HUMAN TECHNOLOGY

ISSN: 1795-6889

www.humantechnology.jyu.fi

Volume 16(3), November 2020, 257–282

COMMUNICATING THROUGH ANCILLARY GESTURES: EXPLORING EFFECTS ON COPERFORMERS AND AUDIENCES

Anna Siminoski
Department of Psychology, Neuroscience & Behaviour
McMaster University
Hamilton, ON
Canada

Erica Huynh
Schulich School of Music
McGill University
Montréal, QC
and
Department of Psychology, Neuroscience & Behaviour
McMaster University
Hamilton, ON
Canada

Michael Schutz
School of the Arts and
Department of Psychology, Neuroscience & Behaviour
McMaster University
Hamilton, ON
Canada

Abstract: *Musicians make elaborate movements while performing, often using gestures that might seem extraneous. To explore these movements, we motion-captured and audio-recorded different pairings of clarinetists and pianists performing Brahms' Clarinet Sonata No. 1 with two manipulations: (a) allowing the performers full vs. no visual feedback, and (b) allowing the performers full vs. partial auditory feedback (i.e., the clarinetist could not hear the pianist). We found that observer ratings of audio–visual point-light renditions discriminated between manipulations and refined this insight through subsequent audio-alone and visual-alone experiments, providing an understanding of each modality's contribution. This novel approach of evaluating point-light displays of performances under systematically manipulated conditions provides new perspective on the ways in which ancillary gestures contribute to both performer communication and audience reception of live performances.*

Keywords: *music performance, ancillary gestures, expression, performer cohesion, audio–visual, point-light displays.*

DOI: https://doi.org/10.17011/ht/urn.202011256765

INTRODUCTION

Performing musicians often move dramatically: swaying in synchronization with the music, raising their hands with a flourish, or subtly nodding to one another. Are these extraneous movements superfluous or do they influence coperformer communication 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 multimodal stimulation from the auditory, visual, and tactile domains to create a cohesive performance with another musician. Musicians must adjust their playing dynamically 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.

Ancillary Gestures Shape Audience "Listening"

Previous researchers have examined which domain—visual or auditory—takes precedence when shaping viewer perception in different situations or conditions (Broughton & Stevens, 2009; Platz & Kopiez, 2012; Schutz, 2008; Thompson, Graham, & Russo, 2005; Vines, Krumhansl, Wanderley, & Levitin, 2006; Wanderley, Vines, Middleton, McKay, & Hatch, 2005). In one such study, Vines et al. (2006) investigated the cross-modal interactions of sound and vision during clarinet performances. Research participants were played 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 of the research participants 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 could communicate different emotional information, but both are integrated into overall perceived emotion, thus creating an emergent experience. Conversely, Vines et al. (2006) also found that vision and sound conveyed similar structural information as indicated by participants' judgments of the phrasing. 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 the 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.

Additional support for the influence that vision can have on an audience's evaluation of musical performance was presented by Tsay (2013): Both novice and professional musicians were 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. Because highly expert musicians play extremely well, auditory output would be similar across pianists. Their movements, however, were more likely to vary. Therefore, the differences in participant ratings may reflect most strongly the variability of competitors' gestures. Mehr, Scannell, and Winner (2018) expanded on the Tsay (2013) study and tested the generalizability of their findings. When using the exact stimuli from Tsay (2013), the results were replicable. However, when Mehr et al. (2018) used other stimuli, even video clips that presented a greater distinction of skill, the sight-over-sound conclusions did not hold and selecting a piano competition winner with visuals were at or below chance. This research suggests that the amount of variability and information in the auditory and visual modalities determines which is most useful in any given situation. This confusion regarding the role of visual information in evaluating musical performances illustrates challenges with understanding gestures' role in naturalistic musical performances. Additionally, it shows the value of exploring new approaches to manipulating performance conditions to clarify the precise contributions of sight and sound to musical evaluations.

Interest in the importance of visual information in the assessment and perception of musical performances is longstanding for both practical and theoretical reasons. For example, Davidson (1993) suggested that vision plays an even more important role than sound in certain musical conditions. In that study, violinists performed a musical excerpt in three different manners that varied in degree of expression—deadpan, standard, or exaggerated. Study 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 stimuli 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, but extreme mismatched stimuli did not show cross-modal effects. When discussing music, sound is usually the main focus. However, these studies demonstrate the importance of considering vision.

In the current study, we ran three experiments that used either audio–visual (Experiment 1), audio-only (Experiment 2), and visual-only (Experiment 3) stimuli to examine the influence of auditory and visual information on study 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 the musicians' original performances. For instance, we did not cross visual recordings with different audio recordings to create our stimuli; rather, we used corresponding audio–visual material. We also balanced our musician pairings, having three clarinetists perform with each of three pianists. This allowed for performers' individual variety and magnitude of movements to be presented multiple times within unique performer pairs.

Ancillary Gestures Shape Intermusician Communication and Expression

Movement in musical performances can be categorized for the purpose they serve: to produce sound, to coordinate an ensemble, or to present expressive ideas (Bishop & Goebl, 2018). The latter two categories can be grouped under the term ancillary gestures. 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 current study focuses on the communicative and expressive quality of ancillary gestures when analyzing visual-cue contributions in a musical performance. This complements previous research on the degree to which performers' visual communication affects the precision of their synchronization (D'Amario, Daffern, & Bailes, 2018) by exploring whether this communication can shape audience evaluation of their movements and sound.

Communication between performers has been shown to occur through visual information in the form of head movements and body sway, both of which are 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 also can 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 intrinsically recognized by audiences (Davidson, 1993). Dahl and Friberg (2007) took videos of musicians expressing various 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 in the audio–visual and visual-only experiments to analyze the impact of biological motion isolated from the facial expressions, physical appearances, and other noticeable features of performers (Figure 1).

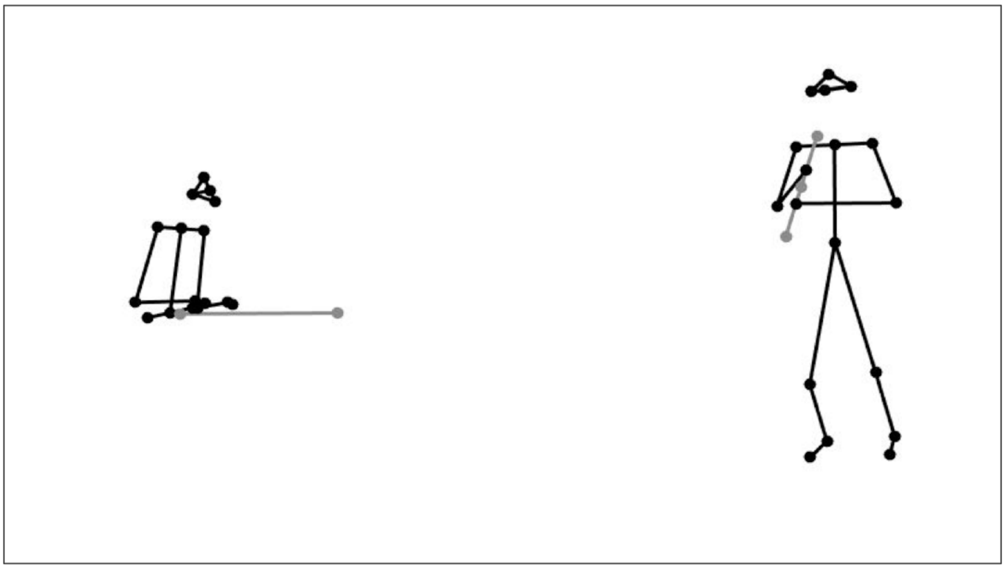

Figure 1. Screenshot of a point-light display video. Representations show the torso of the pianist and the full clarinetist in black. The clarinet appears in grey, and two points from the keyboard connected with a line provide a visual reference point for the pianist's instrument.

