Article

When they listen and when they watch: Pianists' use of nonverbal audio and visual cues during duet performance

Laura Bishop
Austrian Research Institute for Artificial Intelligence (OFAI), Austria

Werner Goebl
Austrian Research Institute for Artificial Intelligence (OFAI), Austria
Institute of Music Acoustics (IWK), University of Music and Performing
Arts Vienna, Austria

Abstract

Nonverbal auditory and visual communication helps ensemble musicians predict each other's intentions and coordinate their actions. When structural characteristics of the music make predicting co-performers' intentions difficult (e.g., following long pauses or during ritardandi), reliance on incoming auditory and visual signals may change. This study tested whether attention to visual cues during piano–piano and piano–violin duet performance increases in such situations. Pianists performed the secondo part to three duets, synchronizing with recordings of violinists or pianists playing the primo parts. Secondos' access to incoming audio and visual signals and to their own auditory feedback was manipulated. Synchronization was most successful when primo audio was available, deteriorating when primo audio was removed and only cues from primo visual signals were available. Visual cues were used effectively following long pauses in the music, however, even in the absence of primo audio. Synchronization was unaffected by the removal of secondos' own auditory feedback. Differences were observed in how successfully piano–piano and piano–violin duos synchronized, but these effects of instrument pairing were not consistent across pieces. Pianists' success at synchronizing with violinists and other pianists is likely moderated by piece characteristics and individual differences in the clarity of cueing gestures used.

Keywords

Corresponding author:
Laura Bishop, Austrian Research Institute for Artificial Intelligence (OFAI) Freyung 6/6, Vienna, A-1010, Austria.
Email: laura.bishop@ofai.at

Introduction

During ensemble performance, people having different musical intentions and, often, playing different instruments, collaborate to create a single shared output. Effective communication helps ensemble performers predict and synchronize with each other's actions (Keller, 2014). Much of this communication is nonverbal. Musicians indicate their intentions with visual gestures, eye contact, breathing, and acoustic cues such as tempo and loudness changes, instead of through speech (Williamon & Davidson, 2002). Musicians' experience and the acoustic properties of their instruments can both facilitate and impair the effective use of incoming cues. Experience in performing similar actions facilitates synchronization with visual gestures (Keller, Knoblich, & Repp, 2007; Luck & Nte, 2008; Wöllner & Cañal-Bruland, 2010), while imprecise perception of note onsets (Gordon, 1987; Rasch, 1979) might limit how accurately incoming cues are interpreted. Such findings raise questions about how effectively musicians use the cues they receive from their co-performers, especially when the co-performers' instruments are ones the musicians have no prior experience in playing and produce sounds whose onsets might be perceived imprecisely. In the present study, musicians' use of nonverbal audio and visual cues during duet performance was investigated. Synchronization among piano–piano and piano–violin duos was assessed to test whether pianists synchronize as successfully with violinists as they do with other pianists.

Performance expertise facilitates action prediction

Musicians can use their own action planning systems to predict their co-performers' behavior. According to the common-coding theory of perception and action, actions and their perceptual effects are jointly represented in the brain (Hommel, Müsseler, Aschersleben, & Prinz, 2001; Keller & Koch, 2008; Prinz, 1990; Taylor & Witt, 2014). The same representation underlies both anticipation and perception of the effects caused (or likely to be caused) by a particular action. As a result, actions can be primed by either perceiving or anticipating their effects. When actions are observed instead of performed, incoming visual signals can trigger internal simulations of the observed actions (Jeannerod, 2003; Keller et al., 2007; Loehr & Palmer, 2011). Perception–action representations are activated, allowing the effects of the actions to be predicted.

Experience in performing a particular instrument strengthens the associations between actions and their perceptual effects, facilitating the prediction of both one's own actions and observed actions (Baumann, Koeneke, Meyer, Lutz, & Jäncke, 2007; Keller & Koch, 2008). Wöllner and Cañal-Bruland (2010), for instance, found string musicians to synchronize finger-taps with silent, individually-presented violin gestures more precisely than did nonstring musicians. String musicians' superior performance on such a task is likely attributable to the larger repertoires of relevant actions that they have represented in their motor planning systems and the stronger associations that exist between those actions and their perceptual effects. Though action prediction improves with visual experience as well, research suggests that the effects of performance expertise are more substantial (Aglioti, Cesari, Romani, & Urgesi, 2008; Calvo-Merino, Grezes, Glaser, Passingham, & Harrad, 2006; Luck & Nte, 2008; Wöllner & Cañal-Bruland, 2010). In the current study, participants' piano performance expertise was expected to facilitate their prediction of observed piano gestures, improving synchronization among piano–piano duos relative to piano–violin duos.

Pianists' synchronization with violinists might also be impaired by imprecision in the perception of incoming cues. Pianists perceptually integrate audiovisual signals from violin-playing

less precisely than they integrate audiovisual signals from piano-playing (Bishop & Goebl, 2014). Ineffective interpretation of violinists' audio and visual cues may reflect a fundamental imprecision in how the human perceptual system processes the sustained bowing gestures and slow-rising sounds produced by string instrument performance. Sustained bowing gestures may not provide as clear an indication of note onsets as the striking actions used in piano performance, rendering visual cues from violinists less informative than visual cues from pianists. For instance, study of violinists' bowing gestures has shown that changes in bow direction lag behind string transitions as performers move between notes (Schoonderwaldt & Altenmüller, 2014), potentially enhancing the ambiguity of observed note onsets. Violinists' audio cues may be less informative as well. For instruments that produce sounds with gradual rise times (e.g., the violin), perceptual onsets tend to occur after the physical onsets of sounded notes, following a delay of variable magnitude that is subject to influence from other simultaneously presented sounds (Gordon, 1987; Rasch, 1979). For instruments with fast rise times (e.g., the piano), perceptual onsets tend to coincide with physical onsets.

Performance expertise can improve the perception of audio signals with gradual rise times (Hofmann & Goebl, 2014). For instance, in the study by Bishop and Goebl (2014), violinists did not show the same imprecise integration of audiovisual violin signals that pianists showed, instead performing similarly for violin and piano stimuli. Performance expertise was not expected to facilitate perception of observed violinist signals in the current study, however, as participants were pianists with little or no string instrument playing experience. Participants accompanied previously recorded performances by either pianists or violinists. Their access to audio and visual signals from the recordings was manipulated so that the effects of instrument pairing on pianists' use of audio and visual cues could be tested independently. Participants' piano performance expertise was primarily expected to affect their use of visual cues, such that the most substantial effects of instrument pairing would occur during visual-only conditions, when synchronization had to be achieved solely on the basis of incoming visual cues.

Hearing the other, seeing the other, and hearing oneself: What do we need to synchronize?

Synchronization with purely visual rhythms tends to be less precise than synchronization with purely auditory rhythms (Hove, Fairhurst, Kotz, & Keller, 2013). Even without visual cues, musically trained and musically untrained listeners can synchronize with sound signals ranging from isochronous rhythms to complex, multi-voiced music (see Repp, 2005; Repp & Su, 2013). Synchronization with isochronous sequences can improve when cues are received through both auditory and visual modalities simultaneously (Elliott, Wing, & Welchman, 2010), but in the more complex context of music ensemble performance, studies investigating the potential benefits of visual cues to synchronization with sounded music have yielded conflicting results.

Some research has shown duet synchronization to be unaffected by the removal of visual contact between pianists (Keller & Appel, 2010), while other research has shown synchronization to improve when pianists can look towards each other (Kawase, 2013). These contrasting findings are likely attributable to differences in the musical material being performed. In the study by Kawase (2013), which showed positive effects of visual contact on synchronization, pianists played a piece containing long pauses and sudden tempo changes. In the study by Keller and Appel (2010), which showed no effects of visual contact, pianists performed a piece with fewer prescribed timing irregularities. Thus, visual communication may be most

important when performers need to synchronize entrances or coordinate abrupt tempo changes. In the present study, it was predicted that synchronization following long pauses and during periods of tempo change would be more successful in the presence of incoming visual cues than in their absence. Though successful synchronization was expected to depend primarily on the presence of incoming audio signals (Goebl & Palmer, 2009), participants were expected to benefit from visual cues at some structurally significant points in the music even when no audio signal was present.

In typical performance situations, ensemble musicians maintain synchrony by monitoring the effects of their own and others' actions and modifying their action plans when asynchronies occur (Repp & Keller, 2008; Repp, Keller, & Jacoby, 2012; Van der Steen & Keller, 2013). Solo piano performance is relatively unimpaired by the removal of auditory feedback (AF), probably because anticipatory imagery strengthens, compensating for the missing sound (Bishop, Bailes, & Dean, 2013; Finney & Palmer, 2003; Repp, 1999). Anticipatory imagery is the experience of actions and their outcomes in advance of their performance or perception, and it contributes to action planning whether AF is present or not (Bishop et al., 2013; Keller, Dalla Bella, & Koch, 2010). Other feedback channels may facilitate anticipatory imagery when AF is unavailable. For instance, imagery may be stronger when motor but no AF is present than when neither auditory nor motor feedback is present (Bishop et al., 2013).

During ensemble performance, sound signals from musicians' own and others' playing might also facilitate imagery. All members of an ensemble must share an integrated representation of the entire piece structure for a coordinated performance to be produced (Keller, 2001). When a performer plays or hears one part, it might be that auditory imagery for the remainder of the music is facilitated, in much the same way that solo performers' auditory imagery is facilitated by the presence of other feedback channels. An aim of the present study was to test whether access to their own AF would improve participants' anticipatory imagery for their co-performers' parts. Improved anticipation of co-performers' actions was expected to occur in the presence of AF, leading to more successful synchronization during normal feedback conditions than during AF deprivation conditions. Benefits of AF were expected even when no sound signal from the co-performer was available. Such a finding would be evidence that AF can improve synchronization by facilitating imagery, and not merely by allowing performers to compare their own and others' overt action outcomes.

