent statistical analyses. A model order of 17 was used, indicating the length of history considered in the G-causality calculation was 1700ms (17 frames). The same model order was used for both excerpts A and B in order to make direct comparisons between the excerpts (see Barnett & Seth,

---

[5] http://www.sussex.ac.uk/sackler/mvgc/

2014 for a detailed explanation of how to choose the appropriate model order). From the nine pairs of musicians performing two excerpts under the four different conditions we calculated 72 G-causalities detailing the information flow from the clarinettist to pianist, 72 G-causalities detailing the information flow from the pianist to clarinettist, and 72 G-causalities stating the causal density, or the averaged flow of information between a pair of musicians.

Results

All of the following statistical analyses were conducted using R (R Core Team, 2013). The Shapiro-Wilk's test reported that the G-causality values from both excerpt A and excerpt B did not follow normal distributions. The Levene test showed that both excerpt A and B did not have homogeneous variation. The assumptions of parametric tests were not met and sample size was small (9 pairs of musicians), so nonparametric tests were performed.

Analysis of Movement Coupling

Pianist head to clarinettist head. Three linear models were created to examine a two-way interaction (2 vision conditions x 2 audio conditions) using causal density collapsed across both excerpts A & B. Wilcoxon's signed-rank test for matched pairs was applied to each linear model. Here we compared the head movements of the pianist to the head movements of the clarinettist. The results showed a significant main effect of vision (W = 143, p = .010), but no significant main effect of audio (W= 54, p= 0.18), and no vision by audio interaction (W = 101, p

= 0.52). Post-hoc Wilcoxon signed-rank tests were performed, comparing all conditions (6 multiple comparisons). G-causalities in condition 3, when musicians could see each other, but the clarinettist could not hear the pianist, were significantly higher than in condition 2, where musicians could not see each other, but could still hear fully (W = 96, p = .037). No other conditions were significantly different (Fig. 2). After using Bonferroni adjustments (alpha = .05/6 = .0083) to correct for multiple comparisons, the results of these post-hoc tests were no longer statistically significant.

Pianist head to clarinet bell. We used the same linear models to examine the head movements of the pianist to clarinet bell movements made by the clarinettist. The results showed no main effect of vision (W = 106 p = 0.39), no main effect of auditory information (W = 56, p = 0.21), and no significant interaction (W = 91, p = 0.83) (Fig. 3).

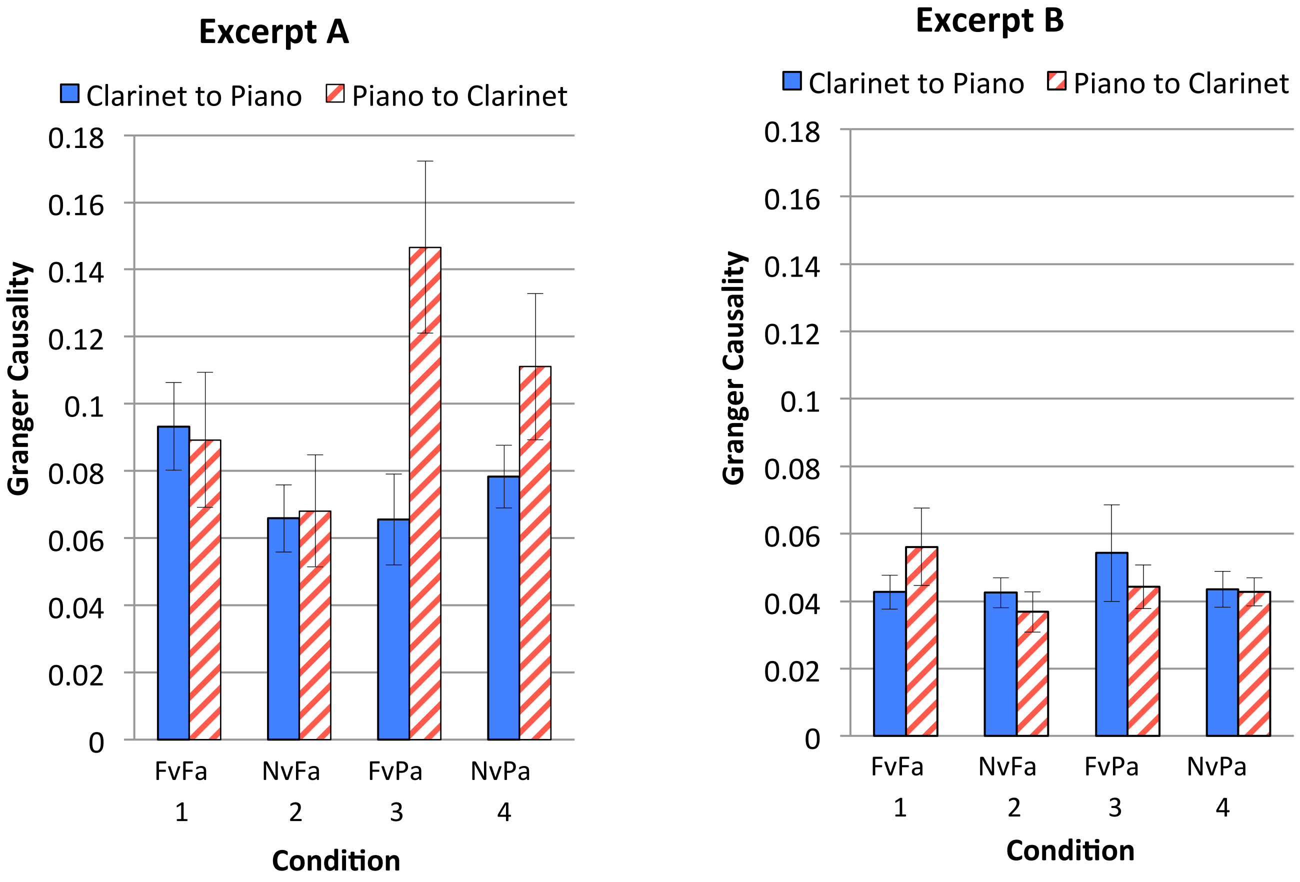
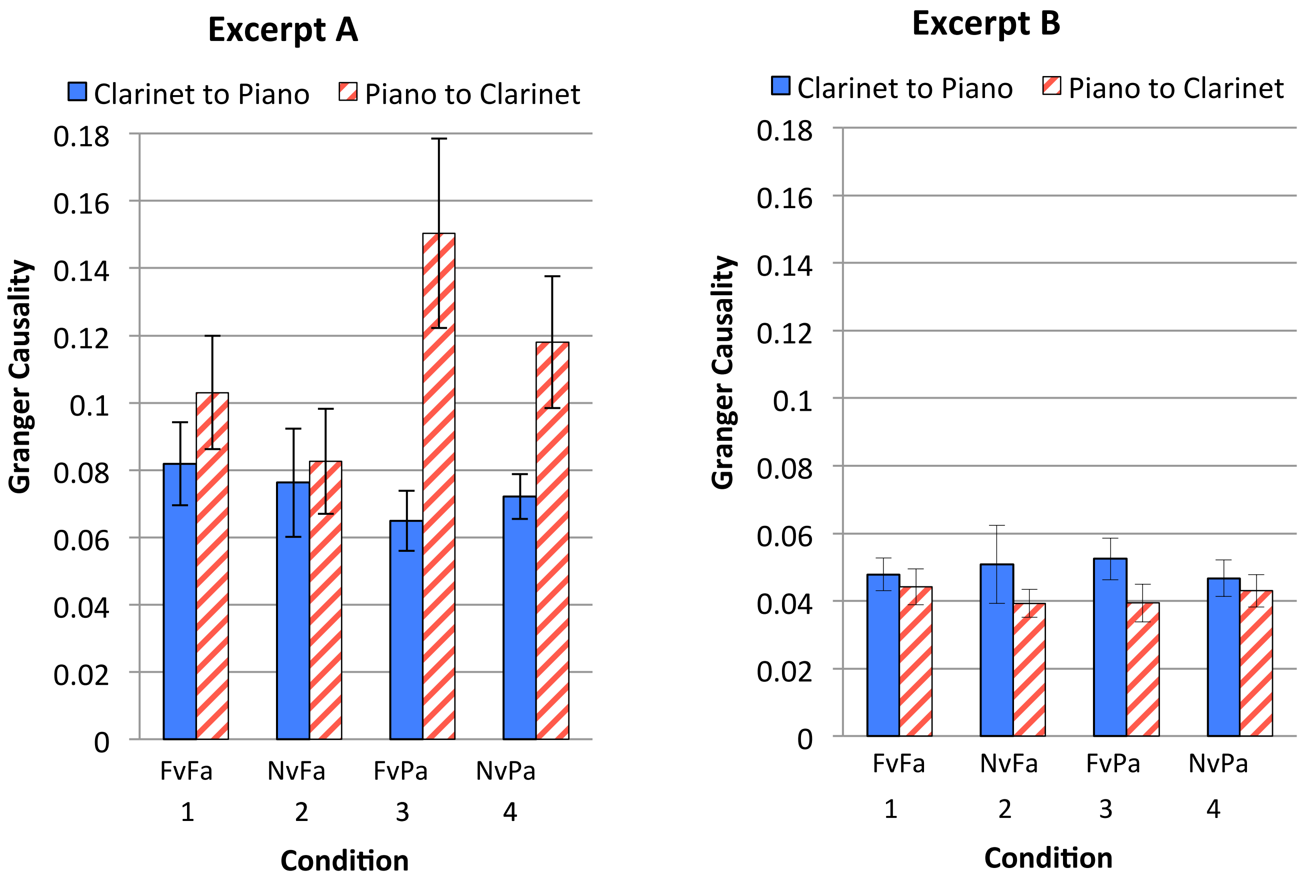


Fig. 2. Granger-causalities (GC) of musicians' head to head couplings. GCs for each condition determining the amount clarinet influences piano (solid bars) and the amount piano influences clarinet (striped bars), for excerpt A and excerpt B. There was a significant main effect of vision (p = .010) when excerpts A and B were averaged together. Error bars represent standard error.


Excerpt A

Excerpt B

Fig 3. Granger-causalities of pianist head to clarinet bell couplings. GCs for each condition determining the amount clarinet influences piano (solid bars) and the amount piano influences clarinet (striped bars), for excerpt A and excerpt B. In excerpt A there was a significant difference bewteen direction of information flow. Error bars represent standard error.

Differences Between Clarinettists and Pianists

Pianist head to clarinettist head. A Wilcoxon signed-rank test compared the information flow between clarinettists and pianists using the forehead markers. A within group test is appropriate since musicians were always manipulated in the same counterbalanced pairings. No significant results were found when we collapsed across excerpts (W = 2403, p = 0.45) or when information flow was inspected separately in excerpt A (W = 523, p = 0.16) and excerpt B (W = 343, p = 0.88). Through visual inspection of the data (Fig. 2), however, there is a directional trend for the information to flow from the pianist to the clarinettist, not vice versa. The trend is especially prevalent in excerpt A, condition 3, where vision is intact but auditory feedback is limited for the clarinettist. A post-hoc comparison of the directions of information flow between the clarinettist and pianist in excerpt A, condition 3, was conducted using a Wilcoxon signed-rank test and showed the pianist was significantly influencing the clarinettists' movements (W = 14, p = .019) (Fig. 2). Adjusting for multiple comparisons (Bonferroni, 0.05/4 = .0125), this result is not significant so has to be interpreted with caution.

Pianist head to clarinet bell. We used Wilcoxon signed-rank tests to compare information flow between the pianists and clarinettists using the head marker of the pianist and the clarinet bell marker of the clarinettist. We found a significant difference in excerpt A (W =387, p = .003), showing the pianist was more influential of the clarinettists' bell movements than vice versa (Fig. 3). No significant differences were found for excerpt B (W =428, p = 0.14) (Fig. 3), or when we collapsed across both excerpts A and B (W = 2404, p = 0.45) (Fig. 3). Multiple

Wilcoxon signed-rank tests were conducted post-hoc to establish the direction of information flow within each condition of excerpt A. The tests revealed that the pianist was significantly influencing the clarinettists' bell movements in condition 3 (full visual, partial audio; $W =$, $p = .014$) and condition 4 ($W =$, $p = .031$). After Bonferroni corrections are applied (0.05/4 = .0125), condition 3 is marginally significant, but condition 4 is no longer significant. Therefore, the results of this exploratory study have to be interpreted with caution.