In this paper, we present three separate experiments that investigated audience perception of a musical performance when experiencing different types of sensory information. The three experiments were based on manipulated sensory feedback available to the performers during the recorded initial musical performances to generate our stimuli to be used in the experiments. To achieve that, we motion-captured and audio-recorded professional clarinetists and pianists performing a duet under distinct experimental conditions to see how reduced sensory feedback influenced interperformer communication abilities, as rated later by naïve participants. Musicians performed under four different conditions: (a) full auditory and visual feedback between performers; (b) no visual but full auditory feedback between performers; (c) full visual but partial auditory feedback; and (d) no visual and partial auditory feedback. In conditions (c) and (d) with partial auditory feedback, the pianist could hear both instruments, whereas the clarinetist could hear only themselves. We created point-light display videos and audio recordings from the musician data. Then, in experiments to examine research participant sensitivity to the performer manipulations, participants were exposed to audio–visual, auditory-only, or visual-only stimuli and were instructed to rate each set of stimuli on how expressive and cohesive they thought the performances were and how much they liked each performance. Participants were not informed of the conditions under which the musicians were initially recorded.

Our study had two principal aims. First, we were interested in whether participant ratings could distinguish between the different conditions under which the musicians performed—whether the musicians could fully see and hear one another. Second, we were interested in how differences in these ratings varied as a function of stimuli (i.e., audio–visual, auditory-only, or visual-only presentations of the performances). To achieve this goal we asked participants to rate three aspects of the performances—performer expression, performer cohesion, and the degree to which participants liked the performance. We predicted that visual information would play a more important role in all ratings than auditory information, based on previous literature (Davidson, 1993; Tsay, 2014; Vines et al., 2006; Vuoskoski et al., 2014). Specifically, we predicted that participant ratings of visual-only and audio–visual stimuli would be more varied across conditions than the audio-only stimuli, indicating that participants were better able to distinguish nuances in performances from visual rather 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 to the initial conditions during the performances, we expected the normal performance where the musicians could both see and hear each other to yield higher ratings of expressiveness, cohesion, and likability than all other conditions because the feedback between the performers was not limited visually or auditorily, as in the other conditions that created fewer means of communication. This prediction is consistent with previous research showing that compensatory methods of communicating with a coperformer are used when visual and/or auditory feedback is diminished (Fulford, Hopkins, Seiffert, & Ginsborg, 2018; Goebl & Palmer, 2009). Goebl and Palmer (2009) found pianists' head movements to have greater synchronization as auditory feedback was reduced, suggesting that visual cuing may function as an alternate means of communication distinct from communicating through sound. Furthermore, limited sensory feedback may make musicians more focused on coperformer communication and synchronization rather than expressing emotions, which could lead to lower expression ratings.

In the next section, we outline how this performer data were collected and our process for creating the point-light stimuli. Then, in subsequent sections, we present the three experiments—each conducted with a new group of participants—based on audio–visual (Experiment 1), audio-alone (Experiment 2), and visual-alone (Experiment 3) presentations of the same performances. All studies met the criteria set by the McMaster University Research Ethics Board.

STIMULI

Performer Participants

We motion-captured and audio-recorded professional clarinetists and pianists performing under four experimental manipulations. Three professional clarinetists (1 female) and three professional pianists (2 female) from the Greater Toronto Area participated in the study for monetary compensation. The clarinetists (mean age = 51.3 years, *SD* = 16.1) had an average of 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. The pianists (mean age = 41.3 years, *SD* = 3.06) had played the piano for an average of 35.0 years (*SD* = 1.00) and had an average of 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, performed regularly in a number of ensembles, and most have taught at a university level. All musicians reported normal hearing and right-handedness.

The performer data collection resulted in 108 duet trials (27 per condition). From these recordings, we selected one trial for each condition and pairing, totaling 36 trials (9 per condition). We selected the trials based on the best audio quality. We then edited the audio recordings of these 36 trials using Reaper software. The motion capture data from the selected trials were cleaned using Qualisys Track Manager software.[1] We used MATLAB (Math Works, Inc.) to create point-light display videos using the Mocap Toolbox (Burger & Toiviainen, 2013; see Figure 1). In one of the pairings, a marker from the pianist was inconsistently rendered in the animations. Therefore, we eliminated that stimulus from experiments with a visual component, leaving 32 point-light display videos for Experiments 1 (audio–visual) and 3 (visual-only) respectively. The corresponding point-light display videos and audio recordings were combined using iMovie software,[2] creating 32 audio–visual clips, 32 visual-only animations, and 36 audio files approximately 40 seconds in length. We used PsychoPy v1.85[3] for all three experiments.

Materials

Clarinetists brought their personal professional-model clarinets and the pianists were provided a Roland FP-80 MIDI keyboard. A directional microphone (AKG C414 XLS) placed in a shield recorded the clarinet to a computer running Reaper software.[4] The clarinetists wore earplugs (noise reduction rating of 32 dB) along with Seinnheiser HDA 200 closed-back headphones. The piano's MIDI output was recorded using the same Reaper program as the clarinet. The pianist wore NADA QH560 open-back headphones. The audio setup allowed for auditory feedback to be adjusted throughout the experiment, depending on the condition.

A Qualisys motion-capture system[5] 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. Eighteen Qualisys cameras recorded the infrared signals reflected off the markers.

Musicians performed an excerpt from the first movement of Brahms' *Clarinet Sonata No. 1* in f minor (bars 1–28; Brahms, 1951). This composition is from the classical–romantic period and allows performers to add emotive expression and timing fluctuations. A copy of the sheet music was provided for musicians with the editor's musical nuances indicated on the score.

Recording Procedure

Each clarinetist 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 the four conditions. To make our results easier to follow, we used abbreviations alongside condition numbers, as outlined in Figure 2. In Condition 1 (FvFa), participants could both hear and see each other, simulating a normal performance setting. In Condition 2 (NvFa), an acoustically transparent screen placed between the musicians blocked visual feedback so they could not see their coperformer. In Condition 3 (FvPa), the pianists could hear both instruments, the clarinetists could hear only themselves, but both musicians could see one another. Condition 4 (NvPa) combined the visual restriction of Condition 2 with the auditory restriction of Condition 3 such that performers were unable to see each other and the clarinetists could hear only themselves while the pianists could hear both instruments. We randomized order of the four performance conditions for each duo in a 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 Audio Feedback	**FvFa** Condition 1	**NvFa** Condition 2
Partial Audio Feedback	**FvPa** Condition 3	**NvPa** Condition 4

Figure 2. Summary and abbreviations of the four performance conditions. In conditions involving partial auditory feedback, the clarinetist could not hear the pianist.

EXPERIMENT 1: AUDIO–VISUAL

Participants observed audio–visual renderings of the professional musicians performing a duet. After each trial, they made ratings of what they had seen and heard, providing the basis for comparison with the audio-only (Experiment 2) and visual-only (Experiment 3) presentations.