Current study

The current study investigated the contributions of incoming audio and visual cues and auditory feedback to successful synchronization during piano–piano and piano–violin duet performance. Pianists performed the secondo part to three duets. They were instructed to synchronize their performance as closely as possible with recordings of a pianist or violinist playing the primo part, as the presence and absence of audio and visual signals from the primo, and the presence and absence of their own auditory feedback were manipulated. Each participant performed the pieces in five audiovisual conditions (within-subject manipulation): (1) while hearing their own auditory feedback and receiving audio but no visual signals from the primo (AO/AF+); (2) while not hearing their own auditory feedback and receiving audio but no visual signals from the primo (AO/AF-); (3) while hearing their own auditory feedback and receiving both audio and visual signals from the primo (baseline); (4) while hearing their own auditory feedback and receiving visual but no audio signals from the primo (VO/AF+); and (5) while not hearing their own auditory feedback and receiving visual but no audio signals from the primo (VO/AO-) (Table 1). Participants performed with either pianist or violinist primo recordings (between-subject manipulation). It was hypothesized that their piano performance experience

Table 1. Audiovisual conditions. AF refers to auditory feedback (i.e., self-audio). In audio-only (AO) conditions, participants could hear but not see the primo; in visual-only (VO) conditions, participants could see but not hear the primo.

Condition	Self-audio	Other-audio	Other-video
AO/AF+	+	+	−
AO/AF-	−	+	−
Baseline	+	+	+
VO/AF+	+	−	+
VO/AF-	−	−	+

would lead pianists to synchronize more precisely with other pianists than with violinists, particularly when the primo audio signal was absent and only visual cues were available. It was also hypothesized that the removal of primos' audio signals would substantially impair synchronization, but participants were expected to achieve some degree of synchronization under visual-only conditions at some structurally-significant points in the music, such as re-entry points. Synchronization success was therefore assessed both at piece-level and at structurally-significant "notes of interest." The combination of aligned audio and visual signals was hypothesized to offer an additional benefit to synchronization, over and above that offered by the audio signal alone. Auditory feedback was expected to facilitate secondos' imagery for both their own and the primos' parts, thus improving synchronization whether the primo's audio signal was available or not.

Methods

Participants

Thirty-one highly-skilled pianists participated in the experiment. All were either current piano students at the University of Music and Performing Arts, Vienna or had completed a university degree in piano performance. Participants were assigned randomly to piano or violin stimulus groups, which were statistically equivalent (all $p > .05$) in terms of age (piano $M = 27.5$, $SD = 8.2$; violin $M = 25.0$, $SD = 4.0$), years of formal training (piano $M = 16.5$, $SD = 4.9$; violin $M = 18.4$, $SD = 4.7$), average number hours of practice per week (piano $M = 23.0$, $SD = 12.5$; violin $M = 25.5$, $SD = 13.7$), and number of performances in the last year (piano $M = 17.1$, $SD = 15.6$; violin $M = 19.9$, $SD = 12.5$). Participants rated their ensemble experience on 5-point scales (1 = "no experience" to 5 = "highly experienced"). With ratings pooled across the three most relevant categories (piano–piano duo, piano–other instrument duo, and small ensemble), piano and violin groups gave equivalent mean self-ratings of "moderately to highly experienced" (out of 15 total points: piano $M = 10.6$, $SD = 1.8$; violin $M = 10.9$, $SD = 2.1$). Fifteen participants completed the experiment as part of the violin group and 16 completed it as part of the piano group. All data from one piano group participant were subsequently disregarded because of problems with equipment, so data from 15 violin and 15 piano group participants were considered for analysis.

Stimuli and equipment

Participants learned the secondo part to three pieces: *Trois Mélodies de 1886,* "Les Anges", a song for voice and piano by E. Satie; *Kirchenmusik*, Op. 23, "Aus tiefer Noth schrei' ich zu dir", a four-voice chorale by F. Mendelssohn; and *Winterabend*, a piano duet for four hands by L. Schytte

(see the appendices). These pieces were selected because they present certain challenges to duet performers, including phrases that end with fermatas in the Mendelssohn, a slow tempo and scope for substantial expressive timing in the Satie, and an exchange of leader/follower roles in the Schytte. Some modifications were made to the original scores to simplify voice-leading and structure and to reduce the size of large chords (see Bishop & Goebl, 2014). Participants played the bottom three lines of the Mendelssohn chorale, the piano accompaniment for the Satie song, and the secondo part to the Schytte duet. A single-voiced melody extracted from each piece served as the primo parts; specifically, the soprano line of the Mendelssohn, the vocal line from the Satie, and the main primo melody from the Schytte.

Stimulus performances had been pre-recorded during a separate session in which a pianist provided live accompaniment to clarinettist, pianist, and violinist soloists. The soloists played the primo part and the accompanying pianist played the secondo part to each of the three pieces. Performers had been instructed to play together with normal expression. Excerpts from these recordings were used as stimuli in an earlier experiment to investigate the audiovisual integration of instrumental music (Bishop & Goebl, 2014). In the current study, recordings of pianist and violinist soloists were used as experimental stimuli, and recordings of clarinettist soloists were used as practice stimuli. During the recording sessions, audio from primo and secondo were recorded in separate channels for each performance. MIDI data from the accompanist and primo pianists, who performed on separate Yamaha CLP-470 Clavinovas, were likewise recorded for each performance. The accompanist sat facing the primo, at a right angle to his or her line of sight. Videos featured only the primo, and were captured by a camera placed just behind the accompanist, so that the primo's glances towards her appeared to be directed towards the camera.

The stimulus set comprised one piano and one violin performance of each piece, plus one clarinet performance of each piece to be used during practice trials. Sample frames from a pianist performance and a violinist performance are shown in Figure 1. Audio-video recordings were imported into Final Cut Pro (FCP) and synched according to their timecodes (which had been aligned at recording using an Ambient Clockit Lanc Logger ALL 601). Selected performances were then exported as MOV files and converted to AVI (XviD codex; 25 frames/second) with Wondershare Converter. The audio track for each video was exported separately from FCP as a WAV file (sampling rate 44.1 kHz).

Stimulus audio and video were presented to participants on an HP Ultrabook running Windows 7, at a 1600 × 900 pixel screen resolution, with Presentation software (Neurobehavioural Systems). Participants performed on a Yamaha CLP470 Clavinova and wore AKG K520 headphones, through which they could hear their own playing and sound from Presentation. Audio from the Clavinova and Presentation were collected using a Focusrite Scarlett 18i8 USB interface. So that video timing could be monitored, videos were modified in FCP to include a small square in the upper right corner that alternated between black and white every 400 ms (10 frames; see Figure 1). An EG & G Vactec VT935G photoresistor was attached to the screen over this square and covered, so as not to distract the participant, and its output was received by the Scarlett interface as an audio signal. The Scarlett interface was connected to a Dimotion Intel Core i5 running Windows 8, and audio from the Clavinova, Presentation, and photoresistor as well as MIDI data from the Clavinova were recorded in Ableton Live 9.1. To check the synchronization of audio-video stimuli presentation, a click track with 400 ms inter-onset intervals and a stimulus video excerpt were played together in Presentation. Sounded clicks and video flashes registered by the photoresister were expected to align exactly in time. The mean asynchrony for the hardware used in the experiment was found to be 3 ms (SD = 10), indicating high precision in the presentation of audio-visual stimuli.
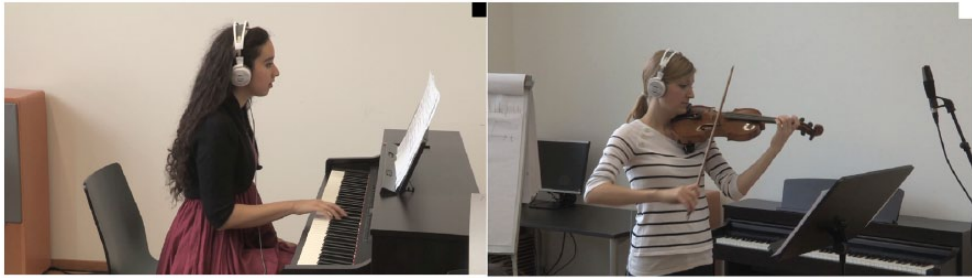
Figure 1. Sample frames from pianist and violinist videos. For both instruments, participants were able to see primos' upper body movements and hand/finger movements. The small square in the upper right corner of the images flashed black and white at 400 ms intervals throughout the duration of each performance, so that precision in audiovisual display could be monitored. This square was covered during the experiment so as not to distract participants.

Design

A 2 (stimulus instrument) × 5 (audiovisual condition) mixed-model design was used. Stimulus instrument was the between-subjects variable; half the participants accompanied piano recordings while the other half accompanied violin recordings. Audiovisual condition was the within-subject variable. There were five audiovisual conditions (Table 1), which participants completed in one of three pseudorandomized orders, always beginning with the baseline condition.