Differences Between Musical Excerpts

Excerpt A was compared to Excerpt B using a Wilcoxon sign-rank test for matched pairs. We performed tests on both head-to-head and head-to-bell G-causality data and found the same results. The excerpts were significantly different when tested using causal density (head-to-head, $W = 629$, $p < 0.001$; head-to-bell, $W = 633$, $p < 0.001$), G-causality of the clarinettist influencing the pianist (head-to-head, $W = 570$, $p < 0.001$; head-to-bell, $W = 532$, $p = 0.001$), and G-causality of the pianist influencing the clarinettist (head-to-head, $W = 613$, $p < 0.001$; head-to-bell, $W = 641$, $p < 0.001$) (Fig. 4).

In order to explore whether the head-to-head G-causality values observed in this experiment differed from chance, we evaluated values arising from mismatched trials in which true coordination would have been impossible. This involved systematically pairing movements of the clarinettist in one trial with movements of the pianist in all other trials. This approach yielded 1260 artificial pairings per excerpt: 36 clarinettist trials X 35 pianist trials (omitting all matched pairings). We calculated G-causality values for all permutations of mismatched trials, finding bi-directional information flow yielded a G-causality of 0.0660 for excerpt A and 0.0438 for excerpt B. Using Wilcoxon sign-rank tests for matched pairs, we found our experimental G-causalities to be significantly higher than this chance level for excerpt A ($W = 516$, $p < 0.01$), but not for excerpt B ($W = 309$, $p = 0.72$).

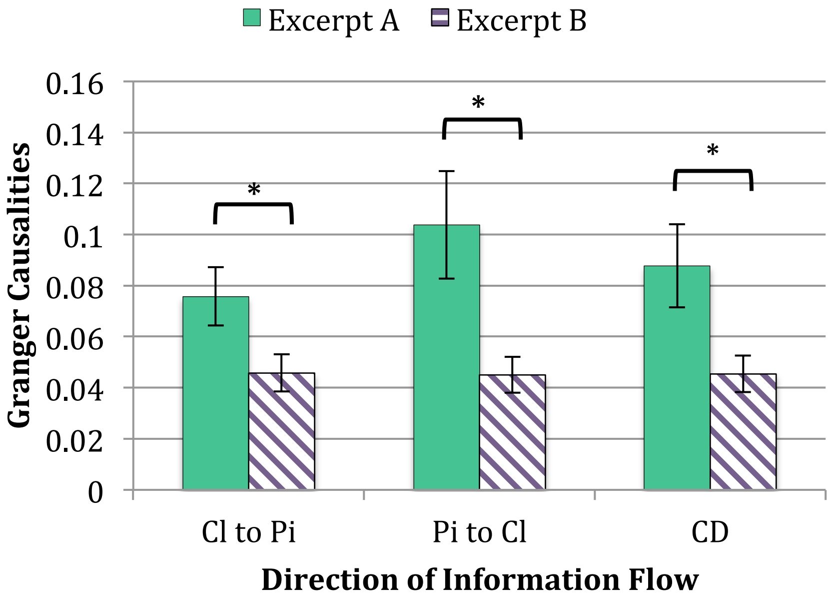


Fig. 4. Differences between musical excerpts for head to head couplings. Granger-causalities averaged for each direction of information flow (clarinet to piano, piano to clarinet, and causal density – the average of both directions) for each excerpt. Errors bars represent standard error and asterisks denote significant differences (* $p < .001$).

Discussion

The results indicate that musicians are communicating nonverbally through the use of body movements. Bi-directional information flow is occurring between pairs of musicians and is dependent on available sensory cues and musical structure. Information flow occurring between co-performers was greater when performers could see each other than when they could not, regardless of the auditory feedback being received (see fig. 2). These findings support past research showing that seeing a co-performer helps musicians coordinate their auditory-motor interactions (Chang et al., 2017). It is important to note however, that visual and auditory feedback manipulations were not equally balanced in our experiment. The visual feedback manipulation was stronger in that it affected both musicians, while the auditory feedback manipulation only directly affected the clarinettist. This imbalance is likely influencing our results.

Previous work has shown that musicians communicate nonverbally through the use of sound (e.g. audibly breathing in time or dynamic changes in instrument sound production) and vision (e.g. head nodding or circling of the clarinet bell to cut off a phrase) (Lähdeoja et al., 2009; Wanderley et al., 2005). The current study explicates which cues may take precedence under different sensory conditions. The highest overall G-causality was present in condition 3 (FvPa), where musicians could see each other but the clarinettist could not hear the pianist. This is consistent with our prediction that restricting auditory feedback would increase the musicians' reliance on gestures for communication. The musicians are affected by the auditory domain deficit and adapt using gestures. Musicians switch from auditory cue-dependent to more visual cue-dependent behaviour. This is similar to what happens in a conversation when the quality of auditory feedback is limited; for example, in a crowded noisy room, people rely more on looking at the lips of the speaker to comprehend words than when in a quiet room (Sumby & Pollack, 1954). When visual cues are more informative, people use this information. Condition 3 (FvPa) was only significantly different from condition 2 (NvFa; when the musicians could not see each other, but they could both fully hear each other) when multiple comparisons were not controlled for. Due to the complex nature of assembling balanced pairs of highly trained musicians, we had a small sample size. This could be contributing to our low signal and null results. Future research with additional musician pairs to increase sample size would afford greater power to explore this issue, clarifying the role of both sight and sound in co-performer communication.

Members of a musical duo comprised of a clarinettist and a pianist are traditionally thought to have predetermined roles, with the clarinettist being the solo musician while the pianist is the collaborator. These roles inherently imply that the clarinettist is the leader while the pianist is the follower. Past research has taken interest in leader-follower dynamics during musical ensemble performance and shows that the leader of an ensemble is more influential than the followers (Badino et al., 2014; Chang et al., 2017). However, to the best of our knowledge previous research has not examined the direction of information flow in the complex soloist-collaborator relationship. Surprisingly, our results indicated that there was a trend for the pianist (the collaborator) to be more influential on the clarinettist's (the soloist's) movements. The pianist was particularly influential of the clarinettist's movements when the clarinettist could not hear the pianist, but the musicians could still see each other (condition 3; FvPa). It may be the case that the pianist was receiving all of the sensory information, while the clarinettist was lacking information, so the pianist altered movements to take on the role as leader. From our experimental manipulation and data, there was no direct evidence that they were deliberately switching roles, so future research is needed to examine this hypothesis.

There is another possibility as to why the trend was for the pianist to be more influential over the clarinettist's gestures than vice versa. Brahms' *Clarinet Sonata No. 1* is a standard piece in the clarinet repertoire, and therefore all of the clarinettists in this experiment were familiar with both the piano and clarinet parts. The pianists, on the other hand, had never played the piece prior to receiving it for the experiment (pianists did receive music with both instrument parts). The clarinettists' greater familiarity with the musical score may have allowed them to be more attentive to the pianists' nonverbal communication. Ragert et al., (2013) studied the effect of familiarity with musical parts on interpersonal coordination and concluded that having knowledge about the structure of the music may be more important than familiarity with co-performers' individual playing styles. Therefore, the imbalance of familiarity with the musical parts between the pianists and clarinettists in our study may have influenced our results. A future study to address this question could use familiar and novel stimuli under the same paradigm as the current study to see how familiarity influences information flow in a soloist-collaborator dynamic.

In addition to our main analysis of head to head movements, we included a secondary analysis examining pianists' head to clarinet bell movements (see fig. 3). We wanted to see if clarinet bell movement reflected clarinettists' head movements, but our findings do not lead to a conclusive answer. When comparing information flow between head to bell movements the sensory manipulations did not have a significant effect, but we did observe a main effect of vision with the head to head analysis. Interestingly, for the head to bell analysis we saw movement coupling where the pianist was significantly more influential of the clarinettists' bell movements than vice versa in excerpt A. This significance was not observed in the head to head movements. Therefore, we conclude that both the head movement of the clarinettist and the movement of the bell capture gestural patterns used for communication between performers, but further research is needed to explain the discrepancies in effect between clarinettist head and bell movements.

A limitation in our experimental manipulation of auditory feedback was that only the piano was silenced for the clarinettist, while the clarinet was never silenced for the pianist. The clarinet's structure makes muting it nearly impossible without significantly interfering with the performer's movements. Future studies could experiment with different ways to 'silence' a clarinet, such as having the clarinettist play in a separate room while their image is projected onto a screen for the pianist to see in the lab in real-time. For this initial study we prioritized the ecological validity of the 'real' performance over such control. Here we found pianists tending towards a leading role when the clarinettists' auditory cues were diminished. However, reducing the pianists' auditory cues in a future study might lead to clarinettists taking on a more controlling role. An even auditory feedback design would help shed light on when and why musicians are exchanging dominant roles in musical performance.

The style of musical piece also had an effect on gestural information flow between performers. Excerpt A had higher G-causality values amongst musicians than excerpt B (see fig. 4). The G-causalities values for excerpt A were also significantly higher than our permutation results, meaning the experimental G-causalities values were significantly different from chance. In contrast, the G-causality values for excerpt B were not significantly different from chance. Excerpt A is considered more rhythmically complex and typically musicians use more *rubato* – musically expressive slowing down and speeding up – than excerpt B. Badino et al. (2014) formed similar conclusions, where more difficult-to-coordinate music increased gestural coupling between co-performers. The present study used contrasting excerpts that were both from the same Classical music composition. Chang et al. (2017) used Baroque and Classical pieces in a string quartet setting, and found differences between the two musical eras. Body sway coupling was more equal among the musicians for Baroque excerpts compared to Classical pieces. Additionally, Chang et al. (2017) used a leader-follower manipulation where one of the musicians was assigned to serve as the secret leader of each trial. The other musicians were better able to detect whom the leader was during the Baroque compared to the Classical music. Even though we did not use excerpts from two different musical eras, we still found robust differences between excerpts. Also, the current study did not manipulate leader/follower roles, but rather let roles emerge organically as a function of performance condition within the context of a collaborative duo. We chose Classical music, which in recent research has resulted in lower G-causality values and therefore lower bi-directional information flow. It would be interesting to examine musical pieces that have more equally distributed parts, more regularly paced tempi, and consistent rhythmic phrases using a similar paradigm to the present study. Using these different styles of pieces might lead to more equal information flow between the two performers.

In addition to contributing to theoretical knowledge of inter-musician communication, this research holds implications for musical pedagogy and education. Music instructors often tell musicians to watch one another and coordinate movements before actually making sound. It is taught that visual cues exchanged amongst performers help improve the auditory aspect of a performance. Our study supports this traditional pedagogical approach as we found vision to play an important role in nonverbal communication. Musicians who can see each other on stage and make an effort to coordinate through gestures may create more cohesive and potentially expressive performances compared to musicians who rely on the auditory component alone.

In conclusion, the present study demonstrated how Granger causality could be a useful quantitative approach in determining information flow between two individuals. Furthermore, we showed that a musical ensemble is a useful and ecologically realistic scenario to study joint action. Head movement kinematics were shown to be used for nonverbal communication between co-performers. Furthermore, we found musical characteristics and the type of sensory feedback received by musicians to influence the amount and direction of communicative gestures between musicians. When performers could see each other, nonverbal communication increased. When the musical score was intricate, nonverbal communication increased. Ancillary gestures seem to be an important component in inter-musician communication and facilitate achieving the common goal of a cohesive musical performance.

Acknowledgements

We would like to thank our research assistants, Erica Huynh, Maxwell Ng, Max Maglal-Lan, and Veronica Panchyshyn, for helping with this project. Also, thank you to the LIVE Lab staff, Dave Thompson, Dan Bosnyak, and J.J. Booth, for helping with the technology. This research was financially supported through grants to Dr. Michael Schutz from the Natural Sciences and Engineering Research Council of Canada (NSERC RGPIN/386603-2010), Ontario Early Researcher Award (ER10-07-195), and the Canadian Foundation for Innovation (CFI-LOF-30101).