Method

Participants and Stimuli

Sixty-three undergraduate students from McMaster University completed the study for course credit or monetary compensation. We presented them with the 32 audio-visual stimuli previously described. Eight participants were excluded from analysis due to incorrectly completing the task.

Evaluation Procedure

Participants watched and listened to all 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 regarding (a) expression, (b) cohesion, and (c) likability. We defined expression as how well the musicians conveyed emotion during the performance. For cohesion ratings, participants evaluated how well the musicians worked together during the performance. Likability was defined as how much participants liked the performance. Each of these three ratings were measured using a continuous scale from 1 (*not expressive/cohesive/likeable*) to 100 (*very expressive/cohesive/likeable*). Participants were asked to consider the entire performance when assigning ratings. After completing each expression and cohesion rating, participants indicated their confidence in their rating on a 7-point Likert scale, from 1 (*not confident*) to 7 (*very confident*). Participants completed ratings for each stimulus in the following order: expression, confidence (of expression rating), cohesion, confidence (of cohesion rating), and likability. The stimuli were presented in a random order for each participant. Participants completed two practice trials consisting of audio–visual recordings of a different Brahms excerpt before starting the experimental trials.

Analyses

Due to some technical difficulties with the PsychoPy interface, 166 of the expression and cohesion ratings were not saved, representing approximately 3.1% of the 5,280 evaluations. This included 61 expression ratings (of 1,760 in total), and 105 cohesion ratings (of 1,760 in total). All participant ratings of likability were successfully recorded. To account for these missing expression and cohesion ratings, we conducted multiple imputations (MI) using the mice (van Buuren & Groothuis-Oudshoorn, 2011) and mitml (Grund, Robitzsch, & Luedtke, 2019) packages on R. MI replaces missing data plausibly based on the observations themselves and a specified statistical model. We ran MI using five imputations, which is the default, and 10 iterations, the number of iterations for which the imputed data reach considerable convergence. In other words, the convergence of imputed data does not improve beyond 10 iterations. We did not run MI on the likability ratings because none of them were missing.

Our experimental design consisted of repeated-measures, therefore we used the lme4 package (Bates, Maechler, Bolker, & Walker, 2015) to fit our imputed expression and cohesion ratings on two separate multilevel models with two predictors as fixed effects: (a) the availability of visual feedback between performers and (b) the availability of auditory feedback between performers. Using a multilevel model allowed us to pool ANOVA-like estimates with imputed data (Grund, 2018). To maintain consistency with the analyses for the imputed expression and cohesion ratings, we also fit the likability ratings on a multilevel model with the same two predictors, even though we did not impute them. Thus, our analyses assessed participants' sensitivity to performance conditions based on audio–visual information (i.e., sound and gestures). 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 Conditions 3 (FvPa) and 4 (NvPa) auditory feedback was only partial. Because we collected data (i.e., ratings) from all levels of interest for each independent variable (e.g., visual feedback between performers: full vs. none; auditory feedback between performers: full vs. partial) and our manipulations pertaining to each level were consistent for each participant, we did not add random slopes in the model. However, to control for random individual differences, we included participant ID as a random intercept in each of our models. This informs the models that there are multiple responses from participants that vary according to their respective baseline levels. Thus, each model assumes an intercept that varies between participants. For each type of rating, we also report the estimated intraclass correlation (ICC) for the intercept-only model with no predictors (i.e., intercept model). ICC indicates the extent to which ratings from the same participant are more similar than those between participants. Then, we report the estimates of the multilevel model with the availability of visual and auditory feedback as predictors.

Results

Expression Ratings

The intercept-only model of the imputed expression ratings with no predictors estimated an ICC of 0.32, which indicates fair individual differences between participants. Consequently, participant ID was added as a random intercept to our multilevel model to control for individual differences. We fit the multilevel model to the imputed expression ratings with two repeated-measures predictors as fixed effects: the availability of visual feedback (full vs. none) and auditory feedback (full vs. partial) between performers. The estimated intercept, which represents the mean expression rating of the FvFa condition, is $\hat{\gamma}_0$=70.66, 95% CI [67.39, 73.92]. The model estimated a significant effect of having no visual feedback between performers, $\hat{\gamma}_1$=–3.78, *t*(28907.01[6])=–3.54, *p*<.001, 95% CI [–5.87, –1.68]: When performers were unable to see each other, expression ratings decreased by 3.78 units, on average, controlling for the effect of the availability of auditory feedback. A significant effect of partial auditory feedback was estimated also, $\hat{\gamma}_2$ = –3.61, *t*(22438.62) = –3.38, *p* = .001, 95% CI [–5.70, –1.52]: If the clarinetists were able to hear themselves only while the pianists heard both instruments, then on average the expression ratings decreased by 3.61 units, controlling for the availability of visual feedback. The estimated effect of the interaction between sight and sound was nonsignificant, $\hat{\gamma}_{1\times2}$ = 2.35, *t*(22905.67) = 1.56, *p* = .120, 95% CI [–0.61, 5.31].

Post hoc multiple comparisons using Tukey adjustments showed the mean expression rating of FvFa to be significantly higher than in all other conditions: NvFa ($p < .001$, 95% CI [–5.87, –1.68]), FvPa; ($p = .001$, 95% CI [–5.70, –1.52]), and NvPa ($p < .001$, 95% CI [–7.19, –2.94]; Figure 3). Table 1 shows mean expression ratings for each condition.

Cohesion Ratings

In the intercept-only model of the imputed cohesion ratings, the estimated ICC was 0.29. Given that the ICC value is quite high, we controlled for individual differences by including participant ID as a random intercept in our multilevel model. The imputed cohesion ratings were fit to the multilevel model with the same two predictors as fixed effects. The model estimated an intercept of $\hat{\gamma}_0 = 72.50$, 95% CI [69.70, 75.30], which represents the mean cohesion rating for the FvFa condition. Significant effects of both no visual feedback and partial auditory feedback were estimated: $\hat{\gamma}_1 = –2.58$, $t(1197.46) = –2.59$, $p = .010$, 95% CI [–4.54, –0.63] and $\hat{\gamma}_2 = –3.19$, $t(3176.92) = –3.42$, $p = .001$, 95% CI [–5.12, –1.26], respectively. Cohesion ratings decreased by 2.58 units on average when the clarinetist and pianist could not see each other, controlling for the effect of auditory feedback. Furthermore, cohesion ratings decreased by an average of 3.19 units when there was partial auditory feedback between performers, controlling for the effect of visual feedback. No significant effect was found for an interaction between the availability of visual and auditory feedback, $\hat{\gamma}_{1\times2} = 0.73$, $t(2677.90) = 0.53$, $p = .599$, 95% CI [–2.00, 3.47].