Procedure

At the start of the session, participants received written and spoken instructions and gave informed consent. They were given hard copies of the scores for the three pieces and were asked to practice the secondo part alone, then with audio-video recordings of a clarinettist performing the primo part. They were told that they would accompany a different instrument during the experimental trials, but were not told which instrument it would be. Clarinettists' audio and video were displayed with Presentation during the practice phase, and participants played with normal auditory feedback. They were free to play through each piece as many times as they needed to feel comfortable playing the music with the score. The experimenters wore headphones and monitored participants' playing, pointing out pitch and rhythm errors as necessary and prompting participants to continue practicing until they had learned the music thoroughly and correctly.

During both practice and experimental trials, participants played from scores placed on the music stand on top of the Clavinova. The laptop displaying stimulus videos sat just to the left of the music stand, so that gestures in the video could be either seen in the periphery of participants' visual fields or viewed directly. This set-up is analogous to naturalistic ensemble situations, in which performers get some movement cues via their peripheral vision, but must look away from their musical scores in order to view their co-performers directly.

The experimental conditions were completed as blocks. Participants performed each piece once during each condition. The Satie was always played first, followed by the Mendelssohn, followed by the Schytte. At the start of each condition, participants were given brief written

and spoken instructions so that they knew which auditory and visual channels would be available to them. When they indicated that they were ready to begin a performance, the experimenter started the recording of stimulus and participant performance data (including audio and MIDI data from the participant and audio and photoresistor data from the computer presenting stimuli) by pressing a single key in Ableton. Participants then initiated the trial by pressing a key on the computer displaying stimuli. There was a 2-second pause before the video began.

Videos began immediately prior to primos' initial cueing-in gestures. The baseline condition was always completed first, as the other conditions would have been too difficult to do without any prior audio-visual exposure to the recordings. Research has previously shown that familiarity with co-performers' playing styles increases rapidly across joint rehearsals (Ragert, Schroeder, & Keller, 2013); thus, only one baseline performance was recorded in order to minimize participants' exposure to the stimulus recordings, keeping participants' familiarity with primos' playing styles low and their reliance on primos' audio and visual signals high.

During baseline and VO conditions, videos were shown in their entirety, from start to finish. During AO conditions, the first few seconds of each video were displayed so that participants could synchronize the start of their performance with the stimulus. Videos cut out prior to primos' second note-onset, and the computer screen remained black for the remainder of the trial (cue lengths for Satie 1924 ms (piano) and 2520 ms (violin); Mendelssohn 1380 ms (piano) and 2228 ms (violin); Schytte 1818 ms (piano) and 2125 ms (violin)). Visual rather than audio cues were given at the beginning of trials so that pianists' precision in synchronizing with visual cueing-in gestures could be assessed.

Following each block, participants were asked to reflect verbally on the difficulty of the condition, how successfully they thought they synchronized, and at which locations in the music synchronization was particularly easy or difficult to achieve. At the end of the experiment, participants were debriefed and asked to complete a musical background questionnaire.

Data analysis

Reference note onset profiles were constructed for each piano and violin stimulus recording. Note onsets for piano stimuli were taken from MIDI data recorded during the original performances. Note onsets for violin stimuli were identified manually in Sonic Visualiser by six independent, musically-trained judges, as automatic methods of onset detection proved too inaccurate. The judges averaged 18.3 years of musical training (SD = 3.3) and were primarily pianists (4, including both authors) or woodwind players (2). The mean standard deviation of the judges' identified onsets was 23.2 ms for the Satie, 23.1 ms for the Mendelssohn, and 17.7 ms for the Schytte. A single reference profile was created for each violin stimulus using the median of the onsets identified for each note. Participants' MIDI performances were aligned with the score for each piece using the performance-score matching system developed by Flossmann, Goebl, Grachten, Niedermayer, and Widmer (2010), which pairs performed pitches with notated pitches based on pitch sequence information. Mismatches, which can occur as a result of errors present in the performance or misinterpretation of the score by the matching algorithm, can be identified and corrected using a graphical user interface. Score-matched performances thus comprised all pitches that could be paired with pitches in the score and excluded any erroneously inserted or substituted notes.

So that participant and stimulus note onset profiles could be aligned, both were expressed in terms of time (in milliseconds) since the start of the trial. The start of each trial was taken to be

the initial impulse recorded by the photoresistor, which occurred with the display of the first frame of video. The start of the stimulus video track was used as the reference "time 0" instead of the start of the audio track because video was displayed at the start of all trials (even in AO conditions), while no audio was presented at all during VO trials. Since the experimenter started recording each trial before the participant pressed "start," there could be up to several seconds between the start of the recording and the start of the video display. The time of the initial photoresistor impulse was identified manually for each trial in Audacity, and all stimulus and participant note onsets were measured with respect to this point.

Asynchronies between reference note onset profiles and score-matched participant profiles were calculated in R (R Core Team, 2013) for all notes that should have been simultaneous, according to the score. Primo parts were all single-voiced, but the secondo parts played by participants contained chords with up to six notes; thus, multiple secondo notes were sometimes corresponded to a single primo note. When such chords were present in the secondo part, asynchronies were calculated between the primo note and each secondo note individually. Signed asynchronies, which were negative when participants led and positive when they lagged behind the soloists, were used to identify outliers. The remainder of the analyses used absolute asynchronies, which indicated the magnitude of asynchrony regardless of which performer was leading. Because the stimulus pieces differed substantially in terms of style and tempo, data were never combined across pieces; rather, analyses were always run for each piece individually.

No participant data were excluded on the basis of high pitch error rates (see below), but data from three participants were excluded on the basis of high asynchrony. The mean signed asynchrony was calculated for all trials, and an absolute threshold of 2000 ms was selected as the criterion for identifying outliers. One VO/AF− and four VO/AF+ participant performances were found to exceed this threshold. These performances had mean signed asynchronies below −2000 ms, meaning that the participants were, on average, more than 2 seconds ahead of the soloists. Z-scores for these outliers (calculated within conditions for each piece) ranged between −2.94 and −5.14. Asynchrony during these performances was high enough to merit doubt that the participants had taken the task seriously. As including only partial data from these participants led to problems with the analyses, whenever a participant achieved an outlying mean signed asynchrony in one condition, data from their other performances of that piece also had to be excluded. Thus, all of one piano group participant's Satie performances, all of a second piano group participant's Schytte performances, and all of a violin group participant's Satie and Mendelssohn performances were excluded on the basis of high asynchrony. The same violin group participants' data for Schytte also had to be excluded because one of their performances had failed to record. In total, 28 participants' performances of the Satie, 29 participants' performances of the Mendelssohn, and 28 participants' performances of the Schytte were analyzed.

Linear mixed-effects modelling (LME) was run using the lme command in R ("nlme" package) to investigate the potential effects of group and condition on synchronization. LME tests a combination of fixed effects, which derive from manipulated variables, and random effects, which derive from sources of random error such as subjects or experimental items (Barr, Levy, Scheepers, & Tily, 2012). Subjects can be nested within predictors in the case of repeated-measures designs, and both main effects and interactions between predictors can be tested as in ANOVA. Post-tests were run using the glht command in R ("multcomp" package), which allows comparisons of interest to be specified. All post-test results were evaluated at Bonferroni-adjusted levels of significance.
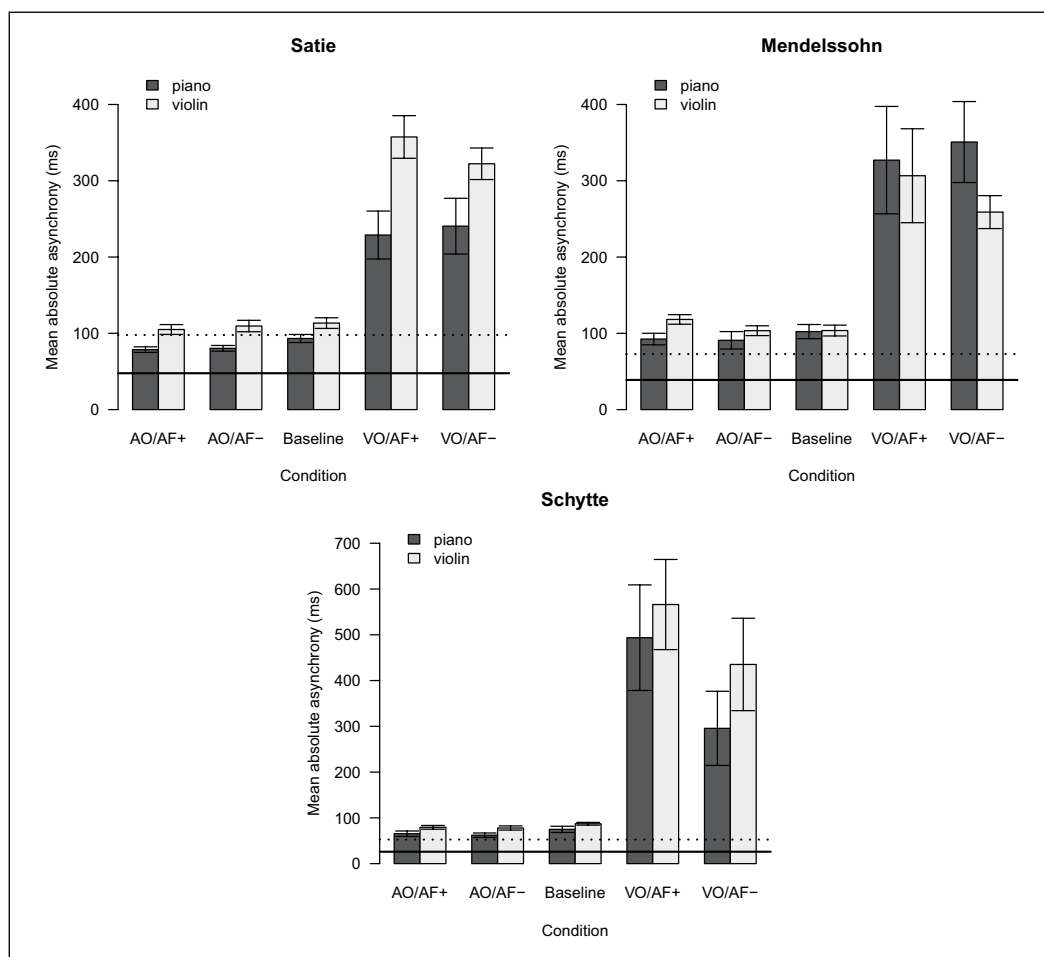
Figure 2. Mean absolute asynchronies by primo group and performance condition. The horizontal lines in each plot indicate the mean absolute asynchronies achieved by the piano–piano duos (solid line) and piano–violin duos (dotted line) during the original recording sessions. Error lines indicate standard errors of the means.