References

Badino, L., D'Ausilio, A., Glowinski, D., Camurri, A., & Fadiga, L. (2014). Sensorimotor communication in professional quartets. *Neuropsychologia*, *55*(1), 98–104. https://doi.org/10.1016/j.neuropsychologia.2013.11.012

Barrett, A. B., Barnett, L., & Seth, A. K. (2010). Multivariate Granger causality and generalized variance. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, *81*(4). https://doi.org/10.1103/PhysRevE.81.041907

Castellano, G., Mortillaro, M., Camurri, A., & Volpe, G. (2008). Automated analysis of body movement in emotionally expressive piano performances. *Music Perception*, *26*(2), 103–119.

Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, *50*(1), 62–81. https://doi.org/10.1016/j.jml.2003.08.004

D'Ausilio, A., Badino, L., Li, Y., Tokay, S., Craighero, L., Canto, R., … Fadiga, L. (2012a). Communication in orchestra playing as measured with granger causality. In *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering* (Vol. 78 LNICST, pp. 273–275). https://doi.org/10.1007/978-3-642-30214-5_37

D’Ausilio, A., Badino, L., Li, Y., Tokay, S., Craighero, L., Canto, R., … Fadiga, L. (2012b). Leadership in orchestra emerges from the causal relationships of movement kinematics. *PLoS ONE*, *7*(5). https://doi.org/10.1371/journal.pone.0035757

D’Ausilio, A., Novembre, G., Fadiga, L., & Keller, P. E. (2015). What can music tell us about social interaction? *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2015.01.005

Dahl, S., & Friberg, A. (2007). Visual perception of expressiveness in musicians’ body movements. *Music Perception*, *24*(4), 433–454. https://doi.org/10.1525/rep.2008.104.1.92.This

Davidson, J. W. (1993). Visual Perception of Performance Manner in the Movements of Solo Musicians. In *Psychology of Music* (Vol. 21, pp. 103–113). https://doi.org/10.1177/030573569302100201

Davidson, J. W. (2012). Bodily movement and facial actions in expressive musical performance by solo and duo instrumentalists: Two distinctive case studies. *Psychology of Music*, *40*(5), 595–633. https://doi.org/10.1177/0305735612449896

Flanagan, J. R., & Johansson, R. S. (2003). Action plans used in action observation. *Nature*, *424*(6950), 769–771. https://doi.org/10.1038/nature01861

Glowinski, D., & Badino, L. (2012). Analysis of leadership in a string quartet. *… Workshop on Social …*, (January). Retrieved from http://www.researchgate.net/publication/232237802_Analysis_of_Leadership_in_a_String_Quartet/file/9fcfd50ae49023179e.pdf

Goebl, W., & Palmer, C. (2009). Synchronization of timing and motion among performing musicians. *Music Perception*, *26*(5), 427–438.

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, *37*(3), 424–438. https://doi.org/10.2307/1912791

Keller, P. E., & Appel, M. (2010). Individual Differences, Auditory Imagery, and the Coordination of Body Movements and Sounds in Musical Ensembles. *Music Perception: An Interdisciplinary Journal*, *28*(1), 27–46. https://doi.org/10.1525/mp.2010.28.1.27

Keller, P. E., Knoblich, G., & Repp, B. H. (2007). Pianists duet better when they play with themselves: On the possible role of action simulation in synchronization. *Consciousness and Cognition*, *16*(1), 102–11. https://doi.org/10.1016/j.concog.2005.12.004

Lähdeoja, O., Wanderley, M. M., & M. Malloch, J. (2009). Instrument augmentation using ancillary gestures for subtle sonic effects. *Proceedings of the Sound and Music Computing Conference*, (July), 327–330.

Loehr, J. D. (2013). Sensory attenuation for jointly produced action effects. *Frontiers in Psychology*, *4*(April), 172. https://doi.org/10.3389/fpsyg.2013.00172

Palmer, C., Koopmans, E., Carter, C., Loehr, J. D., & Wanderley, M. M. (2009). Synchronization of motion and timing in clarinet performance. *International Symposium on Performance Science*, (Davidson 1995), 1–6. Retrieved from http://www.mcgill.ca/files/spl/Palmer-ISPS-final.pdf

Palmer, C., & Loehr, J. D. (2013). Meeting of two minds in duet piano performance. In *Musical Implications* (pp. 323–337).

Papiotis, P., Marchini, M., Perez-Carrillo, A., & Maestre, E. (2014). Measuring ensemble interdependence in a string quartet through analysis of multidimensional performance data. *Frontiers in Psychology*, *5*(AUG). https://doi.org/10.3389/fpsyg.2014.00963

Ragert, M., Schroeder, T., & Keller, P. E. (2013). Knowing too little or too much: The effects of familiarity with a co-performer's part on interpersonal coordination in musical ensembles. *Frontiers in Psychology*, *4*(JUN), 1–15. https://doi.org/10.3389/fpsyg.2013.00368

Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R. L., & Schmidt, R. C. (2007). Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *Human Movement Science*, *26*(6), 867–891. https://doi.org/10.1016/j.humov.2007.07.002

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, *10*(2), 70–76. https://doi.org/10.1016/j.tics.2005.12.009

Seth, A. K. (2010). A MATLAB toolbox for Granger causal connectivity analysis. *Journal of Neuroscience Methods*, *186*(2), 262–273. https://doi.org/10.1016/j.jneumeth.2009.11.020

Sumby, W., & Pollack, I. (1954). Visual contributions to speech intelligibility in noise. *The Journal of the Acoustical Society of America*, *335*, 212–215. https://doi.org/10.1121/1.1907309

Teixeira, E. C. F., Loureiro, M. A., Wanderley, M. M., & Yehia, H. C. (2014). Motion analysis of clarinet performers. *Journal of New Music Research*, *8215*(July), 1–15. https://doi.org/10.1080/09298215.2014.925939

Teixeira, E. C. F., Yehia, H. C., & Loureiro, M. A. (2015). Relating movement recurrence and expressive timing patterns in music performances. *The Journal of the Acoustical Society of America*, *138*(3), EL212-EL216. https://doi.org/10.1121/1.4929621

Verfaille, V., Quek, O., & Wanderley, M. M. (2006). Sonification of musicians' ancillary gestures. *Proceedings of the 12th International Conference on Auditory Display (ICAD 2006)*, 194–197.

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., & Levitin, D. J. (2006). Cross-modal interactions in the perception of musical performance. *Cognition*, *101*(1), 80–113. https://doi.org/10.1016/j.cognition.2005.09.003

Vuoskoski, J. K., Thompson, M. R., Clarke, E. F., & Spence, C. (2014). Crossmodal interactions in the perception of expressivity in musical performance. *Attention, Perception, & Psychophysics*, *76*(2), 591–604. https://doi.org/10.3758/s13414-013-0582-2

Wanderley, M. M. (2002). Quantitative analysis of non-obvious performer gestures. *Gesture and Sign Language in Human-Computer Interaction*, 241–253. https://doi.org/10.1007/3-540-47873-6_26

Wanderley, M. M., Vines, B. W., Middleton, N., McKay, C., & Hatch, W. (2005). The musical significance of clarinetists' ancillary gestures: An exploration of the field. *Journal of New Music Research*, *34*(1), 97–113. https://doi.org/10.1080/09298210500124208

Zamm, A., Pfordresher, P. Q., & Palmer, C. (2014). Temporal coordination in joint music performance: effects of endogenous rhythms and auditory feedback. *Experimental Brain Research*, *233*(2), 607–615. https://doi.org/10.1007/s00221-014-4140-5

CHAPTER 3

RELATIVE CONTRIBUTIONS OF SENSORY INFORMATION TO AUDIENCE SENSITIVITY TO PERFORMER COMMUNICATION ABILITIES

Abstract

Musicians make elaborate movements while performing. Some of these directly influence sound production, but others lack acoustic consequences making them seem unnecessary or even inappropriate. These extraneous movements, however, have been attributed to expressive intentions and communication amongst co-performers (Davidson, 1993; Wanderley, 2002). To explore these movements, we motion-captured and audio-recorded nine pairings of clarinettists and pianists performing Brahms' *Clarinet Sonata No. 1* under four different experimental conditions: (1) normal performance setting, (2) no visual/full auditory feedback, (3) full visual/partial auditory feedback, and (4) no visual/partial auditory feedback. In conditions with a partial auditory manipulation, the clarinettist could not hear the pianist, but the pianist could still hear both parts. Point-light display videos were created from the motion-capture data and sound clips were rendered from the audio recordings. We then examined naïve participants' perception of these musical performances when given either audio-visual, audio-only, or visual-only stimuli. This offered insight into the role of each sensory modality with regards to expression conveyed, cohesion between performers, and overall liking of performances. We found the normal performance setting yielded the highest values across all ratings and all types of sensory stimuli. Additionally, the visual-only and audio-visual stimuli resulted in significant differences in ratings across many of the performer manipulation conditions, while the audio-only experiment resulted in no significant differences. This indicates that musicians' sound outputs were not impaired by the varying abilities to

communicate to their co-performer, but musicians altered their ancillary gestures. Furthermore, viewing musician gestures may play a role in audiences' capability to distinguish features in a musical performance. Our research highlights the importance of live music performances for audience enjoyment and informs music pedagogy practices by stressing the need for intentionally incorporating ancillary gestures.

Introduction

Performing musicians often move around dramatically, sway in synchronization with the music, raise their hands with a flourish, or subtly nod to one another. Are these extraneous movements superfluous or do they influence co-performer communication and/or audience perception? Not all movements made by musicians are mandatory for sound production, so how are these *ancillary* gestures affecting performances (Wanderley, 2002)? Musical performance entails a rich exchange of nonverbal social interactions. Musicians must execute fine motor control under multi-modal stimulation from the auditory, visual, and tactile domains to create a cohesive performance with another musician. Musicians must dynamically adjust their playing in order to compensate for differences in performer timing, timbre, expression, and many other collaborative aspects. Audience members simultaneously receive sound and visual information that they process to create a coherent perception of the performance.

Previous research has examined the circumstances under which either the visual or auditory modality takes precedence in shaping viewer perception (Broughton & Stevens, 2009; Platz & Kopiez, 2012; Schutz, 2008; Thompson, Graham, & Russo, 2005; Vines et al., 2006; Wanderley et al., 2005). One such study, Tsay (2013), presented a provocative idea: both novice and professional musicians are more accurate at selecting the winner of a competition between highly skilled pianists when presented with only visual clips of the musicians rather than with audio alone or both audio and visual presented together. Since highly expert musicians all play extremely well, auditory output would be deemed high quality across pianists. Their movements, however, may be more likely to vary. Therefore, the differences in participant ratings may most strongly reflect the variability of competitors' gestures. This suggests that the amount of variability in the auditory and visual information may be determining which sensory modality is most useful in any given situation.

Vines, Krumhansl, Wanderley, & Levitin (2006) examined the cross-modal interactions of sound and vision during a clarinet performance. Participants were played either auditory, visual, or both auditory and visual recordings from the performances, and were asked to judge the emotional and structural content of the music. The authors found that vision could either strengthen emotional responses when visual information was consistent with the auditory information, or dampen emotional responses when sensory information did not match. They concluded that vision and sound can communicate different emotional information, which are both integrated into overall perceived emotion, creating an emergent experience. Conversely, in the Vines et al. (2006) recordings, vision and sound conveyed similar structural information as indicated by participants' judgements of the phrasing of the musical pieces. Platz and Kopiez (2012) conducted a meta-analysis of the effect audio-visual manipulations have on perceived quality, expressiveness, and preferences for music. Fifteen studies were surveyed, including Vines et al. (2006), and an effect size of d= 0.51 (Cohen's d; 95% CI 0.42, 0.59) was found for influence of the visual component. Given that this is a medium effect size, it suggests that vision is an important aspect of a musical performance.

The importance of visual information in the assessment and perception of musical performances was also examined by Davidson (1993), who suggested that vision plays an even more important role than sound in certain musical conditions. Violinists performed a musical excerpt in three different manners – deadpan, standard, or exaggerated – that varied in degree of expressive exaggeration. Participants rated the performances when presented with audio-visual, audio, or visual recordings, and their ratings showed that differentiating the degree of expressiveness was most accurate with vision alone. When audio was presented alone, participants had difficulty distinguishing between the expressiveness of the performances. Vuoskoski, Thompson, Clarke, and Spence (2014) used a similar design, but also created mismatching audio-visual clips to examine cross-modal interactions. They found that auditory and visual cues both contribute important information to the perception of expressivity in a musical performance, but visual kinematic cues may be more important depending on individual performer's success at communicating through gestures. Furthermore, they observed cross-modal interactions when sensory information could be integrated, indicating auditory expressivity ratings were influenced by visual cues and vice versa. However, extreme mismatched stimuli did not show cross-modal effects. When discussing music, sound is usually the main focus, but these studies demonstrate that vision can play an important role in our perception of sound under certain circumstances.