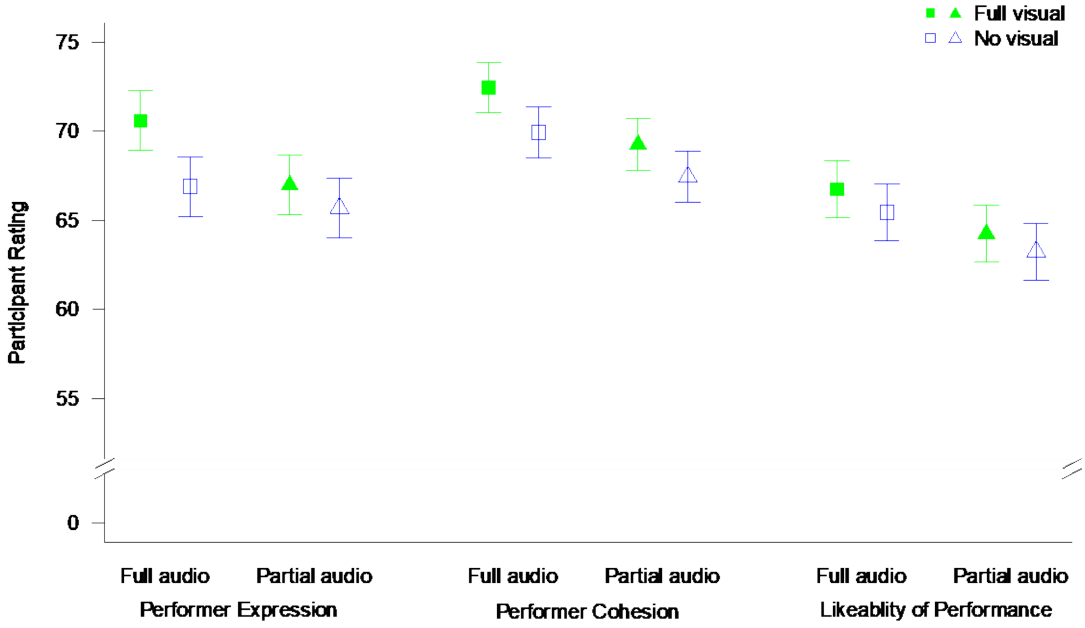

Figure 3. Participant ratings of audio–visual stimuli. Error bars represent standard error about the mean for evaluations of performer expression (left), performer cohesion (middle), and participant liking of performance (right).

Table 1. Descriptive Statistics for Audio–Visual Stimuli.

Parameter	Condition	Description	*M*	*SD*
Expression	1	FvFa	70.61	12.40
	2	NvFa	66.90	12.38
	3	FvPa	66.98	12.40
	4	NvPa	65.69	12.39
Cohesion	1	FvFa	72.46	10.56
	2	NvFa	69.94	10.56
	3	FvPa	69.27	10.66
	4	NvPa	67.44	10.54
Likability	1	FvFa	66.74	11.92
	2	NvFa	65.45	11.92
	3	FvPa	64.25	11.92
	4	NvPa	63.25	11.92

Note. n = 55

Post hoc comparisons using Tukey adjustments found that the mean cohesion rating for FvFa was significantly higher than NvFa (*p* =.010, 95% CI [–4.54, –0.63]), FvPa (*p* = .001, 95% CI [–5.12, –1.26]), and NvPa (*p* < .001, 95% CI [–6.97, –3.11]). The mean cohesion rating of NvFa was also significantly higher than that of NvPa (*p*=.013, *95% CI* [–4.40, –0.52]; Figure 3). No other conditions were significantly different from each other. See Table 1 for means of cohesion ratings for each condition.

Likability Ratings

Because none of the likability ratings were missing, we did not run MI on them. The intercept-only model of the likability ratings estimated an ICC of 0.33, which indicates considerable individual differences between participants. To be consistent with the analyses presented for the expression and cohesion ratings, we fit the likability ratings to a multilevel model with the same two predictors as fixed effects and included participant ID as a random intercept to control for individual differences. The estimated intercept of the model (i.e., the mean likability rating for FvFa) was $\hat{\gamma}_0$ = 66.74, 95% CI [63.57, 69.90]. There was an estimated significant effect of partial auditory feedback, $\hat{\gamma}_2$ = –2.49, *t*(1702) = –2.42, *p* = .016, 95% CI [–4.50, –0.48], which meant that controlling for the effect of visual feedback availability, when there was partial auditory feedback between performers, likability ratings decreased by 2.49 units on average. The effect of no visual feedback between performers and the interaction between visual and auditory feedback were nonsignificant: $\hat{\gamma}_1$ = –1.29, *t*(1702) = –1.26, *p* = .208, 95% CI [–3.31, 0.72], and $\hat{\gamma}_{1\times2}$ = 0.29, *t*(1702) = 0.20, *p* = .840, 95% CI [–2.55, 3.14], respectively.

Tukey adjusted post hoc comparisons showed FvFa to have higher likability ratings than NvPa, p < .004, 95% CI [–6.13, –0.85; Figure 3]. No other conditions were significantly different from each other (see Table 1 for condition means).

Discussion

Participants consistently rated performances where musicians could see and hear each other (FvFa) as being most expressive, cohesive, and likable. When musicians could not see each other and the clarinetist could not hear the pianist (NvPa), participants rated these performances as less cohesive, expressive, and likable. Although this follows our predictions and provides support for past research (Vines, Krumhansl, Wanderley, Dalca, & Levitin, 2011), the ratings decreased only by two to six units on a scale of 1 to 100. Moreover, some of the confidence intervals were close to zero, indicating that the effects were small.

Overall, participants' expression and cohesion ratings were sensitive to the lack of visual and auditory feedback between performers. Likability ratings decreased when there was partial auditory feedback between performers, meaning participants were 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 likability ratings. Specifically, participants rated performances in which the musicians could not see one another as lower in expression and cohesion. Similarly, participants rated performances in which the clarinetist could not hear the pianist as lower in expression, cohesion, and likability.

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 clarinetists and pianists performing a duet using the same procedure and instructions as previously described.

Method

Participants

A new group of 63 undergraduate students from McMaster University completed the study for course credit. We removed eight 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 instead of watching videos, they listened to the auditory component of the audio–visual stimuli described previously. As the audio-only recordings were not affected by issues with motion capture markers described previously, this experiment contained the audio from all 36 performances, rather than the 32 used for Experiments 1 (audio–visual) and 3 (visual-only). 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.

Analyses

As in Experiment 1, some of the expression and cohesion ratings were missing due to technical difficulties. Specifically, 29 of the 1,980 (1.5%) of expression ratings and 83 (4.2%) of the 1,980 cohesion ratings were missing, with all ratings of likability successfully captured. We used MI to estimate the missing expression and cohesion ratings with five imputations and 10 iterations. Similar to the analyses for the previous experiment, five imputations are the default in the mice package and the imputed data converge well with 10 iterations. Then, we fit the imputed expression and cohesion ratings into separate multilevel models with the availability of visual feedback and auditory feedback between performers as fixed effect predictors. None of the likability ratings were missing, so we did not run MI on them. We also fit the likability ratings into a multilevel model with the same predictors. The three multilevel models mentioned included participant ID as a random intercept to control for individual differences. We did not include random slopes in the models because we intend to generalize our findings to the population and ratings were collected from all levels of interest for each predictor. Similar to Experiment 1, we report the ICC for the intercept-only model of the expression, cohesion, and likability ratings in addition to the estimates of the multilevel models with the availability of visual and auditory feedback as fixed effect predictors.

Results

Expression Ratings

The intercept-only model of the imputed expression ratings estimated an ICC of 0.38, which means there are fair individual differences between participants. Thus, we added participant ID as a random intercept in our multilevel model to control for random individual differences. Additionally, we fit the imputed expression data to our multilevel model to examine if the availability of visual (full vs. none) and auditory (full vs. partial) feedback between performers can predict participants' expression ratings. The estimated intercept, which represents the mean expression rating of the FvFa condition, was $\hat{\gamma}_0$ = 72.34, 95% CI [69.40, 75.28]. The model estimated no significant effects of no visual feedback between performers, $\hat{\gamma}_1$ = –0.27, *t*(182956.72) = -0.33, *p* = .745, 95% CI [–1.91, 1.37], or partial auditory feedback between them, $\hat{\gamma}_2$ = –0.84, *t*(89237.19) = –1.00, *p* = .316, 95% CI [–2.48, 0.80]. The estimated effect of the interaction between visual and auditory feedback was nonsignificant, $\hat{\gamma}_{1\times2}$ = –0.11, *t*(15459.89) = –0.10, *p* = .925, 95% CI [–2.44, 2.22; Figure 4]. Mean expression ratings can be seen in Table 2.