Results

Effects of primo group and performance condition on mean absolute asynchrony

Participants were expected to synchronize more successfully with other pianists than with violinists, especially when only visual cues from the primo were available. The mean absolute asynchronies achieved by the piano and violin groups during the five experimental conditions are shown in Figure 2. Superimposed on these graphs are horizontal lines indicating the mean absolute asynchronies achieved by the original piano–piano and piano–violin duos during the recording sessions, following an interactive practice period and with full, two-way audio and visual contact. Differences between these asynchronies and those achieved by participants in the experiment are noticeable, though they are not analyzed statistically, as each horizontal line only corresponds to a sample of one.

LME was run for each piece using audiovisual condition, primo group, and the interaction between condition and group as the fixed effects variables, subjects nested within condition as the random effects variable, and the mean absolute asynchrony per performance as the dependent variable. The main effect of condition was significant for all pieces ($F(4, 104) = 69.08$, $p <$ .001, $\eta^2 = .85$ (Satie); $F(4, 108) = 25.94$, $p < .001$, $\eta^2 = .70$ (Mendelssohn); $F(4, 104) = 26.86$, $p < .001$, $\eta^2 = .71$ (Schytte)). A significant main effect of group was found only for the Satie, $F(1, 26) = 13.64$, $p = .001$, $\eta^2 = .56$; group was not significant for either the Mendelssohn or the Schytte. A significant interaction between condition and group was likewise found for the Satie, $F(4, 104) = 3.48$, $p = .01$, $\eta^2 = .31$, but neither of the other two pieces.

Series of post-tests were run for each piece, addressing three specific questions. First, to test the importance of auditory and visual cues for successful synchronization, the combination of AO/AF+ and AO/AF− conditions and the combination of VO/AF+ and VO/AF− conditions were compared to the baseline condition for piano and violin groups individually. Synchronization was significantly worse during VO conditions than during the baseline condition for both piano ($z = 5.85$, $p < .001$, $d = 1.65$ (Satie); $z = 5.59$, $p < .001$, $d = 1.19$ (Mendelssohn); $z = 2.69$, $p =$ .007, $d = 1.02$ (Schytte)) and violin groups ($z = 8.30$, $p < .001$, $d = 1.88$ (Satie); $z = 3.37$, $p <$ .001, $d = 1.25$ (Mendelssohn); $z = 4.24$, $p < .001$, $d = 1.10$ (Schytte)), in all pieces. AO conditions did not differ from the baseline condition for either group, for any of the pieces. Thus, synchronization was unimpaired when visual contact between performers was removed, but declined substantially in the absence of auditory cues from the primo performers.

The second set of post-tests assessed whether synchronization success differed between piano and violin primo groups. Groups were compared within each performance condition. Although the piano group outperformed the violin group in almost all conditions across the three pieces, the difference between groups only achieved significance during the VO/AF+ and VO/AF− conditions of the Satie (at $\alpha = .01$; $z = 4.71$, $p < .001$, $d = 1.16$ (VO/AF+); $z = 2.99$, $p = .003$, $d = .74$ (VO/AF−)). No significant between-group differences were observed in VO conditions for either the Mendelssohn or the Schytte. None of the between-group differences for AO or baseline conditions were significant in any of the pieces. Differences in how successfully pianists synchronized with violinists versus other pianists were therefore only slight.

A third pair of post-tests assessed the effects of auditory feedback deprivation on duet synchronization. The combination of AO/AF+ and VO/AF+ conditions and the combination of AO/AF− and VO/AF− conditions were compared to each other for the piano and violin groups individually. Auditory feedback deprivation had no significant effects for either piano or violin groups in the Satie and Mendelssohn. In the Schytte, both piano and violin groups synchronized better with auditory feedback than without, but the difference was only significant for the piano group, $z = 2.41$, $p = .02$, $d = .31$ ($\alpha = .03$). These results suggest that auditory feedback did not make a substantial contribution to synchronization success.

Effects of primo group and performance condition on mean absolute asynchronies at notes of interest

Long pauses in the music and tempo changes were expected to encourage increased reliance on visual cues. To test this hypothesis, notes of interest (NOI) were identified a priori in each piece, and mean absolute asynchronies were calculated at these locations for each performance. Seven NOI were identified for the Satie: the first note of the piece, the last note of the piece, and five additional "restarts," where a new phrase began following a pause of 2–4 beats. Eight NOI were identified for the Mendelssohn: the first note of the piece, the last note of the piece, and six additional restarts, where a new phrase began following a fermata. Four NOI and four

Table 2. Results of LME testing effects of condition, primo group, and NOI.

Condition	28.47** (.74)	9.10** (.51)	25.05** (.71)
NOI	7.22** (.46)	4.36* (.38)	8.35** (.50)
Primo group	1.23	.21	.97
Condition × NOI	9.57** (.53)	3.83** (.77)	3.10** (.33)
NOI × primo group	16.15** (.56)	1.07	1.14
Primo group × condition	1.31	.52	.32
Condition × NOI × primo group	1.24	.35	1.10

Note. **$p < .001$, *$p < .01$. Values in parentheses are effect sizes (Cohen's *d*).

note passages of interest were identified for the Schytte: the first note of the piece, three restarts following pauses of two beats in the secondo part, two "secondo lead" passages, where the main melody was in the secondo part (bars 22–25 and 30–33), and two ritardando passages (one of which included the last note of the piece; bars 34–37 and 60–61). NOI are indicated with boxes on the scores displayed in the appendices (but were not visible on participants' scores).

LME was run for each piece using performance condition, primo group, NOI, and the interactions between them as fixed effects variables, subjects nested within condition and NOI as the random effects variable, and mean absolute asynchronies per NOI category as the dependent variable (results in Table 2). There was no main effect of group, but significant main effects of condition and NOI for all pieces. A significant interaction between condition and NOI was also found for all pieces, while the interaction between group and NOI was only significant for the Satie. Neither the interaction between group and condition nor the interaction between group, condition, and NOI was significant for any of the pieces. Figure 3 shows the mean absolute asynchrony achieved across each NOI category, for each piece, by each primo group, under the different experimental conditions.

Three series of post-tests were run (results in Table 3). The first set of tests assessed the effect of condition on synchronization across NOI categories by comparing both AO conditions and both VO conditions to the baseline condition for each of the NOI categories in each piece, with data pooled across primo groups. Synchronization differed between AO and baseline conditions only at the first note of the Satie, where performance was worse during AO conditions than during the baseline. This was surprising, given that the same visual information was available to participants at that point in all conditions. Synchronization did not differ between AO conditions and the baseline at any other NOI, for any of the pieces, suggesting that successful synchronization could be achieved in the absence of primos' visual cues. Synchronization was also worse in VO conditions than during the baseline at the first note of the Satie, and across restarts in all pieces, final notes in the Satie and Mendelssohn, and secondo lead and ritardando passages in the Schytte, indicating that removal of primos' audio signals impaired synchronization everywhere in the music.