In the current study, we ran three experiments – audio-visual, audio-only, and visual-only – to examine the influence of auditory and visual information on participants' perception of musical performances. We tried to maintain as much ecological validity as possible when designing our experiments, aside from manipulating auditory and visual feedback during performer recordings. We presented participants with stimuli that preserved musicians' original performance intentions. For instance, we did not cross visual recordings with different audio recordings to create our stimuli; instead we used corresponding audio-visual material. We also balanced our musician pairings with all three clarinettists performing with all three pianists, allowing performers' natural variety in movement to be presented multiple times.

The current study focuses on ancillary gestures when analysing visual cue contributions in a musical performance. Ancillary gestures – gestures that do not directly influence sound production on an instrument – can be thought of as a form of nonverbal communication (Dahl & Friberg, 2007; Wanderley et al., 2005). In speech, people punctuate and emphasize certain aspects of their dialogue through body language, such as hand movements or shrugging the shoulders. Body language used in speech is analogous to ancillary gestures utilized in a musical performance, both of which have the ability to convey additional information to the viewer. The viewer in a musical performance can either be a co-performer or the audience.

Inter-performer communication has been shown to occur through visual information in the form of head movements and body sway, which are both types of ancillary gestures (Badino, D'Ausilio, Glowinski, Camurri, & Fadiga, 2014; Chang, Livingstone, Bosnyak, & Trainor, 2017; Volpe, D'Ausilio, Badino, Camurri, & Fadiga, 2016). Ancillary gestures can also influence audiences in the way they perceive, understand, and interpret a musical piece (Vines et al., 2006). It is apparent that gestures possess expressive content that is pre-attentively registered by audiences (Davidson, 1993). Dahl and Friberg (2007) took videos of musicians expressing different emotions when playing a piece and asked participants to rate the expressive content when presented with different views of a silent video. Viewing conditions varied across videos in the amount of the body visible in the frame. Participants correctly identified the performers' intent of conveying happiness, sadness, and anger in all viewing conditions, suggesting that movement alone is enough to impart intended emotions. Furthermore, other studies have shown that point-light displays, which present only physical movements in the form of stick figure videos, convey enough information to discern emotional intent and other salient features of a musical performance (Davidson, 1993; Schutz & Kubovy, 2009; Sevdalis & Keller, 2011, 2012; Vuoskoski et al., 2014). In the current study we used point-light display videos (see Fig. 1) in the audio-visual and visual-only experiments to analyse the impact of biological motion isolated from the facial expressions, physical appearances, and other noticeable features of performers.

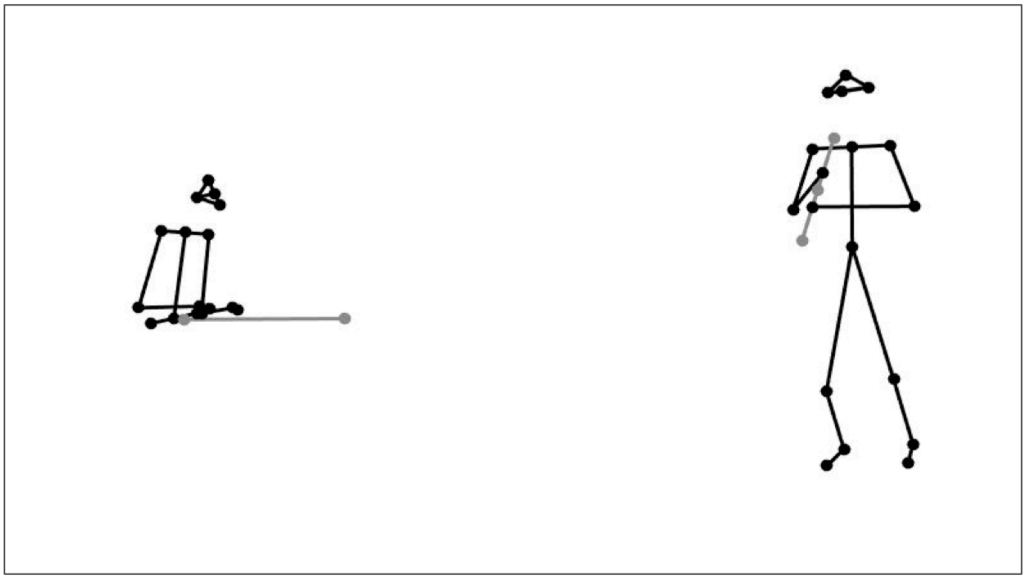


Fig. 1. Screenshot of a point-light display stimuli evaluated by participants. Musicians are shown in black and instruments are shown in grey (two points on the piano and two points on the clarinet).

We independently manipulated the visual and auditory communication between performers. Musicians were either in a normal performance setting, could not see each other, the clarinettist could not hear the pianist, or there was a combination of the latter two conditions in that the musicians could not see each other and the clarinettist could not hear the pianist (but the pianist could always hear everything). We explored the effects of these manipulations in two separate studies. The first examined how the sensory manipulations affected inter-musician communication abilities measured using head movements (Siminoski & Schutz, under review). Here we explore a different question – the degree to which naïve participant evaluations of performances are affected by the different manipulations under which the musicians performed. Additionally, we independently manipulated the type of sensory information participants rated in order to examine the relative contributions of modalities. Participants evaluated performances presented in three different manners across three different experiments: (1) audio-visual, (2) audio-alone, and (3) visual-alone. We instructed them to rate stimuli on how expressive and cohesive they thought the performances were and how much they liked each performance.

We aimed to compare participant ratings of the audio-visual, audio-only, and visual-only stimuli to infer (1) how co-performers changed their audio output and visual communication styles based on the performer manipulations, (2) if naïve participants could detect differences in expressiveness and cohesion based on performer condition, and (3) assess the relative contributions of sight and sound to this ability. We predicted that visual information would play a more important role in all ratings than auditory information based on published literature (Davidson, 1993; Tsay, 2014; Vines et al., 2006; Vuoskoski et al., 2014). Specifically, we predicted that participant ratings of visual-only stimuli would be more varied across conditions than auditory information, indicating that participants were better able to distinguish nuances in performances from visual than auditory information. We predicted that participants would be less able to discern differences in expressiveness and cohesion between performance conditions when listening to music without any visual information in comparison to the conditions where visuals were available. As far as the initial conditions during the performance, we expected the normal performance where the musicians could both see and hear each other to yield higher ratings of expressiveness, cohesion, and likeability than all other conditions, since the limited visual or auditory feedback between performers in the other conditions created fewer means of communication. Additionally, limited sensory feedback may make musicians more focused on co-performer communication than performing expressively, which could lead to lower expression ratings.

Stimuli

The stimuli used were recorded in a previous study (Siminoski & Schutz, under review). In this study, we motion-captured and audio-recorded highly trained professional clarinettists and pianists performing under four different experimental manipulations and examined movement using Granger causality analyses (Siminoski & Schutz, in prep). Here we outline how this performer data was collected (see supplemental material for additional details), and in the subsequent sections present the three experiments conducted from the performer data. All studies met the criteria set by the McMaster University Research Ethics Board.

Materials and Procedure for Stimuli Collection

Clarinettists brought their personal professional model clarinets and the pianists were provided with a Roland FP-80 MIDI keyboard. A directional microphone (AKG C414 XLS) placed in a directional microphone shield recorded the clarinet. We used Reaper software[1] to record both the MIDI output of the piano as

---

[1] http://www.reaper.fm/

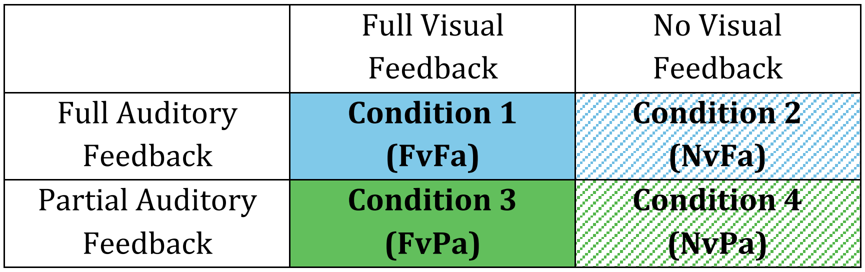
well as the acoustic output of the clarinet. The audio setup allowed for auditory feedback to be adjusted throughout the experiment depending on the condition. A Qualisys motion-capture system[2] recorded participants' movements when performing in the LIVE Lab at McMaster University. The entire body of the clarinettist and the upper half of the pianists' body were motion captured. The clarinet had two markers: one on the bell and another on the barrel. The piano had two markers on either side of the keyboard.

Musicians performed two excerpts from Brahms' *Clarinet Sonata No. 1* in f minor, but for this study we are focusing on the first excerpt (i.e., movement 1, bars 1-28). This composition is from the classical-romantic period and allows performers to add emotive expression and timing fluctuations. Unmarked sheet music was provided for musicians with the editor's musical nuances indicated on the score. Each clarinettist performed with each pianist, forming nine pairings of musicians. On the day of the experiment, each musician first played the excerpt solo three times before playing duets. Each duo performed the excerpt three times under four different conditions. To make our results easier to follow we have used abbreviations alongside condition numbers; condition 1 full vision and full audio (FvFa), condition 2 no vision and full audio (NvFa), condition 3 full vision and partial audio (FvPa), and condition 4 no vision and partial audio (NvPa) (Fig. 2). In condition 1, participants could both hear and see each other like in a normal performance setting. In condition 2, an acoustically transparent screen placed between the musicians blocked visual feedback so musicians could not see their co-

---

[2] http://www.qualisys.com/

performer. In condition 3, the clarinettist could not hear the pianist, but could still hear himself/ herself. The pianist was able to hear both the clarinet and piano. Both musicians could see one another. Condition 4 was a combination of the latter two conditions, no visual feedback and partial auditory feedback. Musicians could not see each other, the clarinettist could not hear the pianist but could still hear himself/ herself, and the pianist was able to hear both instruments. We randomized ordering of the four performance conditions for each duo in this 2 (visual feedback) x 2 (auditory feedback) design. We instructed the musicians to play as if they were performing for an audience. The experiment was conducted over three consecutive days.



	Full Visual Feedback	No Visual Feedback
Full Auditory Feedback	**Condition 1 (FvFa)**	**Condition 2 (NvFa)**
Partial Auditory Feedback	**Condition 3 (FvPa)**	**Condition 4 (NvPa)**

Fig. 2. Summary of the four performance conditions. In conditions involving partial auditory feedback, the clarinettist could not hear the pianist. Colours and shading coordinate with Fig. 3, 4, & 5.

Experiment 1- Audio-Visual

Our overall goal was to assess how naïve participants' evaluations of expression cohesion, and likeability varied as a function of performance condition. Additionally, we aimed to explore the relative contribution of sight and sound to these evaluations by asking participants to assess three representations of the same performances: audio-visual (Experiment 1), audio alone (Experiment 2), and visual alone (Experiment 3).

Method

Participants. Seventy-three undergraduate students (50 female; mean age = 20.0 years, SD = 3.32 years) from McMaster University completed the study for course credit or monetary compensation. Eight participants were excluded from analysis due to incorrectly completing the task.