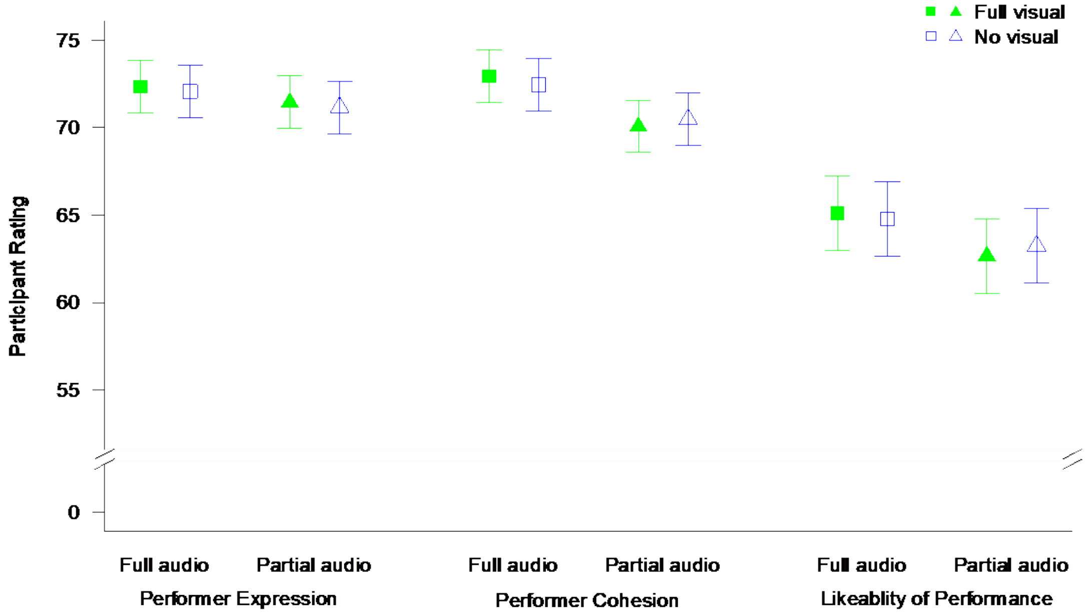

Figure 4. Participant ratings of audio-only stimuli. Error bars represent standard error about the mean for evaluations of performer expression (left), performer cohesion (middle), and participant liking of performance (right).

Table 2. Descriptive Statistics for Audio-Only Stimuli.

Parameter	Condition	Description	*M*	*SD*
Expression	1	FvFa	72.33	11.13
	2	NvFa	72.08	11.13
	3	FvPa	71.45	11.15
	4	*NvPa*	71.17	11.13
Cohesion	1	FvFa	72.95	11.09
	2	NvFa	72.46	11.10
	3	FvPa	70.07	11.07
	4	NvPa	70.48	11.09
Likability	1	FvFa	65.11	15.80
	2	NvFa	64.78	15.80
	3	FvPa	62.67	15.80
	4	NvPa	63.25	15.80

Note. *n* = 55

Cohesion Ratings

The ICC of the intercept-only model of the imputed cohesion ratings was estimated as 0.31. Thus, we added participant ID as a random intercept in our multilevel model to control for random individual differences. The results of fitting the multilevel model to the imputed cohesion ratings with the same two predictors revealed an estimated intercept (i.e., mean cohesion rating for FvFa) of $\hat{\gamma}_0$ = 72.98, 95% CI [70.03, 75.93]. The estimated effect of partial auditory feedback between performers was significant, $\hat{\gamma}_2$ = –2.87, *t*(10063.63) = –3.06, *p* = .001, 95% CI [–4.71, –1.03]. In other words, when the clarinetists could hear themselves only and the pianists could hear both instruments, cohesion ratings decreased by an average 2.87 units, controlling for the effect of the availability of visual feedback. The estimated effect of no visual feedback was nonsignificant, $\hat{\gamma}_1$ = –0.52, *t*(1298.83) = –0.54, *p* = .589, 95% CI [–2.39, 1.36]. The estimated effect of the interaction between visual and auditory feedback was nonsignificant also, $\hat{\gamma}_{1\times2}$ = 0.91, *t*(43570.08) = 0.69, *p* = .489, 95% CI [–1.68, 3.51].

We conducted post hoc comparisons using Tukey adjustments, which found that cohesion ratings were higher for FvFa in comparison to FvPa, *p* = .002, 95% CI [–4.71, –1.03], and NvPa, *p* = .010, 95% CI [–4.36, –0.59]. Moreover, cohesion ratings for NvFa were significantly higher than those for FvPa, *p* = .013, 95% CI [–4.21, –0.50], and NvPa, *p* = .038, 95% CI [–3.80, –0.11; Figure 4]. See Table 2 for mean cohesion ratings for each condition.

Likability Ratings

The intercept-only model of the likability ratings with no predictors revealed an estimated ICC of 0.51. This indicates that ratings from the same participant were more similar than those between participants. To control for these individual differences, we added participant ID as a random intercept in our multilevel model. To be consistent with the analyses conducted for the expression and cohesion ratings, our multilevel model for the likability ratings included the same two repeated-measures predictors as those in the multilevel models of the expression and cohesion ratings. The model revealed an estimated intercept of $\hat{\gamma}_0$ = 65.11, 95% CI [60.91, 69.31] (i.e., the mean likability rating for FvFa). A significant estimated effect of partial auditory feedback was revealed, $\hat{\gamma}_2$ = –2.44, *t*(1922) = –2.60, *p* = .010, 95% CI [–4.28, –0.60], indicating an average decrease of 2.60 units in likability ratings when there was partial auditory feedback between performers, controlling for the effect of visual feedback. The model did not estimate significant effects of no visual feedback, ($\hat{\gamma}_1$ = –0.33, *t*(1922) = –0.35, *p* = .728, 95% CI [–2.17, 1.51], or an interaction between auditory and visual feedback $\hat{\gamma}_{1\times2}$ = 0.91, *t*(1922) = 0.68, *p* = .496, 95% CI [–1.70, 3.51]. With Tukey-corrected post hoc comparisons, we found that the likability ratings were significantly higher for FvFa than for FvPa, *p* = .046, 95% CI [–4.85, –0.03; Figure 4]. Table 2 shows the means for each condition.

Discussion

Participants' mean ratings for expression, cohesion, and likability were all less variable across each condition compared to the same ratings of the audio–visual stimuli. This is consistent with our prediction that our performance condition manipulations would not be easily detected by participants presented with audio-only stimuli. Past research has shown that nonmusicians 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 an effect of the availability of auditory feedback between performers for cohesion and likability ratings, indicating that participant ratings were sensitive to whether or not musicians could fully hear each other during the initial recording sessions. That said, several of the ends of the 95% confidence intervals approached zero (FvFa vs. NvPa, NvFa vs. FvPa, and NvFa vs. NvPa for the cohesion ratings; FvFa vs. FvPa for the likability ratings), which suggest small effects. We conclude that our performance manipulations had minimal effect on our performers' acoustic output, and that listeners were not sensitive to them. This suggests movements helpful for coordination did not lead to differences in their resultant sound, at least among this population. It remains unclear, however, whether greater musical training would help draw attention to these differences—a topic that could provide an interesting avenue for future research.