The second series of post-tests assessed the effect of musical structure on synchronization within conditions, with data again pooled across primo groups. NOI categories were compared to non-NOI for the baseline, both AO conditions, and both VO conditions. In baseline and AO conditions, synchronization did not differ significantly between non-NOI and other NOI categories, for any of the pieces. In VO conditions, better synchronization was achieved at first notes than at non-NOI for all pieces. Synchronization was better at restarts than at non-NOI for the Satie and Mendelssohn, while for restarts in the Schytte, the difference was not significant. There was likewise no significant difference between secondo lead passages and non-NOI in the Schytte. In contrast, synchronization at final notes (for the Satie and Mendelssohn) and during
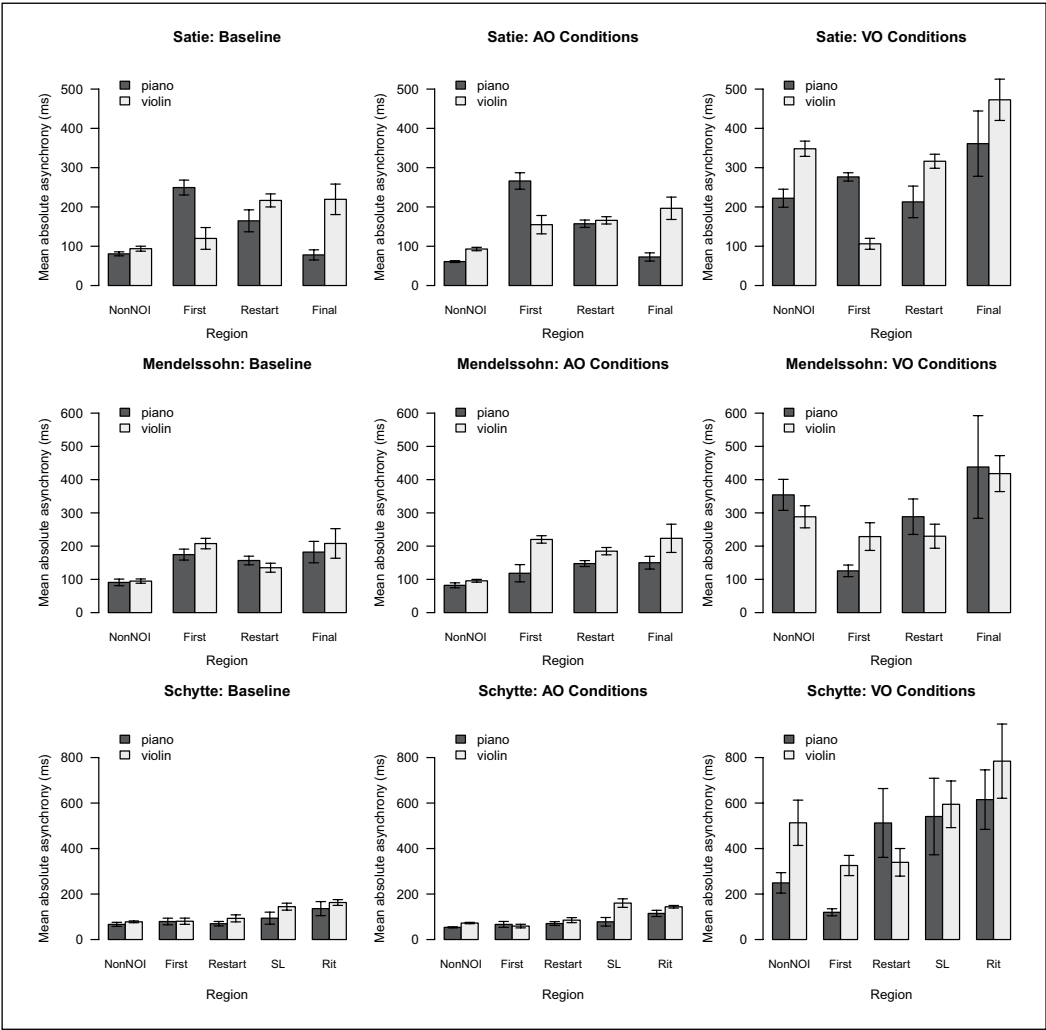
Figure 3. Mean absolute asynchronies at NOI across conditions. "NonNOI" refers to all notes of the piece not included in one of the other categories, "first" refers to the first note or chord of the piece, "restart" refers to phrase restarts following a pause or held note, "SL" refers to secondo lead passages, "rit" refers to ritardando passages, and "final" refers to the final note or chord of the piece. Error lines indicate standard error.

ritardando passages (for the Schytte) was significantly worse than at non-NOI. Thus, while synchronization declined generally in the absence of primo audio, the impairment was much less at entry and re-entry points than elsewhere in the music. Visual cues were not as informative during periods of gradual tempo change as they were at restart points.

The third series of post-tests assessed the differences between piano and violin primo groups at first note, restart, final note, secondo lead, and ritardando NOI. Separate tests were done for the baseline condition, the AO conditions, and the VO conditions. In baseline conditions, the violin group synchronized marginally more successfully than the piano group at the first note of the Satie, while the piano group synchronized more successfully than the violin group at the

Table 3. Results of post-tests (z values).

Comparison	NOI/condition	Satie	Mendelssohn	Schytte
Baseline vs. AO conditions	First note	4.36*** (.22)	.51	.00
	Restarts	1.28	1.13	.05
	Final note	.29	.94	
	Secondo lead			.33
	Ritardando			.36
Baseline vs. VO conditions	First note	3.69*** (.05)	.31	1.16
	Restarts	3.40*** (.55)	2.88** (.57)	3.94*** (.71)
	Final note	7.31*** (.91)	5.25*** (.46)	
	Secondo lead			4.54*** (.77)
	Ritardando			5.80*** (.87)
Non-NOI vs. First note	Baseline	.12	.96	.05
	AO conditions	.87	.17	.11
	VO conditions	4.84*** (.91)	2.44^ (.73)	2.91** (.47)
Non-NOI vs. Restarts	Baseline	1.49	.28	.02
	AO conditions	.09	.26	.11
	VO conditions	4.44*** (.15)	3.03** (.27)	1.51
Non-NOI vs. Final note	Baseline	1.53	.96	
	AO conditions	1.72	.49	
	VO conditions	4.46*** (.51)	3.28** (.23)	
Non-NOI vs. Secondo lead	Baseline			.32
	AO conditions			.39
	VO conditions			2.06
Non-NOI vs. Ritardando	Baseline			.43
	AO conditions			.12
	VO conditions			2.90** (.51)
Baseline: Piano vs. violin	First note	2.32 ^ (1.47)	.36	.01
	Restarts	.93	.17	.17
	Final notes	2.53* (1.27)	.28	
	Secondo lead			.35
	Ritardando			.19
AO conditions: Piano vs. violin	First note	3.75** (.97)	.52	.11
	Restarts	1.12	.23	.05
	Final notes	1.87	.02	
	Secondo lead			.44
	Ritardando			.21
VO conditions: Piano vs. violin	First note	4.87*** (2.64)	.01	.41
	Restarts	.94	2.13	3.57*** (.30)
	Final notes	2.97** (.32)	3.02** (.03)	
	Secondo lead			3.06** (.08)
	Ritardando			2.73** (.23)

Note. ***p < .001, **p < .01, *p = .01. ^p = .02. Values in parentheses are effect sizes (Cohen's d). Bonferroni-adjusted levels of significance are used for each group of contrasts: α = .02 for Satie and Mendelssohn and α = .01 for Schytte.

final note of the Satie. The groups did not differ significantly at any other NOI, for any of the pieces. In AO conditions, the violin group again synchronized more successfully than the piano group at the first note of the Satie. No other between-group differences were significant. In VO
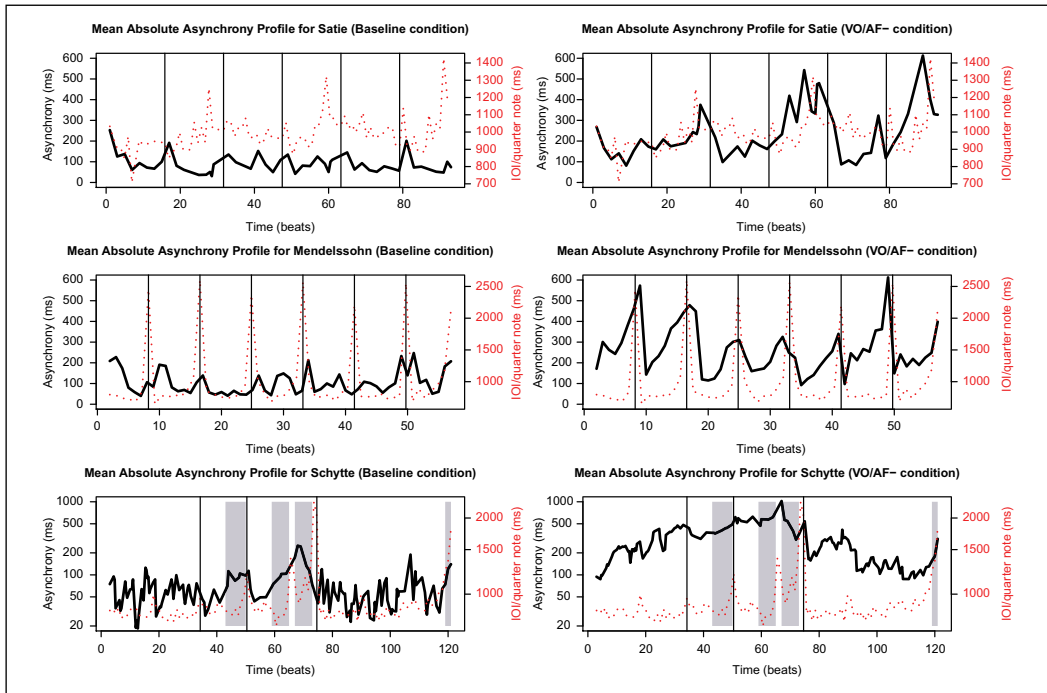
Figure 4. Mean absolute asynchrony profiles for baseline and VO/AF– conditions. The solid line in each graph is the mean absolute asynchrony profile for all participants in the piano group (for Satie and Schytte) or violin group (for Mendelssohn) and corresponds to the left y-axis. The dotted lines are the timing curves for the corresponding stimulus performances, and correspond to the right y-axis in each plot. Vertical lines in the Satie and Mendelssohn graphs indicate restart NOI. In the Schytte, the lines at beats 35, 51, and 75 indicate restarts, the shaded areas between beats 43 and 50 and 119 and 121 indicate ritardando passages, and the shaded areas between beats 59 and 65 and 67 and 73 indicate secondo lead passages. Plots for the Schytte have been rescaled onto log axes so that the full range of asynchronies can be seen in detail.

conditions, the violin group synchronized more successfully than the piano group at the first note of the Satie, at restarts in the Schytte, and at the final note of the Mendelssohn. The piano group synchronized more successfully than the violin group at the final note of the Satie and during secondo lead and ritardando passages in the Schytte. Between-group differences were not significant at the first notes of the Mendelssohn and Schytte or at restarts in the Satie and Mendelssohn. Thus, at times, pianists extracted more accurate information out of violinists' visual gestures than they did out of pianists' visual gestures. The inconsistent pattern of primo group differences observed suggests variation both between and within primo performers in the clarity of cueing gestures used.