Stimuli. From the performer data collection, we selected one trial for each condition and pairing, totalling 36 trials (9 trials per condition), based on the best audio quality. We created 36 high-quality audio recordings using Reaper software to audio edit. The motion capture data from the selected trials were cleaned using Qualisys Track Manager software[3]. We used MATLAB (Math Works, Inc.) to create point-light display videos using the Mocap Toolbox (Burger & Toiviainen, 2013) (see Fig. 1). In one of the pairings, a marker from the pianist was consistently missing. Therefore, we eliminated the pairing from the experiment, leaving 32 point-light display videos. The corresponding point-light display videos and audio

---

[3] http://www.qualisys.com/software/qualisys-track-manager/

recordings were combined using iMovie software[4], creating 32 audio-visual clips of about 40 seconds in length. Stimuli were randomized for each participant. The experiment was programmed and run using PsychoPy v1.85.[5]

Evaluation Procedure. Participants watched and listened to audio-visual stimuli in a sound attenuated booth. The audio was presented through Seinnheiser HDA 300 closed-back headphones and video was presented on a MacBook. After each audio-video clip participants rated the performance on *expression*, *cohesion*, and how much they liked the performance (*likability*). We defined expression as how well the musicians conveyed emotion during the performance. For cohesion ratings, participants were instructed to evaluate how well the musicians worked together during the performance. Ratings were measured using a continuous scale from 1 (e.g. not expressive) to 100 (e.g. very expressive). We asked participants to consider the entire performance when assigning ratings. After each expression and cohesion rating, participants indicated their confidence rating on a 7-point Likert scale (1-not confident to 7- very confident). Participants completed ratings for each stimulus in the following order: expression, confidence, cohesion, confidence, and likability. The presentation order of videos was randomized for each participant. Participants completed two practice trials consisting of audio-visual recordings of a different Brahms excerpt before starting the experimental trials.

Results

---

[4] https://www.apple.com/ca/imovie/
[5] http://www.psychopy.org/

We conducted 2 (visual manipulation) x 2 (auditory manipulation) within-subjects ANOVAs to assess participants' sensitivity to performance condition based on audio-visual information (i.e. sound and gestures) for each of expression, cohesion, and likeability ratings. During the original performances, in conditions 1 (FvFa) and 3 (FvPa) musicians could see each other, but in conditions 2 (NvFa) and 4 (NvPa) vision was blocked. In conditions 1 (FvFa) and 2 (NvFa) auditory feedback was intact, but in condition 3 (FvPa) and 4 (NvPa) auditory feedback was only partial.

Expression ratings. The 2 x 2 within-subjects ANOVA on expression ratings showed a main effect of visual manipulation (F(1,64) = 13.4, p < .001, η^2 = .010), a main effect of auditory manipulation (F(1,64) = 11.6, p < 0.01, η^2 = .012), and a significant interaction between visual and auditory manipulations (F(1,64) = 4.73, p = .033, η^2 = .003). Post-hoc multiple comparisons using Bonferroni adjustments showed ratings for FvFa to be significantly higher than all other conditions (NvFa, p = .012; FvPa, p < 0.01; NvPa, p < .001) (Fig. 3). Effect size was calculated using Cohen's *d* and indicated all comparisons had a small effect (*d* = 0.35; *d* = 0.33; *d* = 0.44, respectively). Table 1 in the appendix shows mean expression ratings for each condition.

Cohesion ratings. The 2 x 2 within-subjects ANOVA on cohesion ratings, revealed a main effect of visual manipulation (F(1,64) = 4.61, p = .036, η^2 = .004), a main effect of auditory manipulation (F(1,64) = 19.1, p < 0.001, η^2 = .013), and a significant interaction (F(1,64) = 6.30, p = .015, η^2 = .003). Post-hoc comparisons using Bonferroni adjustments found that ratings for FvFa were significantly higher than for FvPa ($p < 0.01$, $d = 0.36$) and NvPa ($p < 0.01$, $d = 0.36$) (Fig. 3). No other conditions were significantly different from each other. See Appendix Table 1 for means of cohesion ratings for each condition.

Likability ratings. The 2 x 2 within-subjects ANOVA likeability ratings revealed a main effect of auditory manipulation ($F(1,64) = 13.7$, $p < .001$, $\eta^2 = .009$), but no main effect of visual manipulation ($F(1,64) = 1.63$, $p = 0.21$, $\eta^2 = .002$) and no significant interaction ($F(1,64) = 3.46$, $p = .068$, $\eta^2 = .002$). Bonferroni adjusted post-hoc comparisons showed FvFa to have higher likeability ratings than FvPa ($p = .018$, $d = 0.28$) and NvPa ($p = .024$, $d = 0.29$) (Fig. 3). No other conditions were significantly different from each other (see Appendix Table 1 for condition means).

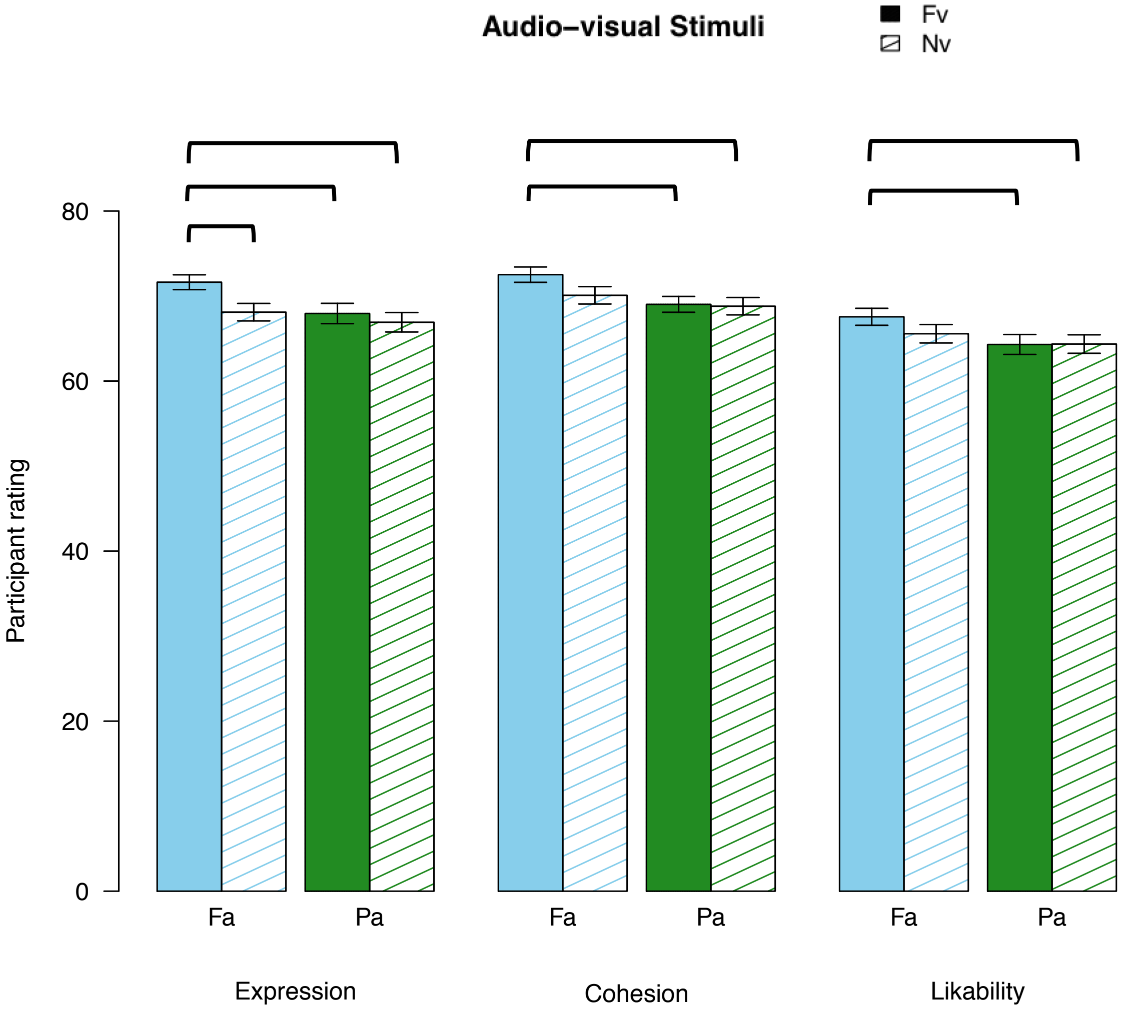


Fig. 3. Participant ratings of audio-visual stimuli. Error bars represent standard error. Brackets indicate a significant difference for multiple comparisons. There was a main effect of performer auditory manipulation for all three rating types and a main effect of performer visual manipulation for expression and cohesion.

Discussion

Consistent with our predictions, participants rated performances where musicians could see and hear each other (condition 1) as most expressive, cohesive, and likeable. When musicians could not see each other and/or the clarinettist could not hear the pianist (condition 4), participants rated these performances as less cohesive, expressive, and likeable.

Whether performers received auditory and visual feedback influenced participant expression and cohesion ratings. An interaction between the two sensory feedbacks for performer conditions was also found to influence participant ratings of expression. Likeability ratings showed a main effect of auditory feedback, meaning participants were only sensitive to whether the musicians could fully hear each other or not when rating how much they liked the performance. Having audio-visual information available to participants led to distinctions between expression, cohesion, and likeability ratings. Expression ratings were most varied, while likeability ratings were more similar across performer conditions. Even though audio-visual stimuli produced both main effects and an interaction, they all had small to very small effect sizes; therefore we recognize that the conclusions may be weak.

Experiment 2- Audio Only

Experiment 2 assessed the auditory component of the audio-visual stimuli used in the first experiment. Participants listened to audio-only recordings of the clarinet and piano players performing a duet using the same procedure and instructions as previously described.

Method

Participants. Seventy-one undergraduate students (49 female) from McMaster University completed the study for course credit. They had a mean age of 19.3 years (SD = 2.92 years). We removed 8 participants from the analysis due to technical difficulties with PsychoPy or participants incorrectly completing task instructions.

Stimuli and evaluation procedure. Participants followed a similar procedure as experiment 1, but participants evaluated the audio component of the 36 audio-visual clips used in experiment 1. Participants completed two practice trials consisting of audio-only recordings of a different Brahms excerpt before starting the experimental trials. The experiment was programmed and run through PsychoPy v1.85 on a MacBook.

Results

Expression ratings. The 2 (visual manipulation) x 2 (auditory manipulation) within-subjects ANOVA on expression ratings revealed no significant main effects of visual manipulation (F(1,62) = .009, p = 0.93, $\eta^2 < .001$), auditory manipulation (F(1,62) = 1.72, p = 0.20, $\eta^2 < .001$), or interaction (F(1,62) = 0.61, p = 0.44, $\eta^2 < .001$) (Fig. 4). Mean expression ratings can be seen in Appendix Table 2.

Cohesion ratings. The 2 x 2 within-subjects ANOVA on cohesion ratings revealed a main effect of auditory manipulation (F(1,62) = 11.5, p < 0.01, η^2 = .008). There was no main effect of visual manipulation (F(1,62) = .061, p = 0.81, η^2 < .001) or interaction (F(1,62) = 1.73, p = 0.19, η^2 < 0.01). We conducted post-hoc comparisons using Bonferroni adjustments, but no significant differences between conditions were found (Fig. 4). See Table 2 in the appendix for mean cohesion ratings for each condition.

Likability ratings. The 2 x 2 within-subjects ANOVA on likability ratings, revealed a main effect of auditory manipulation (F(1,62) = 8.00, p < 0.01, η^2 = .003). No main effect of visual manipulation (F(1,62) = .002, p = 0.89, η^2 < .001) or interaction (F(1,62) = 1.97, p = 0.17, η^2 < .001) were found (Fig. 4). We found no significant differences between conditions with post-hoc multiple comparisons using Bonferroni adjustments (see Appendix Table 2 for condition means).