EXPERIMENT 3: VISUAL ONLY

In Experiment 3, participants rated point-light display videos without sound. This clarified participants' ability to differentiate expression, cohesion, and likability 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

A new group of 63 undergraduates 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 saw only the visual component of the audio–visual stimuli described previously. Therefore participants watched 32 point-light display videos without sound. Before the experimental trials began, participants heard an audio recording of the musical excerpt that was to be performed by the musicians in the point-light displays. For Experiment 3, participants were instructed to focus on the movements of performers when rating videos. We defined expression as how well the performers used their body movements to convey emotion. Cohesion was defined as how well the performers used their body movements to work together during the performance. 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.

Analyses

Due to technical difficulties, 43 of 1,760 (2.3%) expression ratings and 88 of 1,760 (5%) of cohesion ratings were missing. All 1,760 ratings of likability were successfully captured. Consequently, we ran MI (five imputations, 10 iterations) to estimate the expression and cohesion ratings that were missing. To pool ANOVA-like estimates, we fit the imputed expression ratings and imputed cohesion ratings into separate multilevel models. The repeated-measures predictors for each multilevel model were the same as in the previous two experiments: the availability of visual feedback (full vs. none) and the availability of auditory feedback (full vs. partial) between performers, which were specified in the model as fixed effects. Although the likability ratings did not have missing data and we did not run MI on them, we maintained consistency in the analyses by still fitting them to a multilevel model with the same two predictors. In each multilevel model, we did not specify any random slopes. Similar to Experiments 1 and 2, we collected data from every level of interest for each independent variable and we intended to generalize our findings to the population. Participant ID, however, was included in our multilevel models as a random intercept. The ICCs of the intercept-only models are quite high, so adding participant ID as a random intercept controls for individual differences across participants. Thus, we report the ICCs for each intercept-only model and the findings for the multilevel models with our two predictors of interest.

Results

Expression Ratings

When we fit the imputed expression ratings to an intercept-only model with no predictors, we have an estimated ICC of 0.27. To control for individual differences between participants, we included participant ID as a random intercept in our multilevel models. Fitting the imputed expression data to the model with the availability of visual (full vs. none) and auditory (full vs. partial) feedback between performers as predictors revealed an estimated intercept of $\hat{\gamma}_0$ = 65.05, 95% CI [61.73, 68.37], which represents the mean expression rating in the FvPa performance setting. The estimated effect of no visual feedback between performers was significant, $\hat{\gamma}_1$ = –5.23, *t*(6735.93) = –4.41, *p* < .001, 95% CI [–7.55, –2.90]: When performers were unable to see each other, perceived expression decreased by an average of 4.41 units, controlling for the effect of the availability of auditory feedback between the performers. The estimated effect of partial auditory feedback and the estimated interaction effect between visual and auditory feedback were nonsignificant, $\hat{\gamma}_2$ = –2.28, *t*(48770.83) = –4.41, *p* = .053, 95% CI [–4.59, 0.02], and $\hat{\gamma}_{1\times2}$ = –2.76, *t*(55674.50) = –1.66, *p* = .097, 95% CI [–6.02, 0.50], respectively (Figure 5). See Table 3 for mean expression ratings for each condition.

Post hoc comparisons were calculated on expression ratings between the four conditions using Tukey adjustments. Expression ratings were significantly higher for FvFa, *p* < .001, 95% CI [7.94, 12.60], NvFa, *p* < .001), 95% CI [2.72, 7.36], and FvPa, *p* < .001, 95% CI [5.67, 10.30], compared to NvPa. Moreover, expression ratings in FvFa and FvPa performance settings both were higher than those of NvFa: *p* < .001, 95% CI [2.90, 7.55] and *p* = .013, 95% CI [0.62, 5.27], respectively.

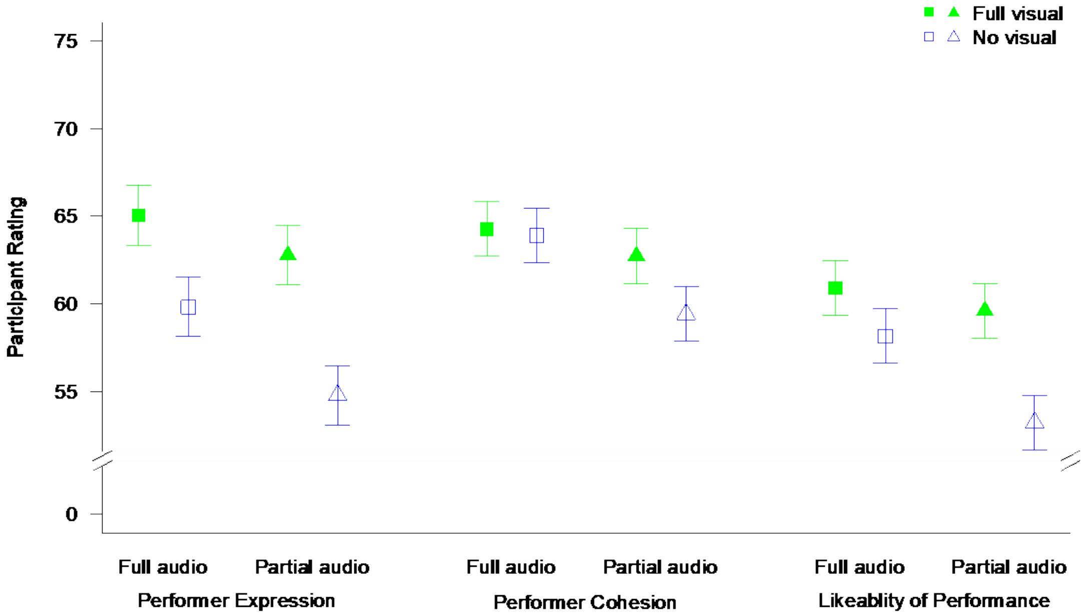

Figure 5. Participant ratings of visual-only stimuli. Error bars represent standard error about the mean for evaluations of performer expression (left), performer cohesion (middle), and participant liking of performance (right).

Table 3. Descriptive Statistics for Visual-Only Stimuli.

Parameter	Condition	Description	M	SD
Expression	1	FvFa	65.05	12.56
	2	NvFa	59.83	12.56
	3	FvPa	62.77	12.56
	4	NvPa	54.79	12.56
Cohesion	1	FvFa	64.27	11.54
	2	NvFa	63.91	11.51
	3	FvPa	62.74	11.61
	4	NvPa	59.42	11.50
Likability	1	FvFa	60.92	11.58
	2	NvFa	58.17	11.58
	3	FvPa	59.61	11.58
	4	NvPa	53.23	11.58

Note. *n* = 55

Cohesion Ratings

The intercept-only model for the imputed cohesion ratings with no predictors estimated an ICC of 0.23. Results from fitting the model to the imputed cohesion ratings with the same two predictors as fixed effect and participant ID as a random intercept revealed an intercept estimate of $\hat{\gamma}_0$ = 64.27, 95% CI [61.22, 67.32], which is the mean cohesion rating for the FvFa performance setting. The model estimated no significant effects of having no visual feedback between performers, $\hat{\gamma}_1$ = –0.36, *t*(742.83) = –0.29, *p* = .772, 95% CI [–2.78, 2.07], or having partial auditory feedback between performers, $\hat{\gamma}_2$ = –1.53, *t*(85345.58) = –1.28, *p* = .199, 95% CI [–3.87, 0.81], and no significant interaction effect between visual and auditory feedback, $\hat{\gamma}_{1\times2}$ = –2.96, *t*(22592.65) = –1.75, *p* = .080, 95% CI [–6.28, 0.35]; Figure 5). See Table 3 for mean cohesion ratings for each condition.