Differences in mean absolute asynchronies between baseline and VO/AF– conditions and between NOI and non-NOI can be seen in more detail in Figure 4. For each piece, mean profiles are shown for the baseline condition when the maximum amount of audiovisual information was available, and for the VO/AF– condition when the least amount of audiovisual information was available. These profiles were constructed by averaging individual absolute asynchronies profiles across all participants within a stimulus group. Mean profiles are shown alongside a timing curve for the corresponding stimulus. Timing curves comprise the series of interonset
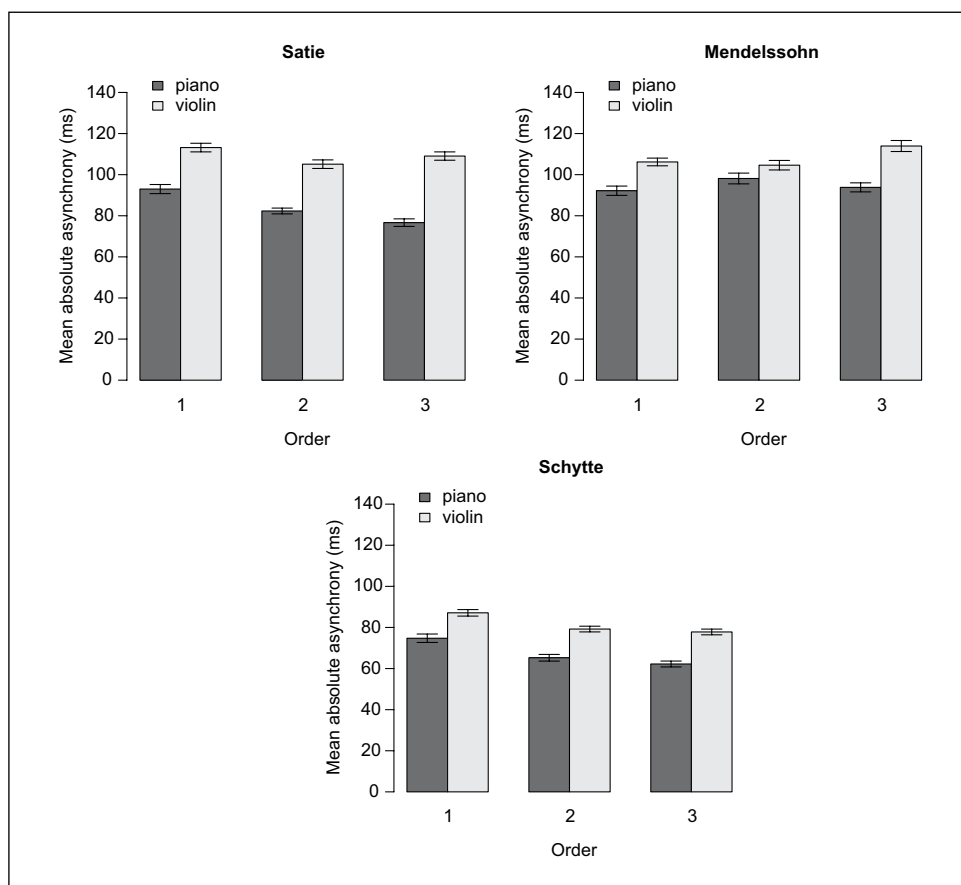
Figure 5. Mean absolute asynchronies across performances with primo audio, in order of completion. Error lines indicate standard errors of the means.

intervals (IOIs) between each quarter note in recorded piano or violin primo performances. IOIs were interpolated wherever notes in the primo melody did not fall on the quarter beat (e.g., half notes in the Satie were subdivided into two equal quarters, corresponding to two points on the timing curve). For the Satie and Mendelssohn in particular—as confirmed with the post-tests detailed above—mean absolute asynchrony tends to peak at phrase boundaries during the baseline condition. In contrast, asynchrony tends to improve at phrase boundaries and worsen within phrases during the VO/AF– condition.

Effects of familiarity with primo playing styles

Most participants reported a subjective increase in familiarity with the primos' playing styles and timing patterns throughout the course of the experiment, so a test was run to investigate whether synchronization improved across performances. Only performances completed with primo audio were considered, as effects of familiarity were less likely to come through during the more unnatural VO conditions. Figure 5 shows the mean absolute asynchrony achieved by piano and violin groups during the conditions with primo audio completed first (always

baseline), second (AO/AF+ or AO/AF−), and third (AO/AF+ or AO/AF−). LME was run for each piece using instrument group, order, and the interaction between them as fixed effects variables, subjects nested within order as the random effects variable, and mean absolute asynchronies as the dependent variable.

The main effect of primo group was significant for the Satie, $F(1, 26) = 12.77$, $p = .001$, $\eta^2 = .38$, and Schytte, $F(1, 26) = 5.63$, $p = .03$, $\eta^2 = .16$, but not for the Mendelssohn. There was also a significant main effect of order for the Satie, $F(2, 52) = 5.15$, $p = .01$, $\eta^2 = .17$, and Schytte, $F(2, 52) = 5.65$, $p = .01$, $\eta^2 = .18$, but not for the Mendelssohn. The interaction between group and order was not significant for any of the pieces. These results provide evidence that, for the Satie and Schytte, instrument pairing affected synchronization during audio conditions. Synchronization also improved across conditions with primo audio for the Satie and Schytte, which is likely attributable to increased familiarity with the primos' playing styles.

Effects of primo group and performance condition on pitch accuracy

Though the effects of primo group and condition on synchronization were of primary interest, the effects of these variables on pitch accuracy were also considered. A count was made of the total number of correctly played notes for each participant performance. Mean rates of pitch accuracy, defined as the proportion of notes in the secondo score to which the matching system could pair a performed note, are shown in Figure 6. Pitch error rates (i.e., 1—pitch accuracy) reflect a combination of omissions (score notes that were not performed) and substitutions (notes performed in place of score notes) (Drake & Palmer, 2000; Flossmann et al., 2010; Palmer & Van de Sande, 1995; Repp, 1996).

LME was run for each piece using condition, primo group, and the interaction between them as fixed effects variables, subjects nested within condition as the random effects variable, and pitch accuracy as the dependent variable. The main effect of condition was significant for all pieces ($F(4, 104) = 11.90$, $p < .001$, $\eta^2 = .31$ (Satie); $F(4, 108) = 12.24$, $p < .001$, $\eta^2 = .31$ (Mendelssohn); $F(4, 104) = 8.08$, $p < .001$, $\eta^2 = .24$ (Schytte)). There was also a marginally significant effect of primo group for the Mendelssohn, $F(1, 27) = 4.17$, $p = .05$, $\eta^2 = .13$, but not for the Satie or Schytte. The interaction between condition and primo group was significant for the Satie, $F(4, 104) = 3.24$, $p = .02$, $\eta^2 = .11$, but not the Mendelssohn or Schytte.

Two series of post-tests were run. The first compared the accuracy achieved by piano and violin primo groups during AO and VO conditions to the accuracy they achieved during the baseline. Both piano and violin groups played with similar accuracy during the AO and baseline conditions for all pieces. In contrast, the piano group performed significantly less accurately during VO conditions than during the baseline for all pieces ($z = 6.87$, $p < .001$, $d = 1.06$ (Satie); $z = 4.79$, $p < .001$, $d = .62$ (Mendelssohn); $z = 5.40$, $p < .001$, $d = .76$ (Schytte)). The violin group performed significantly less accurately during VO conditions than during the baseline for the Satie, $z = 2.51$, $p = .01$, $d = .74$, but not for the Mendelssohn or the Schytte.

The second set of post-tests compared piano group accuracy to violin group accuracy during the baseline, AO conditions, and VO conditions. The groups did not differ during AO or baseline conditions for any of the pieces, but the violin group performed significantly more accurately than the piano group during the VO conditions for all pieces ($z = 5.39$, $p < .001$, $d = .74$ (Satie); $z = 3.87$, $p < .001$, $d = .60$ (Mendelssohn); $z = 4.07$, $p < .001$, $d = .59$ (Schytte)).