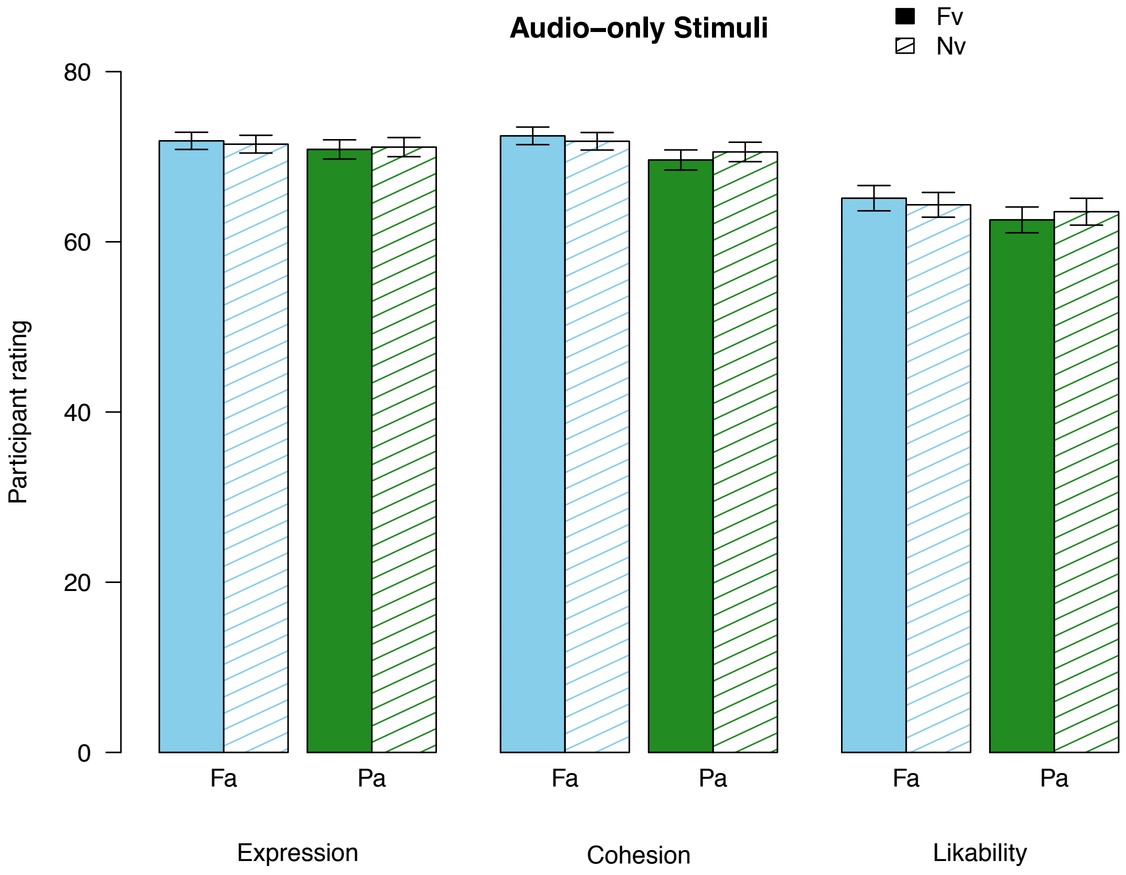


Fig. 4. Participant ratings of audio-only stimuli. Error bars represent standard error. No significant comparisons were found. There was a main effect of performer auditory manipulation for cohesion and likeability.

Discussion

Participants' mean ratings for expression, cohesion, and likeability were all not significantly different from each other across all conditions. This is consistent with our prediction that our performance condition manipulations would not be detected by participants presented with audio-only stimuli. Therefore, we believe differences detected by participants in the audio-visual experiment were not driven by auditory information. Past research has shown that non-musicians and musicians have a difficult time perceiving differences between musical performances based on the auditory component alone (Davidson, 1993; Tsay, 2013; Vines et al., 2011). Experiment 2 provides further support for this concept.

We also observed a main effect of auditory feedback for cohesion and likeability ratings, indicating that participant ratings were influenced by whether or not musicians could fully hear each other during the initial recording sessions. That being said, effect sizes were very small for the main effect of sound in both ratings (cohesion, η^2=0.0082; likeability, η^2=0.0023) and did not lead to significant differences between conditions in post hoc tests. We conclude that our performance manipulations had minimal effect on our performers' acoustic output, and/or it was not detectable by listeners. This provides evidence that differences in musicians' movements between performance conditions were truly ancillary gestures, as they did not lead to differences in participant ratings of the resultant sound when presented independently of the movements used in creating that sound.

Experiment 3- Visual Only

In experiment 3, participants rated point-light display videos of the professional clarinettists and pianists performing together. Videos were presented without sound to clarify participants' ability to differentiate expression, cohesion, and likeability between performer conditions on the basis of movement differences that are by definition ancillary (i.e. they led to no differences in ratings in Experiment 2).

Method

Participants. Sixty-nine undergraduate students (mean age = 18.8 year, SD = 1.12; 58 female) from McMaster University participated in the experiment for course credit. Eight participants were excluded due to technical problems with PsychoPy or incorrectly completing the task.

Stimuli and evaluation procedure. Participants followed a similar procedure as in Experiment 1, but watched only the visual component of the audio-visual stimuli. Before the experimental trials began, participants heard an audio recording of the musical excerpt they were about to watch the point-light display musicians perform. For experiment 3 we defined expression as how well the performers used their body movements to convey emotion. Cohesion was defined as how well performers used their body movements to work together during the performance. Participants were instructed to focus on the movements of performers when rating videos. Participants completed two practice trials consisting of silent point-light display videos of a different Brahms excerpt. The experiment was programmed and run on PyschoPy v1.85.

Results

Expression ratings. The 2 (visual manipulation) x 2 (auditory manipulation) within-subjects ANOVA on expression ratings revealed a main effect of visual manipulation ($F(1,60) = 64.7$, $p < .001$, $\eta^2 = .062$) and a main effect of auditory manipulation ($F(1,60) = 29.4$, $p < .001$, $\eta^2 = .020$), but no significant interaction

(F(1,60) = 2.27, p = 0.14, $\eta^2 = .002$). Post-hoc comparisons were calculated on expression ratings between the four conditions using Bonferroni corrections. Expression ratings were significantly higher for FvFa compared to NvFa ($p < .001$) and NvPa ($p < .001$). Effect size, using Cohen's *d*, indicated a small effect ($d = 0.26$) and a medium effect ($d = 0.50$), respectively. Ratings for NvFa were significantly higher than for NvPa ($p < 0.01$, $d = 0.23$) and ratings for FvPa were significantly higher than for NvPa ($p < .001$, $d = 0.38$) (Fig. 5). Mean expression ratings for each condition are displayed in Appendix Table 3.

Cohesion ratings. The 2 x 2 within-subjects ANOVA on cohesion ratings revealed a main effect of visual manipulation (F(1,60) = 7.51, $p < 0.01$, $\eta^2 = .008$), a main effect of auditory manipulation (F(1,60) = 10.9, $p < 0.01$, $\eta^2 = .014$), but no significant interaction (F(1,60) = 2.46 p = 0.12, $\eta^2 = .004$). Post-hoc comparisons using Bonferroni corrections on cohesion ratings between conditions revealed that ratings for NvPa were significantly lower than for the other three conditions (FvFa, $P < 0.01$; NvFa, $p < 0.01$; FvPa, p = .019). The effect sizes for all three comparisons were small, $d = 0.25$, $d = 0.22$, and $d = 0.19$, respectively (Fig. 5). See Table 3 in the appendix for mean cohesion ratings for each condition.

Likability ratings. The 2 x 2 within-subjects ANOVA on the likability ratings revealed a main effect of visual manipulation (F(1,60) = 51.3, $p < .001$, $\eta^2 = .043$), a main effect of auditory manipulation (F(1,60) = 15.2, $p < .001$, $\eta^2 = .019$, and a significant interaction between visual and auditory manipulations (F(1,60) = 4.68, p = .034, $\eta^2 = .005$). Post-hoc comparisons were conducted on likability ratings between each condition using Bonferroni corrections. Ratings for FvFa were significantly higher than ratings for NvFa (p = .048). Ratings for NvPa were significantly lower than for all other conditions (FvFa, P < .001; NvFa, p < .001; FvPa, p < .001) (Fig. 5). See Appendix Table 3 for mean likeability ratings for each condition.

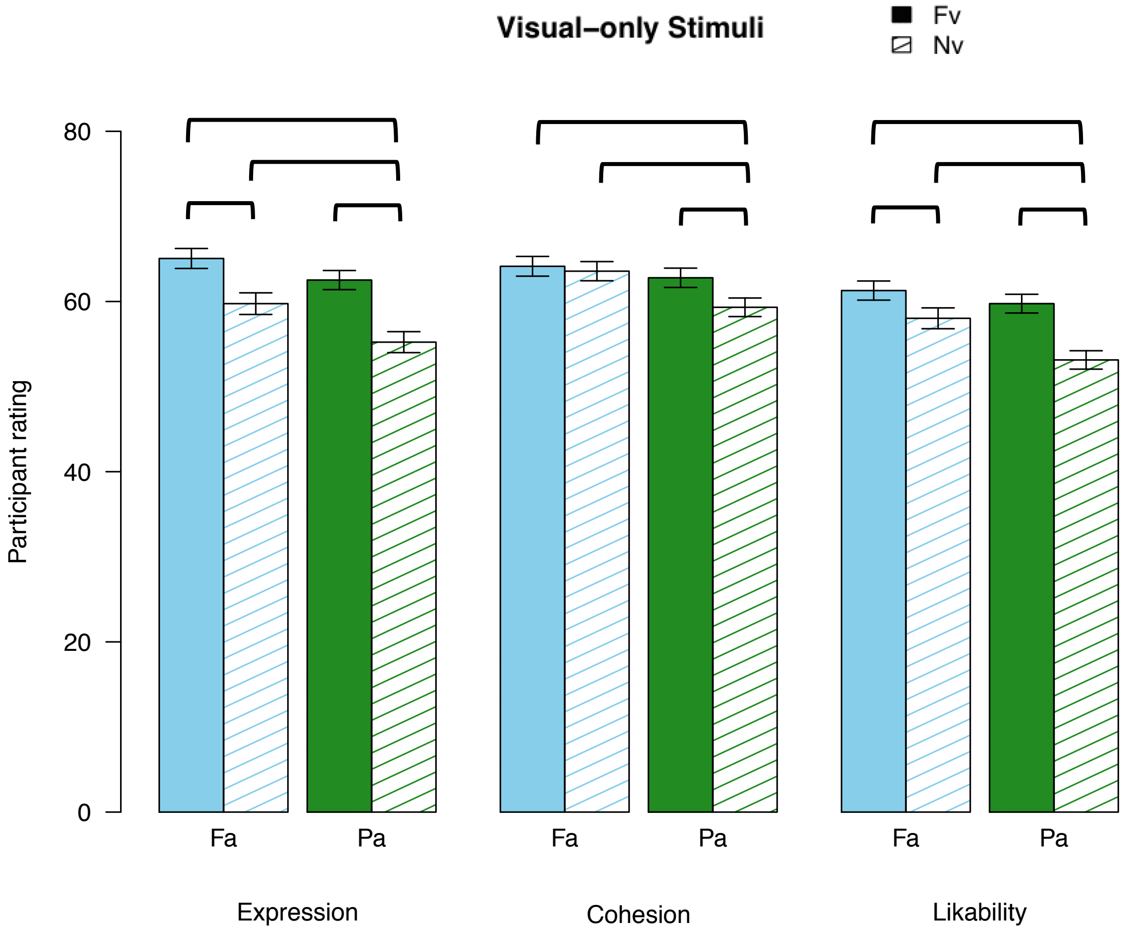


Fig. 5. Participant ratings of visual-only stimuli. Error bars represent standard error. Brackets indicate a significant difference for multiple comparisons. For all three rating types there was a main effect of performer auditory manipulation and a main effect of performer visual manipulation.

Discussion

Participants rated the movements of performances where musicians could see and hear each other (condition 1) as most expressive, cohesive, and likeable. They also rated performances where musicians could not see each other and the clarinettist could not hear the pianist as being least expressive, cohesive, and likeable. This is consistent with our prediction that physical movements alone are sufficient to capture the consequences of our performance manipulations. Experiment 3 shows that vision played a definitive role in participants' sensitivity to performer conditions, and was likely driving responses in Experiment 1. Furthermore, the fact that participant ratings distinguished between performer conditions indicates that musicians moved noticeably different as a result of altering musicians ability to see and/or hear one another. Consistent with previous findings, we found point-light display videos to be an effective method for observing whether participants are influenced by the ancillary gestures of musicians (Dahl & Friberg, 2007; Davidson, 1993).