Likability Ratings

The intercept-only model of the likability ratings revealed an estimated ICC of 0.26. Consequently, we added participant ID as a random intercept in our multilevel model to control for considerable individual differences as estimated by the ICC. Fitting the likability ratings to the multilevel model with the same two predictors revealed an estimated intercept (i.e., mean likability rating in FvFa) of $\hat{\gamma}_0$ = 60.92, 95% CI [57.84, 63.99]. Significant estimates were found for the effect of no visual feedback between performers, $\hat{\gamma}_1$ = –2.75, *t*(1702) = –2.46, *p* = .014, 95% CI [–4.93, –0.56]. This means that likability ratings decreased by an average of 2.75 units when performers could not see each other while controlling for the availability of auditory feedback. There was also a significant interaction effect between visual and auditory feedback, $\hat{\gamma}_{1\times2}$ = –3.63, *t*(1702) = –2.30, *p* = .021, 95% CI [–6.72, –0.54]. The estimated effect of partial auditory feedback between performers, however, was nonsignificant, $\hat{\gamma}_2$ = –1.31, *t*(1702) = –1.17, *p* = .241, 95% CI [–3.49, 0.88] (see Figure 5).

Post hoc pairwise comparisons with Tukey adjustments were conducted on likability ratings between each condition to inspect the significant interaction effect further. The simple main effect of the lack of visual feedback between performers was significant when there was only partial auditory feedback between the performers: Likability ratings were significantly higher when performers could see each other than when they could not, holding partial auditory feedback between performers constant (i.e., FvPa vs. NvPa performance settings), *p* < .001, 95% CI [3.51, 9.25]. However, ratings did not differ depending on whether or not the performers could see each other when there was full auditory feedback between them (i.e., FvFa vs. NvFa), *p* = .066, 95% CI [–0.12, 5.61]. The simple main effect of the lack auditory feedback between performers was significant when there was also no visual feedback between the performers: When performers could not see each other, likability ratings were significantly higher when there was full auditory feedback between performers (NvFa) than when there was partial auditory feedback between them (NvPa), *p* < .001, 95% CI [2.07, 7.81]. However, likability ratings for performance settings where performers could fully or partially hear one another did not differ when they could fully see each other (i.e., FvFa vs. FvPa), *p* = .645, 95% CI [-1.562, 4.176]. Additional pairwise comparisons revealed that participants liked FvFa performances significantly more than NvPa ones, *p* < .001, 95% CI [4.82, 10.55]. See Table 3 for mean likability ratings for each condition.

Discussion

Participants in Experiment 1 (audio–visual stimuli) 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 clarinetist could not hear the pianist (Condition 4), as being least expressive, cohesive, and likeable. Ratings from Experiment 3 (visual–only) suggest this difference is driven more by the performances' visual—rather than auditory—component. Together, these experiments are consistent with previous findings on the efficacy of point-light displays as tools for studying ancillary gestures (Dahl & Friberg, 2007; Davidson, 1993). More importantly however, they provide novel evidence for the musical implications of musician's body movements—showing these movements convey information about the conditions in which musicians performed.

Visual feedback had a main effect on participant ratings for expression and likability. This indicates that our manipulations of visual feedback (i.e., whether the musicians could both see one another or not) may have been detected by participants watching the musicians' movements in the absence of hearing performances. Although these effects were small, they reflect that participants were at least sensitive to the manipulations of visual feedback even though they were uninformed of those manipulations. 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 conditions (NvFa, NvPa), musicians were more affected by the visual manipulation since it disrupted sensory information for both performers rather than just one, as occurred during the partial auditory feedback manipulation.

We find it interesting that visual feedback and auditory feedback interacted for likability 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. Overall, Experiment 3 shows that visual information, specifically ancillary gestures viewed by audiences, can play an important role in evaluating a musical performance.

GENERAL DISCUSSION

The current studies examined two levels of communication that exist in a musical performance. The first level dealt with intermusician communication. We wanted to see how the presence of auditory and visual information affected musicians' performance as perceived by an audience. The second level examined how musician duos communicate with audience members. To test these two aims, 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 resulting from a change in musicians' ability to communicate with their coperformers.

Intermusician Communication

To examine intermusician communication, we explored whether participants detected changes in musicians' gestures and audio output as a result of our manipulations in recording conditions—the degree to which musicians could see and hear one another. As participants rated audio-only stimuli consistently across all performer conditions (Experiment 2), we found no evidence that removing musicians' ability to see one another affected evaluation of their sound's cohesion. Furthermore, we found no evidence that removing clarinetists' ability to hear the pianists affected ratings of their sound. Intriguingly however, participants ratings of both audio–visual (Experiment 1) as well as visual-only (Experiment 3) stimuli suggest they were sensitive to communication between the performers.

We also found evidence that participants are sensitive to musicians' movements, as their ratings distinguished between conditions where the musicians could see one another—in the audio–visual and visual-only stimuli. When evaluating the audio-alone stimuli, only their ratings of cohesion distinguished between conditions (see Table 4). We interpret these results as reflecting changes in ancillary gestures between performer conditions, as participants noticed differences in performer movement even in cases where they did not detect differences in their auditory output. This indicates musicians' modulated movements may be serving a communicative purpose, complementing and extending past studies suggesting ancillary gestures are used mainly for expressive purposes (Teixeira, Loureiro, Wanderley, & Yehia, 2014; Teixeira, Yehia, & Loureiro, 2015).

Table 4. Summary of Results from Experiments 1, 2, and 3.

Experiment	**Parameter**		
	Expressivity	Cohesiveness	Likability
1 (audio–visual)	A V	A V	A NS
2 (audio-only)	NS NS	A NS	A NS
3 (visual-only)	A V	NS NS	NS V (I)

Note. Summary of main effects and interactions. *Expressivity*: We found significant estimated effects (but no interaction effect) for both the auditory (Fa vs. Pa) and visual (Fv vs. Nv) manipulations in the audio–visual and visual-only stimuli, but neither estimated effects (nor an interaction effect) in the audio-only condition. *Cohesion*: We found significant estimated effects for both the auditory and visual manipulations (and no interaction effect) in the audio–visual stimuli, but only a significant estimated effect of the auditory manipulation (and no interaction effect) in the audio-only stimuli. *Likability*: We found only a significant estimated effect of the auditory manipulation (but not visual manipulation, nor an interaction effect) in the audio–visual and audio-only stimuli, and only a significant estimated effect of the visual manipulation (as well as a significant interaction effect) in the visual-only stimuli.

Musician to Audience Communication

To observe musician-to-audience communication, we examined how musicians' ability to communicate among themselves affected participant ratings. When performers played under normal performance settings (full vision, full audio), participants consistently rated musicians as most expressive, cohesive, and likable, regardless of the sensory information available to participants. When performers could not see each other and the clarinetist could not hear the pianist (no vision, partial audio), participants rated musicians as least expressive, cohesive, and likable, across all experiments. It appears that musicians are affected by the experimental manipulations and this subsequently affects participant ratings.