The reduced pitch accuracy observed during VO conditions is unsurprising, as demands on participants' attention were high, and their visual focus had to be divided between the video and the musical score, since the music was not memorized. The higher pitch accuracy achieved by the violin group might reflect differences in the magnitude of movements used by violinist
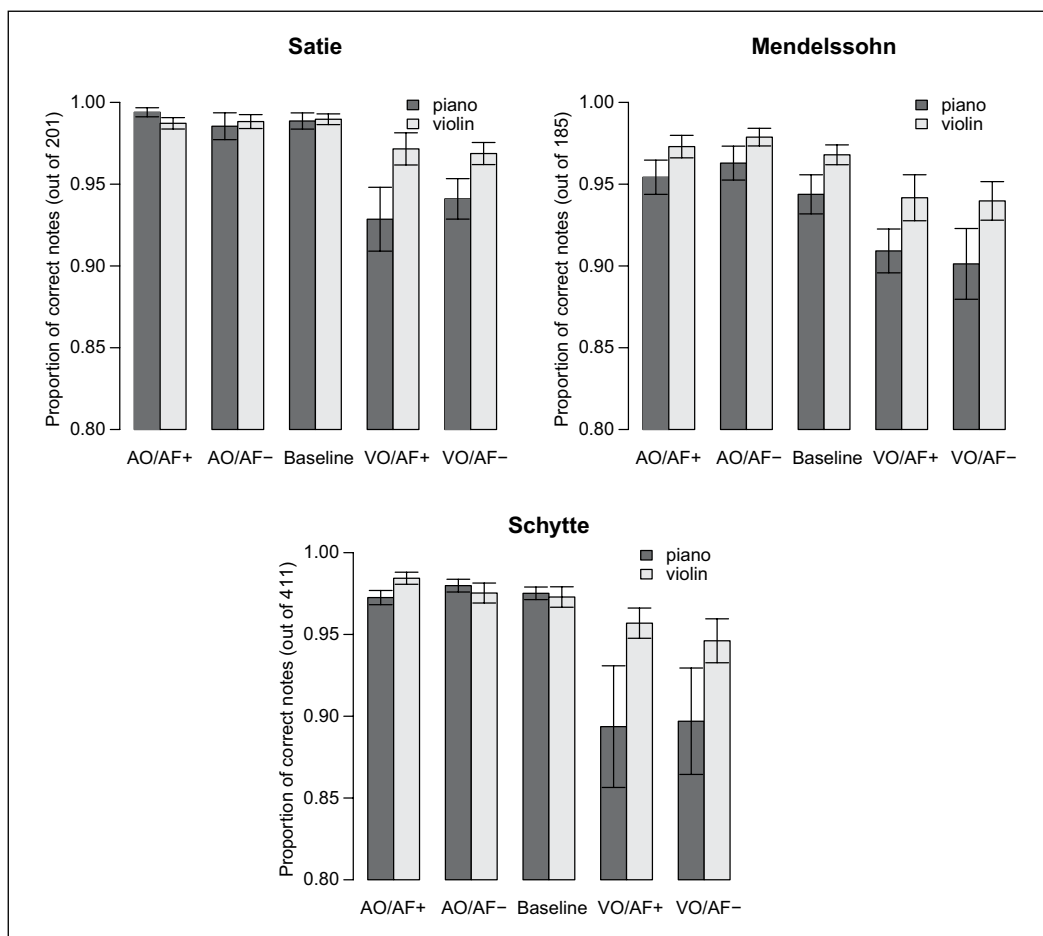
Figure 6. Mean pitch accuracy per condition. Error lines indicate standard errors of the means.

and pianist primos. The larger-magnitude bowing movements used by violinists might have been easier for participants to see in their peripheral vision than the smaller-magnitude hand movements used by pianists, enabling violin group participants to attend more to the score.

Discussion

This study investigated highly-skilled pianists' success at synchronizing with recordings of violinists and other pianists during the performance of three duets (by Satie, Mendelssohn, and Schytte). The presence and absence of audio and video cues from the primo and the presence and absence of participant secondos' own auditory feedback were manipulated in order to test whether reliance on audio and visual signals would change at structurally-significant points in the music ("notes of interest"). Piano–piano duos synchronized more precisely than piano–violin duos during the VO conditions of the Satie; likewise, when only performances completed with primo audio were considered, piano–piano duos synchronized more precisely than piano–violin duos for the Satie and the Schytte. In contrast, while piano–piano duos synchronized

better than piano–violin duos during VO conditions at some notes of interest, piano–violin duos synchronized better than piano–piano duos during AO and VO conditions at other notes of interest. A combination of factors, including performance experience, similarity in performers' playing styles, and the clarity of gestures produced by primos, likely influenced the effective use of visual cues. Participants synchronized less precisely during VO conditions than during conditions with primo audio, but synchrony was better during VO conditions at notes that followed long pauses than elsewhere in the piece, both for Satie and for Mendelssohn. Such effects show that synchronization can be maintained in the absence of visual cues, but visual cues provide important information at times when co-performers' intentions are otherwise difficult to predict. Auditory feedback deprivation had no effect on synchronization, suggesting that pianists do not need to hear the sounds their own actions produce in order to synchronize with a duet partner.

The hypothesis that pianists would synchronize more successfully with other pianists than with violinists was only partly supported. While the first stage of analysis (which assessed synchrony across entire performances) showed superior synchronization among piano–piano duos in some conditions, the second stage of analysis (which assessed synchrony at notes of interest) showed superior synchronization among piano–piano duos at some final notes (Satie) and Schytte secondo lead and ritardando passages, and superior synchronization among piano–violin duos at some first notes (Satie), some final notes (Mendelssohn), and Schytte restart points. Performance experience may have facilitated synchronization to some degree, but in some instances, violinist cueing gestures seem to have been more readily interpretable than pianist cueing gestures. Participants had to divide their visual attention between the video and the music score, and violinists' larger-magnitude gestures might have been more salient at these points than pianists' smaller-magnitude gestures. Also, with only one piano and one violin recording of each piece in the stimulus set, participants' success at synchronizing with visual gestures might have reflected individual differences in the clarity of primos' cueing techniques. Variability within and between primos in the clarity of their cueing gestures probably contributed to the inconsistent pattern of results across pieces and NOI.

For both primo groups, synchronization during AO conditions was substantially better than synchronization during VO conditions. Such a finding aligns with previous research showing that people synchronize finger-taps more successfully with simple sounded rhythms than with simple visual rhythms (Elliott et al., 2010; Hove et al., 2013), and is unsurprising given that temporal acuity is higher within the auditory system than within the visual system. Part of the difficulty in synchronizing during VO conditions may have stemmed from participants' lack of familiarity with the music. Had the music been memorized, participants would not have had to divide their visual attention between the videos and the musical scores. The mean absolute asynchronies for all pieces were less than the duration of a beat in magnitude, though, suggesting that participants extracted enough information from the primos' silent gestures to keep pace with the recordings, even if precise synchronization was not achieved at a note level. Thus, the decline in synchronization seen during VO conditions, compared to AO and baseline conditions, cannot be attributed entirely to the heightened demands on participants' visual attention. Participants were able to make predictions about observed gestures even while dividing their attention between the videos and the scores.

In some performance contexts, it is not unusual for musicians to have to synchronize with silent observed gestures. Orchestral musicians often synchronize with conductor gestures or with performers in other sections whose sound they cannot hear distinctly. In the context of duet performance, in contrast, musicians very rarely need to synchronize with a co-performer they cannot hear. The VO conditions in the current study therefore presented participants with

a particularly unusual task. In normal duet performance situations, musicians monitor the aggregate effects of their own and others' actions in order to maintain synchrony (Keller, 2001). In the absence of primo audio, participants in this study could not monitor synchrony based on a within-modality comparison between their own and their partner's sound. Instead, they would have had to adopt a new monitoring strategy, such as comparing their own sound with an imagined primo sound (VO/AF+ condition), comparing their own imagined sound with an imagined primo sound (VO/AF– condition), or comparing their own visual feedback with the incoming visual signal. Such strategies might be challenging for pianists whose default is to monitor synchrony between their own and others' audio signals. These alternative strategies also depend on pianists' abilities to make accurate inferences about observed gestures. Both difficulty in predicting primos' actions and difficulty in error-correcting, therefore, may have contributed to the superior synchronization observed during conditions with primo audio.

Performance was equally successful during baseline and AO conditions, so the results provide no evidence that simultaneous access to auditory and visual cues facilitates pianists' synchronization. Contrary to previous research (Kawase, 2013), participants did not benefit from combined audiovisual cues even following long pauses. However, visual contact in the current study was only one-way: participant secondos could observe primos' recorded body movements, but not vice versa. Two-way visual contact allows performers to establish eye-contact and exchange visual signals, and likely provides additional benefits. The lack of improved synchronization during the baseline condition might also reflect the fact that this condition was always completed first. That is, at the time of the baseline performances, participants were playing with their pianist or violinist duet partners for the first time. Improved synchronization might have been seen had participants completed a second baseline performance at the end of the testing session.

Though visual cues did not improve synchronization when primo audio was available, in the absence of primo audio, visual cues facilitated synchronization at restart points in the Satie and Mendelssohn. This finding suggests that the most meaningful visual cues for participants were those that indicated primos' intentions to resume playing following a long pause. Similar improvements in synchronization were not observed at points where there was gradual tempo change (e.g., during ritardando passages or on final notes) or at restart points in the Schytte. The Schytte differed from the other pieces in its meter (6/8 instead of 4/4), and the pauses that occurred at its restart points were shorter than those that occurred at restart points in the Satie and Mendelssohn. Secondos may have been more certain about primos' timing in the Schytte, and may have attended less closely to incoming visual cues than they did during performances of the Satie and Mendelssohn.

It is difficult for musicians to predict their co-performers' actions when the music resumes following a pause, and skilled ensemble musicians may attend more closely to others' gestures at these points (especially if following) or exaggerate their own body movements (especially if leading) to ensure synchronization. Further study of musicians' cueing gestures should aim to describe the types of movements used, and should track observers' focus of attention in order to determine whether the more effective visual communication observed at restart points can be attributed to a refocusing of visual attention, greater clarity in body movements, or a combination of both.