Both visual feedback and auditory feedback had a main effect on participant ratings for expression, cohesion, and likeability. This indicates that our manipulations of performance conditions (i.e., whether the musicians could both see and hear one another) were detected by participants watching the musicians' movements even in the absence of hearing the performances. Visual feedback had a larger effect size for expression ($\eta^2 = .062$) and likeability ($\eta^2 = .043$) ratings than auditory feedback (expression, $\eta^2 = .020$; likeability, $\eta^2 = .019$). This could mean that having visual information available to musicians during a performance may be more important for expression than having auditory information available. However, it could also be the case that during the no visual feedback condition (condition 2), musicians were more affected by the manipulation since it disrupted sensory information for both performers rather than just one, as occurred during the partial auditory feedback condition. It is also interesting that the effect of the visual manipulation was different depending on the auditory manipulation for likeability ratings. This shows that participants are sensitive to the sensory information available to performers and the subsequent effect this had on performance qualities. When we removed the musicians' visual communication channel, their performance was affected differently depending upon whether or not their auditory communication was impaired or not. Overall, Experiment 3 shows that visual information, specifically ancillary gestures, viewed by audiences can play an important role in the evaluation of a musical performance.

General Discussion

The current studies aimed to examine two levels of communication that exist in a musical performance. The first level dealt with inter-musician communication and the second level examined how musician duos communicate with audience members. To test this, participants rated various presentations of the musical duos with different sensory modalities (i.e. audio-visual, audio-only, or visual-only stimuli). We observed what happened to participant ratings of performances as a result of changing musicians' ability to communicate with their co-performers.

Inter-musician Communication

To examine inter-musician communication, we observed whether musicians' gestures and audio output changed in order to communicate with one another given the various experimental manipulations. We found no evidence that removing musicians' ability to see one another impaired their capacity to produce cohesive audio signals. Participants rated audio-only stimuli consistently across all performer conditions, so ratings were not affected by restricting performers' ability to communicate through gesture. We also found no evidence that removing clarinettists' ability to hear the pianist affected the musicians' combined audio output. Specifically, participants' ratings of audio stimuli did not change across performer conditions, including the limited auditory communication conditions. Nevertheless, auditory and visual feedback between performers did play an important musical role. The results showed main effects of restricting visual and auditory communication between performers when participants experienced the audio-visual and visual stimuli.

We also found evidence that musicians' ancillary gestures change depending on whether the performers can see each other or not. Participants rated performer conditions differently when presented with the audio-visual and visual-only stimuli, but did not distinguish conditions with audio-only stimuli (see Table 1). The differences in participant ratings can be attributed to ancillary gestures changing between performer conditions since performers kept audio output consistent. The musicians modulated movements based on their ability to visually interact, which is consistent with the idea that ancillary gestures serve a communicative purpose. Past studies have suggested that ancillary gestures are used mainly for expressive purposes (Teixeira et al., 2014, 2015). In contrast, we found that ancillary gestures play a part in communication separate from their expressive role.

Table 1.

Summary of Results from Experiments 1, 2, and 3

Experiment	Evaluation Type					
	Expressivity		Cohesiveness		Likability	
1 (audio-visual)	1-2	A		A		A
	1-3	V	1-3	V	1-3	
	1-4	I	1-4	I	1-4	
2 (audio-only)	NS	NS	NS	A	NS	A
3 (visual-only)	1-2	A		A	1-2	A
	1-4	V	1-4	V	1-4	V
	2-4		2-4		2-4	I
	3-4		3-4		3-4	

Note. Number pairs (e.g. *x-y*) represent significant comparisons between conditions. NS = no significance. A = main effect of performer auditory feedback, V = main effect of performer visual feedback, and I = interaction between performer auditory and visual feedback.

Musician to Audience Communication

To observe musician-to-audience communication we examined how musicians' ability to communicate amongst themselves affected participant ratings. When performers played under normal performance settings (full vision, full audio) participants consistently rated musicians as most expressive, cohesive and likeable regardless of the sensory information available to participants. When performers could not see each other and the clarinettist could not hear the pianist (no vision, partial audio), participants rated musicians as least expressive, cohesive, and likeable, across all experiments. This illustrates that the manipulations to performance conditions affected musicians in ways detectable to even musically untrained listeners.

We found participants to be more sensitive to performer manipulations when presented with visual-only stimuli compared to audio-only and audio-visual stimuli. The visual-only experiment yielded the most differentiation between conditions, while the audio-only stimuli yielded no differences between conditions. This was true for all ratings – expression, cohesion, and likeability – indicating that visual information may allow for better discernment of musical differences than auditory information alone. This was also shown through differences in effect sizes between the visual-only experiment and the audio-only experiment. The main effect of obscuring performers' vision and the main effect of masking the performers' ability to communicate through sound were both highest for the visual-only stimuli, compared to the audio-visual and audio-only experiments. These results follow Davidson's (1993) findings where participants could differentiate performer mannerisms best through visual-only stimuli, and had the hardest time deciphering mannerisms with audio alone.

Measuring expressivity, Vuoskoski, Thompson, Clarke, & Spence (2014) also found visual kinematic cues to contribute more substantially to participant ratings than auditory information. Vuoskoski et al. (2014) created their stimuli using performances of two solo pianists whose natural performance movements varied greatly in style and magnitude. We expanded upon their approach of using two solo musicians by recording nine balanced pairings of three clarinettists and three pianists. The intent was that performer-dependent movement information would be repeated in different musician pairings, so that potentially unique performer movements would be rated multiple times by participants. This design helped control for performer-dependent gestures that might otherwise limit the generalizability of these findings.

The use of point-light displays in the present study allowed us to conclude that participants can detect the effects of the performance conditions on the musicians by watching musicians' body movements regardless of whether they heard the performances. Point-light displays have been used in many experiments to study body movements since they isolate ancillary gestures from other visual influencers such as physical appearance, facial expressions, and lighting cues (Davidson, 1993; Sevdalis & Keller, 2011; Vines et al., 2006; Wanderley et al., 2005). Our study complements this field of research, and confirms that point-light displays are a valuable tool for separating visual kinematic cues from the entirety of musical performances.

Differences Between Sensory Stimuli

Another interesting outcome of our study is that mean ratings for expression, cohesion, and likeability were consistently lower for visual-only stimuli compared to audio-only and audio-visual experiments. This is consistent with Vines et al. (2011) who attributed lower ratings to novelty of stimuli. Participants are familiar with listening to music alone, but are not familiar with watching point-light display videos. The audio-visual stimuli contained point-light displays, but the concept of computer-generated figures moving to sound is somewhat familiar. Therefore the most novel experiment was the visual-alone experiment where stick figures moved in the absence of sound. It is possible that familiarity with stimuli types resulted in more enjoyment in general when sound was present, leading to higher expression, cohesion, and likeability ratings. Vines et al. (2006) also found that visual information strengthens overall expressiveness of performances when musician gestures correspond to the emotion of the auditory component. Our results were consistent with that, as we found higher mean ratings for audio-visual stimuli compared to visual-only stimuli. Vuoskoski et al. (2014) attributed higher participant ratings to cross-modal interactions when visual and auditory information could be integrated in a meaningful way. In our audio-visual experiment, participants should have been able to integrate the information, theoretically leading to cross-modal interactions that led to increased ratings.

Future Investigations

These studies have some limitations that should be considered when interpreting the results. We did not fully balance the performer manipulations due to the nature of the instruments. The visual feedback was balanced in that both performers could either see each other or not, but the auditory feedback was not even. In conditions with partial auditory feedback, the clarinettist could not hear the pianist, but the pianist could always hear both parts. The condition where the pianist could not hear the clarinettist was not included in the protocol as it is hard to mute an acoustic clarinet. Although an electric clarinet that could be silenced might have been used, we wanted to keep our experiment as ecologically valid as possible.

We chose to use a clarinet and piano piece in this study in order to examine how communication abilities between a soloist (the clarinettist) and collaborator (the pianist) were affected by the manipulations, and how audience perception was changed as a result, as data on this type of musical ensemble dynamic is limited in the joint action literature. Future studies could balance performer roles using piano duets, as electronic pianos are easily muted. Goebl and Palmer (2009) used piano duets to examine the role auditory feedback has on synchronization of musical parts. Pianists either heard both parts, the assigned leader heard only themself while the follower heard both parts, or both pianists only heard themselves. The authors found reduced auditory feedback led to decreased auditory output synchronization, but increased head movement synchronization between piano players. Given these findings, we could gain clarity on the current study results if we had fully balanced audio manipulations.

Another interesting avenue of investigation would be testing trained musicians as audience participants using the same experimental paradigm. In the current study, participants on average had low levels of musical training. Musicians may have more fine-tuned perception of expression and cohesion, especially clarinet and piano players. Previous research with similar paradigms have found comparable emotion ratings between non-musicians and musicians, so our expression ratings may be similar regardless of musical training (Vines et al., 2011).

Vines et al. (2011) did not directly measure cohesion, and it is possible that trained musicians recognize what movements are communicative in purpose and provide different ratings than non-trained participants.

Conclusions

In conclusion, this study demonstrated that visual information is an important aspect of musical performance in a solo instrument-accompanist setting, both for inter-performer communication and communication to the audience. We found that musicians changed their ancillary gestures depending on the sensory information available to them, and without noticeable changes to their audio output. We have attributed changing ancillary gestures to the need for musicians to communicate with their co-performers when sensory feedback is obscured. Ancillary gestures can communicate novel information that increases an audience's sensitivity to performer expression and cohesion. Visual information may be more important than auditory information when audiences are asked to indicate distinctions between performances. Our findings strongly suggest that live music performances, where performers interact with one another and with the audience, may be more enjoyable for an audience than recordings. Live audiences are able to see and hear musicians, which adds to overall enjoyment through increased perception of expression and cohesion. Our findings also inform music pedagogy practices. Music students should be taught how to properly implement ancillary gestures in order to create the most expressive and cohesive performances possible.

Acknowledgements

We would like to thank our research assistants, Maxwell Ng, Shanthika Ravindran, Nicole Divincenzo, Brannon Senger, Rachel Heo, Max Maglal-Lan, and Veronica Panchyshyn, for helping with this project. Also, thank you to the LIVE Lab staff, Dave Thompson, Dan Bosnyak, and J.J. Booth, for helping with the technology. This research was financially supported through grants to Dr. Michael Schutz from the Natural Sciences and Engineering Research Council of Canada (NSERC RGPIN/386603-2010), Ontario Early Researcher Award (ER10-07-195), and the Canadian Foundation for Innovation (CFI-LOF-30101).

Supplemental Material

Stimuli Details

Performer Participants. Three professional clarinettists (1 female) and three professional pianists (2 female) from the Greater Toronto Area participated in the study for monetary compensation. Clarinettists (mean age = 51.3 years, SD = 16.1) averaged 39.0 years (SD = 14.9) performance experience and 12.0 years (SD = 2.00) of lessons. They spent an average of 17.7 hours (SD = 4.93) a week playing the clarinet. Pianists (mean age = 41.3 years, SD = 3.06) played the piano for an average of 35.0 years (SD = 1.00), and averaged 31.0 years (SD = 3.61) of lessons. They played the piano for an average of 14.3 hours (SD = 8.14) a week. All musicians were highly trained, regularly performed in a number of ensembles, and most taught at a University level. All musicians reported normal hearing and right-handedness.

Materials. Clarinettists brought their personal professional model clarinets and the pianists were provided with a Roland FP-80 MIDI keyboard. A directional microphone (AKG C414 XLS) placed in a directional microphone shield recorded the clarinet to a computer running Reaper software.[6] We used Reaper to record both the MIDI output of the piano as well as the acoustic output of the clarinet. The pianist wore NADA QH560 open-back headphones allowing them to hear both the direct sound of the clarinet as well as synthesis of their own MIDI data. The clarinettists wore earplugs (noise reduction rating of 32 dB) along with Seinnheiser HDA 200 closed-back headphones. This allowed for manipulation of whether they heard the pianists' performance, while always allowing them to hear themselves at a

---

[6] http://www.reaper.fm/

reasonable level from their sound piped back through the headphones The audio setup allowed for auditory feedback to be adjusted throughout the experiment depending on the condition.

A Qualisys motion-capture system[7] recorded participants' movements when performing in the LIVE Lab at McMaster University. Clarinet players wore 18 reflective markers to allow full body movements to be captured. Markers were placed bilaterally at the ankle, knee, hip, shoulder, elbow, and wrist; one marker was placed centrally on the nape of the neck; and a solid cap was worn containing four markers: one on top of the head, one centrally on the forehead, and two on the temples. The clarinet had two markers: one on the bell and another on the barrel. Piano players wore 14 reflective markers to capture the movements of the upper half of the body: bilaterally on the hip, shoulder, elbow, and wrist; one centrally on the nape of the neck; and a cap with four markers: one on top of the head, one centrally on the forehead, and two on the temples. The piano had two markers on either side of the keyboard. 18 Qualisys cameras recorded the infrared signals reflected off the markers.