We found participants to be more sensitive to performer manipulations when presented with audio–visual stimuli compared to audio-only and visual-only stimuli. The visual-only experiment yielded more differentiation between conditions than the audio-only stimuli. This was true for all ratings—expression, cohesion, and likability—indicating that visual information may allow for better discernment of musical differences than auditory information alone. It appears that participants' perception of expression and likability was influenced by whether or not the musicians could see each other when they watched the point-light displays without sound. On the other hand, participants' cohesion and likability ratings were influenced by whether the performers had full or partial auditory information between them when listening to the audio recordings.

Measuring expressivity, Vuoskoski et al. (2014) also found visual kinematic cues contributing 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 addressed this limitation by using three clarinetists and three pianists, creating nine balanced pairings. 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 skew results.

The use of point-light displays in the present study allowed us to conclude that observed differences in the visual-only stimuli are attributed solely to the performers' body movements. 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 the removal of audio information lowered participants' ratings. Specifically, mean ratings for expression, cohesion, and likability in the visual-only experiment were lower than those in either the audio-only or audio–visual experiments. This is similar to what Vines et al. (2011) found, who attributed lower ratings to novelty of stimuli. Participants are not familiar with watching point-light display videos without sound but are familiar with listening to music alone. Even though the audio–visual stimuli contained point-light displays, the concept of the figures moving to actual sound is familiar; the novel condition is stick-figures moving 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 likability 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 clarinetist could not hear the pianist, but the pianist could always hear everything. The condition where the pianist could not hear the clarinetist 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 clarinetist) 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 synchronizing musical parts. Pianists heard both parts, the assigned leader heard only themselves 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 a more fine-tuned perception of expression and cohesion, especially clarinet and piano players. Previous research with similar paradigms have found comparable emotion ratings between nonmusicians and musicians (Vines et al., 2011), so our expression ratings may be similar regardless of musical training. However, 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 nontrained participants.

CONCLUSIONS

This study demonstrated that visual information is an important aspect of musical performance in a solo instrument–accompanist setting, both for interperformer communication and communication to the audience. Musicians change ancillary gestures depending on the sensory information available to them but are able to keep audio output consistent regardless. We have attributed changing ancillary gestures to the need for musicians to communicate with their coperformers 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.

IMPLICATIONS FOR THEORY AND APPLICATION

By connecting research on ancillary gestures and interpersonal synchronization, these experiments complement and extend previous work, showing how future studies exploring the complex relationship between physical gesture, interperformer coordination, and audience response could shed new light on interpersonal communication. Refining our understanding of how musicians' gestures simultaneously affect performances on numerous levels provides useful insight for musical training. Previous research on musical movements typically explores either ancillary gestures' effects on audiences or on coperformers. Although the specificity afforded by this bifurcation is helpful from a theoretical perspective, musicians' decisions regarding ancillary movements simultaneously affect both their audiences and their coperformers. By exploring these issues in tandem, our research results offer insight useful in applying such research to musical performances. This topic is timely given the recent explosion of interest in socially distanced performances, where musicians may appear together despite having recorded individual parts in isolation.

ENDNOTES

1. See http://www.qualisys.com/software/qualisys-track-manager/ for the particulars on the movement tracking software used in the study.
2. For more information on the software go to https://www.apple.com/ca/imovie/
3. The software details can be accessed at http://www.psychopy.org/
4. See http://www.reaper.fm/ for more information on the software.
5. A more detailed description of the software can be found at http://www.qualisys.com/
6. The degrees of freedom (*df*) for many of these tests appear to be very large as a result of using the mice package on R to impute missing data. The *df* values vary drastically due to the combination of the number of imputations and maximum iterations as specified in the mice function. Consequently, we ran the MI with five imputations (i.e., the default in the mice package) and 10 maximum iterations. We chose 10 iterations because the imputed data appeared to have good convergence with this value.

REFERENCES

Badino, L., D'Ausilio, A., Glowinski, D., Camurri, A., & Fadiga, L. (2014). Sensorimotor communication in professional quartets. *Neuropsychologia, 55*(1), 98–104.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Bishop, L., & Goebl, W. (2018). Performers and an active audience: Movement in music production and perception. *Jahrbuch Musikpsychologie, 28,* 1–17.

Brahms, J. (1951). *Sonata in F minor, for clarinet and piano,* op. 120, no. 1 [Musical score]. New York, NY USA: G. Schirmer. (Original work published 1895)

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.

Burger, B., & Toiviainen, P. (2013). MoCap Toolbox: A Matlab toolbox for computational analysis of movement data. In R. Bresin (Ed.), *Proceedings of the Sound and Music Computing Conference 2013* (SMC 2013; pp. 172–178). Stockholm, Sweden: Logos Verlag Berlin.

Chang, A., Livingstone, S. R., Bosnyak, D. J., & Trainor, L. J. (2017). Body sway reflects leadership in joint music performance. *Proceedings of the National Academy of Sciences, 144*(21), E4134–E4141.

Dahl, S., & Friberg, A. (2007). Visual perception of expressiveness in musicians' body movements. *Music Perception, 24*(4), 433–454.

D'Amario, S., Daffern, H., & Bailes, F. (2018). Synchronization in singing duo performances: The roles of visual contact and leadership instruction. *Frontiers in Psychology, 9,* 1–17.

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

Fulford, R., Hopkins, C., Seiffert, G., & Ginsborg, J. (2018). Reciprocal auditory attenuation affects looking behaviour and playing level but not ensemble synchrony: A psychoacoustical study of violin duos. *Musicae Scientiae, 24*(2), 168–185.

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

Grund, S. (2018, June 25). *Dealing with missing data in ANOVA models* [Web log post]. Retrieved from https://simongrund1.github.io/posts/anova-with-multiply-imputed-data-sets/

Grund, S., Robitzsch, A., & Luedtke, O. (2019). mitml: Tools for multiple imputation in multilevel modeling. R package version 0.3-7.

Mehr, S. A., Scannell, D. A., & Winner, E. (2018). Sight-over-sound judgments of music performances are replicable effects with limited interpretability. *Plos One, 13*(9), e0202075.

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.

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.

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.

Teixeira, E. C. F., Loureiro, M. A., Wanderley, M. M., & Yehia, H. C. (2014). Motion analysis of clarinet performers. *Journal of New Music Research, 44*(2), 97–111. 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.

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

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), 14580–14585.

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

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software, 45*(3), 1–67.

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.

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.

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), article 20150377. http://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.

Wanderley, M. M. (2002). Quantitative analysis of non-obvious performer gestures. In I. Wachsmuth & T. Sowa (Eds.), *Gestures and sign languages in human computer interaction* (pp. 241–253). Berlin, Germany: Springer Verlag.

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.

Authors' Note

We 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 and the LIVE Lab staff, Dave Thompson, Dan Bosnyak, and J. J. Booth, for helping with the technology. This research was supported financially 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). All authors state that they have no conflicts of interest.

All correspondence should be addressed to
Anna Siminoski
McMaster University
424 Togo Salmon Hall
1280 Main Street West
Hamilton, ON, Canada, L8S 4M2
annasiminoski@gmail.com

Human Technology
ISSN 1795-6889
www.humantechnology.jyu.fi