AF deprivation had little effect on synchronization, whether primo audio or visual cues were present or not. Previously, auditory imagery and feedback from other modalities have been found to compensate in the absence of AF, allowing solo musicians to perform accurately in silence (Bishop et al., 2013; Keller et al., 2010; Repp, 1999; Wöllner & Williamon, 2007). In the present study, AF had been expected to improve synchronization during visual-only conditions

by facilitating secondos' imagery for the primo part. With performance not substantially different between VO/AF+ and VO/AF− conditions, though, there is no evidence to suggest that access to one's own sound facilitates either imagery for another's part or synchronization in the absence of a primo audio signal. Instead, the results align with other recent research suggesting that pianists do not need to hear their own playing in order to coordinate with a duet partner (Goebl & Palmer, 2009; Zamm, Pfordresher, & Palmer, 2014).

At the end of each condition, participants were asked to reflect verbally on the difficulty of the experimental manipulations. Though they almost universally described the VO conditions as the most difficult, there was little consensus over whether AF helped or hindered synchronization. Some participants also reported benefitting from motor feedback during the AF deprivation conditions, while others said that they had trouble monitoring their own actions without the kinaesthetic feedback normally provided by an acoustic piano. There were also conflicting opinions in both piano and violin groups over whether it was harder or easier to synchronize with the clarinettists during the practice phase than the pianists/violinists during the experimental phase. Participants' opinions over this question—as well as their opinions about which pieces were easier or harder to synchronize—likely relate to how similar their own interpretations were to those of the soloists, and therefore how predictable they found the soloists. The lack of consensus between participants highlights how important it is for the factors that contribute to successful synchronization to be investigated empirically. Besides the factors tested in the current study, duet synchronization is likely affected by the degree of temporal ambiguity present in the music being performed and the compatibility between primo and secondo playing styles (Ragert et al., 2013).

This study provides some evidence that pianists can synchronize more precisely with other pianists during duet performance than with violinists, but participants were also found to use violinists' silent visual cues more effectively than pianists' silent visual cues at some entry and re-entry points, and at the final note of one of the pieces. The effects of performance experience on synchronization success are likely mitigated by factors such as individual differences in the clarity of performers' cueing gestures. The results also show that while synchronization with observed performances is poor without a corresponding audio signal, pianists use visual cues with particular effectiveness at entry and re-entry points. This study focused on follower behavior, with participants playing the role of accompanists. Future research should also consider how leaders and followers jointly modify their behavior in different instrument pairing and audiovisual situations.

References

Aglioti, S. M., Cesari, P., Romani, M., & Urgesi, C. (2008). Action anticipation and motor resonance in elite basketball players. *Nature Neuroscience*, *11*, 1109–1116.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2012). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278.

Baumann, S., Koeneke, S., Meyer, M., Lutz, K., & Jäncke, L. (2007). A network for audio-motor coordination in skilled pianists and non-musicians. *Brain Research*, *1161*, 65–78.

Bishop, L., Bailes, F., & Dean, R. T. (2013). Musical imagery and the planning of dynamics and articulation during performance. *Music Perception*, *31*(2), 97–116.

Bishop, L., & Goebl, W. (2014). Context-specific effects of musical expertise on audiovisual integration. *Frontiers in Cognitive Science*, 5, 1123.

Calvo-Merino, B., Grezes, J., Glaser, D. E., Passingham, R. E., & Harrad, P. (2006). Seeing or doing? Influence of visual and motor familiarity in action observation. *Current Biology*, *16*, 1905–1910.

Drake, C., & Palmer, C. (2000). Skill acquisition in music performance: Relations between planning and temporal control. *Cognition*, *74*(1), 1-32.

Elliott, M., Wing, A., & Welchman, A. (2010). Multisensory cues improve sensorimotor synchronisation. *European Journal of Neuroscience*, *31*, 1828–1835.

Finney, S., & Palmer, C. (2003). Auditory feedback and memory for music performance: Sound evidence for an encoding effect. *Memory & Cognition*, *31*(1), 51–64.

Flossmann, S., Goebl, W., Grachten, M., Niedermayer, B., & Widmer, G. (2010). The Magaloff project: An interim report. *Journal of New Music Research*, *39*(4), 363–377.

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

Gordon, J. W. (1987). The perceptual attack time of musical tones. *Journal of the Acoustical Society of America*, *82*(1), 88–105.

Hofmann, A., & Goebl, W. (2014). Production and perception of legato, portato and staccato articulation in saxophone playing. *Frontiers in Cognitive Science*, 5. doi: 10.3389/fpsyg.2014.00690

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioural and Brain Sciences*, *24*, 849–937.

Hove, M. J., Fairhurst, M. T., Kotz, S. A., & Keller, P. E. (2013). Synchronizing with auditory and visual rhythms: An fMRI assessment of modality differences and modality appropriateness. *NeuroImage*, *67*, 313–321.

Jeannerod, M. (2003). The mechanism of self-recognition in humans. *Behavioural Brain Research*, *142*, 1–15.

Kawase, S. (2013). Gazing behavior and coordination during piano duo performance. *Attention, Perception, & Psychophysics*, *76*(2), 527–540.

Keller, P. E. (2001). Attentional resource allocation in musical ensemble performance. *Psychology of Music*, *29*, 20–38.

Keller, P. E. (2014). Ensemble performance: Interpersonal alignment of musical expression. In D. Fabian, R. Timmers, & E. Schubert (Eds.*), Expressiveness in Music Performance: Empirical Approaches across Styles and Cultures* (pp. 260–282). Oxford: Oxford University.

Keller, P. E., & Appel, M. (2010). Individual differences, auditory imagery, and the coordination of body movements and sounds in musical ensembles. *Music Perception*, *28*(1), 27–46.

Keller, P. E., Dalla Bella, S., & Koch, I. (2010). Auditory imagery shapes movement timing and kinematics: Evidence from a musical task. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(2), 508–513.

Keller, P. E., Knoblich, G., & Repp, B. (2007). Pianists duet better when they play with themselves: On the possible role of action simulation in synchronization. *Consciousness and Cognition*, *16*, 102–111.

Keller, P. E., & Koch, I. (2008). Action planning in sequential skills: Relations to music performance. *The Quarterly Journal of Experimental Psychology*, *61*(2), 275–291.

Loehr, J. D., & Palmer, C. (2011). Temporal coordination between performing musicians. *The Quarterly Journal of Experimental Psychology*, *64*, 2153–2167.

Luck, G., & Nte, S. (2008). An investigation of conductors' temporal gestures and conductor-musician synchronization, and a first experiment. *Psychology of Music*, *36*(1), 81–99.

Palmer, C., & Van de Sande, C. (1995). Range of planning in music performance. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(5), 947–962.

Prinz, W. (1990). A common coding approach to perception and action. In O. Neumann & W. Prinz (Eds.), *Relationships between Perception and Action* (pp. 167–201). Berlin Heidelberg: Springer.

R Core Team. (2013). *R: A language and environment for statistical com*puting [Computer software manual]. Vienna, Austria. Retrieved from http://www.R-project.org/

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 Auditory Cognitive Neuroscience*, *4*. doi: 10.3389/fpsyg.2013.00368

Rasch, R. A. (1979). Synchronization in performed ensemble music. *Acustica*, *43*(2), 121–131.

Repp, B. (1996). The art of inaccuracy: Why pianists' errors are difficult to hear. *Music Perception*, *14*(2), 161–183.

Repp, B. (1999). Effects of auditory feedback deprivation on expressive piano performance. *Music Perception*, *16*, 409–438.

Repp, B. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review*, *12*, 969–992.

Repp, B., & Keller, P. E. (2008). Sensorimotor synchronization with adaptively timed sequences. *Human Movement Science*, *27*, 423–456.

Repp, B., Keller, P. E., & Jacoby, N. (2012). Quantifying phase correction in sensorimotor synchronization: Empirical comparison of three paradigms. *Acta Psychologica*, *139*, 281–290.

Repp, B., & Su, Y. (2013). Sensorimotor synchronization: A review of recent research (2006–2012). *Psychonomic Bulletin & Review*, *20*(3), 403–452.

Schoonderwaldt, E., & Altenmüller, E. (2014). Coordination in fast repetitive violin-bowing patterns. *PLoS ONE*, *9*(9), e106615.

Taylor, J. E. T., & Witt, J. K. (2014). Listening to music primes space: Pianists, but not novices, simulate heard actions. *Psychological Research*. Advance online publication. Retrieved February 8, 2014. doi: 10.1007/s00426-014-0544-x

Van der Steen, M. C., & Keller, P. E. (2013). The ADaptation and Anticipation Model (ADAM) of sensorimotor synchronization. *Frontiers in Human Neuroscience*, *7*. doi: 10.3389/fnhum.2013.00253

Williamon, A., & Davidson, J. W. (2002). Exploring co-performer communication. *Musicae Scientiae*, *6*(1), 53–72.

Wöllner, C., & Cañal-Bruland, R. (2010). Keeping an eye on the violinist: Motor experts show superior timing consistency in a visual perception task. *Psychological Research*, *74*, 579–585.

Wöllner, C., & Williamon, A. (2007). An exploratory study of the role of performance feedback and musical imagery in piano playing. *Research Studies in Music Education*, *29*(1), 39–54.

Zamm, A., Pfordresher, P., & Palmer, C. (2014). Temporal coordination in joint music performance: Effects of endogenous rhythms and auditory feedback. *Experimental Brain Research*, *233*(2), 607–615.

Appendix 1: *Trois Mélodies de 1886*, "Les Anges".

Appendix 2: *Kirchenmusik*, Op. 23, "Aus tiefer Noth schrei' ich zu dir".

Kirchenmusik, Op. 23
Aus tiefer Noth schrei' ich zu dir

Felix Mendelssohn

Appendix 3: *Winterabend.*

Appendix 3. (Continued)