Musicians performed two excerpts from Brahms' *Clarinet Sonata No. 1* in f minor, but for this study we are focusing on the first excerpt (i.e., movement 1, bars 1-28). This composition is from the classical-romantic period and allows performers to add emotive expression and timing fluctuations. Unmarked sheet music was provided for musicians with the editor's musical nuances indicated on the score.

---

[7] http://www.qualisys.com/

References

Badino, L., D'Ausilio, A., Glowinski, D., Camurri, A., & Fadiga, L. (2014). Sensorimotor communication in professional quartets. *Neuropsychologia*, *55*(1), 98–104. https://doi.org/10.1016/j.neuropsychologia.2013.11.012

Barnett, L., & Seth, A. K. (2014). The MVGC multivariate Granger causality toolbox: A new approach to Granger-causal inference. *Journal of Neuroscience Methods*, *223*, 50–68. https://doi.org/10.1016/j.jneumeth.2013.10.018

Barrett, A. B., Barnett, L., & Seth, A. K. (2010). Multivariate Granger causality and generalized variance. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, *81*(4). https://doi.org/10.1103/PhysRevE.81.041907

Broughton, M., & Stevens, C. (2009). Music, movement and marimba: An investigation of the role of movement and gesture in communicating musical expression to an audience. *Psychology of Music*, *37*(2), 137–153. https://doi.org/10.1177/0305735608094511

Castellano, G., Mortillaro, M., Camurri, A., & Volpe, G. (2008). Automated analysis of body movement in emotionally expressive piano performances. *Music Perception*, *26*(2), 103–119.

Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, *50*(1), 62–81. https://doi.org/10.1016/j.jml.2003.08.004

D'Ausilio, A., Badino, L., Li, Y., Tokay, S., Craighero, L., Canto, R., … Fadiga, L. (2012a). Communication in orchestra playing as measured with granger causality. In *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and*

Telecommunications Engineering (Vol. 78 LNICST, pp. 273–275). https://doi.org/10.1007/978-3-642-30214-5_37

D'Ausilio, A., Badino, L., Li, Y., Tokay, S., Craighero, L., Canto, R., … Fadiga, L. (2012b). Leadership in orchestra emerges from the causal relationships of movement kinematics. *PLoS ONE*, *7*(5). https://doi.org/10.1371/journal.pone.0035757

D'Ausilio, A., Novembre, G., Fadiga, L., & Keller, P. E. (2015). What can music tell us about social interaction? *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2015.01.005

Dahl, S., & Friberg, A. (2007). Visual perception of expressiveness in musicians' body movements. *Music Perception*, *24*(4), 433–454. https://doi.org/10.1525/rep.2008.104.1.92.This

Davidson, J. W. (1993). Visual Perception of Performance Manner in the Movements of Solo Musicians. In *Psychology of Music* (Vol. 21, pp. 103–113). https://doi.org/10.1177/030573569302100201

Davidson, J. W. (2012). Bodily movement and facial actions in expressive musical performance by solo and duo instrumentalists: Two distinctive case studies. *Psychology of Music*, *40*(5), 595–633. https://doi.org/10.1177/0305735612449896

Flanagan, J. R., & Johansson, R. S. (2003). Action plans used in action observation. *Nature*, *424*(6950), 769–771. https://doi.org/10.1038/nature01861

Glowinski, D., & Badino, L. (2012). Analysis of leadership in a string quartet. *… Workshop on Social …*, (January). Retrieved from http://www.researchgate.net/publication/232237802_Analysis_of_Leadership _in_a_String_Quartet/file/9fcfd50ae49023179e.pdf

Goebl, W., & Palmer, C. (2009). Synchronization of timing and motion among performing musicians. *Music Perception*, *26*(5), 427–438.

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, *37*(3), 424–438. https://doi.org/10.2307/1912791

Keller, P. E., & Appel, M. (2010). Individual Differences, Auditory Imagery, and the Coordination of Body Movements and Sounds in Musical Ensembles. *Music Perception: An Interdisciplinary Journal*, *28*(1), 27–46. https://doi.org/10.1525/mp.2010.28.1.27

Keller, P. E., Knoblich, G., & Repp, B. H. (2007). Pianists duet better when they play with themselves: On the possible role of action simulation in synchronization. *Consciousness and Cognition*, *16*(1), 102–11. https://doi.org/10.1016/j.concog.2005.12.004

Lähdeoja, O., Wanderley, M. M., & M. Malloch, J. (2009). Instrument augmentation using ancillary gestures for subtle sonic effects. *Proceedings of the Sound and Music Computing Conference*, (July), 327–330.

Loehr, J. D. (2013). Sensory attenuation for jointly produced action effects. *Frontiers in Psychology*, *4*(April), 172. https://doi.org/10.3389/fpsyg.2013.00172

Palmer, C., Koopmans, E., Carter, C., Loehr, J. D., & Wanderley, M. M. (2009). Synchronization of motion and timing in clarinet performance. *International Symposium on Performance Science*, (Davidson 1995), 1–6. Retrieved from http://www.mcgill.ca/files/spl/Palmer-ISPS-final.pdf

Palmer, C., & Loehr, J. D. (2013). Meeting of two minds in duet piano performance. In *Musical Implications* (pp. 323–337).

Papiotis, P., Marchini, M., Perez-Carrillo, A., & Maestre, E. (2014). Measuring ensemble interdependence in a string quartet through analysis of multidimensional performance data. *Frontiers in Psychology*, *5*(AUG). https://doi.org/10.3389/fpsyg.2014.00963

Platz, F., & Kopiez, R. (2012). What the eye listens: A meta-analysis of how audio-visual presentation enhances the appreciation of music performance. *Music Perception: An Interdisciplinary Journal*, *30*(1), 71–83. https://doi.org/10.1525/rep.2008.104.1.92.

Ragert, M., Schroeder, T., & Keller, P. E. (2013). Knowing too little or too much: The effects of familiarity with a co-performer's part on interpersonal coordination in musical ensembles. *Frontiers in Psychology*, *4*(JUN), 1–15. https://doi.org/10.3389/fpsyg.2013.00368

Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R. L., & Schmidt, R. C. (2007). Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *Human Movement Science*, *26*(6), 867–891. https://doi.org/10.1016/j.humov.2007.07.002

Schutz, M. (2008). Seeing music? What musicians need to know about vision. *Empirical Musicology Review*, *3*(3), 83–108.

Schutz, M., & Kubovy, M. (2009). Deconstructing a musical illusion: Point-light representations capture salient properties of impact motions. *Canadian Acoustics*, *37*(1), 23–28.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, *10*(2), 70–76. https://doi.org/10.1016/j.tics.2005.12.009

Seth, A. K. (2010). A MATLAB toolbox for Granger causal connectivity analysis. *Journal of Neuroscience Methods*, *186*(2), 262–273. https://doi.org/10.1016/j.jneumeth.2009.11.020

Sevdalis, V., & Keller, P. E. (2011). Perceiving performer identity and intended expression intensity in point-light displays of dance. *Psychological Research*, *75*(5), 423–434.

Sevdalis, V., & Keller, P. E. (2012). Perceiving bodies in motion: Expression intensity, empathy, and experience. *Experimental Brain Research*, *222*(4), 447–453. https://doi.org/10.1007/s00221-012-3229-y

Sumby, W., & Pollack, I. (1954). Visual contributions to speech intelligibility in noise. *The Journal of the Acoustical Society of America*, *335*, 212–215. https://doi.org/10.1121/1.1907309

Teixeira, E. C. F., Loureiro, M. A., Wanderley, M. M., & Yehia, H. C. (2014). Motion analysis of clarinet performers. *Journal of New Music Research*, *8215*(July), 1–15. https://doi.org/10.1080/09298215.2014.925939

Teixeira, E. C. F., Yehia, H. C., & Loureiro, M. A. (2015). Relating movement recurrence and expressive timing patterns in music performances. *The Journal of the Acoustical Society of America*, *138*(3), EL212-EL216. https://doi.org/10.1121/1.4929621

Thompson, W. F., Graham, P., & Russo, F. A. (2005). Seeing music performance:

Visual influences on perception and experience. *Semiotica*, *2005*(156), 203–227. https://doi.org/10.1515/semi.2005.2005.156.203

Tsay, C.-J. (2013). Sight over sound in the judgment of music performance. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(36), 14580–5. https://doi.org/10.1073/pnas.1221454110

Tsay, C.-J. (2014). The vision heuristic: Judging music ensembles by sight alone. *Organizational Behavior and Human Decision Processes*, *124*(1), 24–33. https://doi.org/10.1016/j.obhdp.2013.10.003

Verfaille, V., Quek, O., & Wanderley, M. M. (2006). Sonification of musicians' ancillary gestures. *Proceedings of the 12th International Conference on Auditory Display (ICAD 2006)*, 194–197.

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., Dalca, I. M., & Levitin, D. J. (2011). Music to my eyes: Cross-modal interactions in the perception of emotions in musical performance. *Cognition*, *118*(2), 157–170. https://doi.org/10.1016/j.cognition.2010.11.010

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., & Levitin, D. J. (2006). Cross-modal interactions in the perception of musical performance. *Cognition*, *101*(1), 80–113. https://doi.org/10.1016/j.cognition.2005.09.003

Volpe, G., D'Ausilio, A., Badino, L., Camurri, A., & Fadiga, L. (2016). Measuring social interaction in music ensembles. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1693). https://doi.org/10.1098/rstb.2015.0377

Vuoskoski, J. K., Thompson, M. R., Clarke, E. F., & Spence, C. (2014). Crossmodal interactions in the perception of expressivity in musical performance. *Attention,* *Perception, & Psychophysics*, *76*(2), 591–604. https://doi.org/10.3758/s13414-013-0582-2

Wanderley, M. M. (2002). Quantitative analysis of non-obvious performer gestures. *Gesture and Sign Language in Human-Computer Interaction*, 241–253. https://doi.org/10.1007/3-540-47873-6_26

Wanderley, M. M., Vines, B. W., Middleton, N., McKay, C., & Hatch, W. (2005). The musical significance of clarinetists' ancillary gestures: An exploration of the field. *Journal of New Music Research*, *34*(1), 97–113. https://doi.org/10.1080/09298210500124208

Zamm, A., Pfordresher, P. Q., & Palmer, C. (2014). Temporal coordination in joint music performance: effects of endogenous rhythms and auditory feedback. *Experimental Brain Research*, *233*(2), 607–615. https://doi.org/10.1007/s00221-014-4140-5

Appendix

Table 1.

Evaluation Type	Description	M	SD
Expression	FvFa	71.63	9.30
	NvFa	68.10	10.93
	FvPa	67.95	12.70
	NvPa	66.92	12.16
Cohesion	FvFa	72.52	9.63
	NvFa	70.08	10.87
	FvPa	69.02	9.91
	NvPa	68.81	10.87
Likeability	FvFa	67.56	10.63
	NvFa	65.56	11.51
	FvPa	64.30	12.46
	NvPa	64.35	11.59

Descriptive Statistics for Audio-visual Stimuli

Note. n = 65

Table 2.

Evaluation Type	Description	M	SD
Expression	FvFa	71.87	10.57
	NvFa	71.28	10.91
	FvPa	70.86	11.81
	NvPa	71.13	11.77
Cohesion	FvFa	72.45	10.82
	NvFa	71.82	10.78
	FvPa	69.62	12.38
	NvPa	70.56	11.97
Likeability	FvFa	65.13	15.58
	NvFa	64.35	15.23
	FvPa	62.58	15.88
	NvPa	63.54	16.53

Descriptive Statistics for Audio-only Stimuli

Note. n = 63

Table 3.

Evaluation Type	Description	M	SD
Expression	FvFa	65.06	12.09
	NvFa	59.75	13.12
	FvPa	62.52	11.59
	NvPa	55.22	12.71
Cohesion	FvFa	64.14	11.96
	NvFa	63.56	11.63
	FvPa	62.79	11.71
	NvPa	59.32	11.25
Likeability	FvFa	61.28	11.58
	NvFa	58.03	12.62
	FvPa	59.75	11.36
	NvPa	53.13	11.16

Descriptive Statistics for Visual-only Stimuli

Note. n = 61