Individual Differences, Auditory Imagery, and the Coordination of Body Movements and Sounds in Musical Ensembles

Peter E. Keller
Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

Mirjam Appel
Eberhard Karls Universität, Tübingen, Germany

THE ROLE OF ANTICIPATORY AUDITORY IMAGERY IN musical ensemble performance was investigated by examining the relationship between individual differences in auditory imagery and temporal coordination in piano duos. Vividness of imagery for upcoming sounds was assessed in 14 pianists using a task that required the production of rhythmic sequences with or without auditory feedback. Ensemble coordination was assessed by examining temporal relations between body movements (recorded by a motion capture system) and sound onsets (triggered by key strokes on two MIDI pianos) in seven duos playing two contrasting pieces with or without visual contact. Sound synchrony was found to be related to anterior-posterior body sway coordination in a manner that depended upon leader/follower relations between pianists assigned to 'primo' and 'secondo' parts. Furthermore, the quality of coordination, which was not affected markedly by whether pianists were in visual contact, was correlated with individual differences in anticipatory auditory imagery. These findings suggest that auditory imagery facilitates interpersonal coordination by enhancing the operation of internal models that simulate one's own and others' actions during ensemble performance.

Received November 6, 2009, accepted April 15, 2010.

Key words: ensemble coordination, music performance, auditory imagery, movement kinematics, individual differences

THE SYNCHRONIZATION OF SOUNDS PRODUCED by different individuals in musical ensembles requires the coordination of their body movements. Achieving the desired temporal relations in the auditory domain thus depends upon the group's collective ability to control interpersonal spatiotemporal relations in motor activity. Differences in the quality of ensemble coordination that can be observed even among groups of highly trained musicians suggest that there is considerable variation in this ability. These individual differences are not well understood despite burgeoning interest in the behavioral and brain bases of interpersonal coordination in music-related contexts (e.g., Goebl & Palmer, 2009; Lindenberger, Li, Gruber, & Müller, 2009; Maduell & Wing, 2007; Merker, Madison, & Eckerdal, 2009; Schögler, 1999–2000, 2003).

One consideration that may affect ensemble coordination is the degree to which ensemble members share a common performance goal, i.e., a unified conception of the ideal integrated sound. A shared goal requires that performers are familiar with one another's parts and that they agree on stylistic matters such as tempo, dynamics, articulation, and phrasing. Carefully conducted observational studies have revealed that the development of shared performance goals is affected by a variety of social factors—such as personality, interpersonal relationships, and verbal and nonverbal communication styles—that influence the effectiveness of information exchange while preparing for performance (Blank & Davidson, 2007; Davidson & King, 2004; Ginsborg, Chaffin, & Nicholson, 2006; Goodman, 2002; Williamon & Davidson, 2002).

Once shared goals are established (and the requisite technical skills are acquired), the quality of ensemble coordination is presumably constrained by the efficiency and reliability of cognitive/motor processes that allow a performer to anticipate, attend, and adapt to the actions of co-performers in real time. Accordingly, Keller (2008) proposed a framework comprised of three core cognitive/motor skills that interact to determine the quality of musical coordination. The most fundamental skill—*adaptive timing*—involves adjusting the timing of one's movements in order to maintain synchrony in the face of small-scale random (unintentional) and expressively motivated (intentional) deviations in local tempo, as well as large-scale global tempo changes and unexpected temporal perturbations (see Repp, 2002, 2005, 2006; Repp & Keller, 2004, 2008). A more advanced cognitive skill—which has been termed *prioritized integrative attending*

Music Perception volume 28, issue 1, pp. 27–46, issn 0730-7829, electronic issn 1533-8312 DOI:10.1525/mp.2010.28.1.27

(Keller, 2001; Keller & Burnham, 2005)—involves dividing attention between one's own actions (high priority) and those of others (lower priority) while monitoring the integrated ensemble sound. Another advanced skill—*anticipatory auditory imagery*—involves the use of mental imagery in planning the production of one's own sounds (Keller & Koch, 2006, 2008) and predicting upcoming sounds of other performers (Pecenka & Keller, 2009a). The focus of the current paper is on the relationship between anticipatory auditory imagery and interpersonal coordination in musical ensembles.

The temporal precision that can be observed in competent musical ensembles suggests the involvement of anticipatory mechanisms. The typical degree of asynchrony between nominally synchronous sounds in such groups, which is around 30–50 ms (Rasch, 1988; Shaffer, 1984), is smaller than would be expected if performers were merely reacting to the sounds of an individual serving as the leader. Ensemble musicians may instead generate online predictions about events in other parts by running *internal models* (see Keller, 2008; Wolpert, Doya, & Kawato, 2003) that simulate the ongoing productions of their co-performers. In social behaviors more generally, it has been proposed that such action simulation plays a role in understanding others' intentions and affective states, and in predicting an observed action's immediate outcome and future course (e.g., Gallese, Keysers, & Rizzolatti, 2004; Knoblich & Sebanz, 2008; Leman, 2007; Overy & Molnar-Szakacs, 2009; Schubotz, 2007; Sebanz & Knoblich, 2009; Wilson & Knoblich, 2005). A recent study of musical action simulation required pianists to record one part from several duets and then, months later, to play the complementary part in time with either their own or others' recordings (Keller, Knoblich, & Repp, 2007). The finding that pianists were more accurate at synchronizing with their own recordings can be explained by assuming that the match between simulated event timing and actual timing in the complementary part is best when both are products of the same cognitive/motor system.

In ensembles such as experienced piano duos, where each performer can potentially execute the movements required to play the other's part, the simulation of co-performers' actions may involve a combination of auditory and motor imagery. The continuity and specificity of both forms of imagery presumably vary, however, as a function of familiarity with the other part(s). In mixed ensembles where musicians are required to synchronize with instruments that they cannot themselves play, accurate motor imagery of others' performances is necessarily limited to general, instrument-independent forms of body motion (e.g., swaying and expressive gesturing) and articulatory activity that could potentially approximate others' sounds (see Schubotz, 2007).[1] Auditory imagery is most likely paramount under such conditions. Indeed, introspective accounts by professional ensemble musicians tend to focus upon the importance of auditory, more so than motor, imagery for performance excellence (see Trusheim, 1993). Self-reports on the ability to imagine co-performers' sounds during private practice are particularly enlightening. For example, elite orchestral musicians are prone to making claims such as, "The sound of what is going on in the rest of the orchestra is always in my imagination" (Trusheim, 1993, p. 145).

The current study investigates the relationship between individual differences in musicians' use of auditory imagery to anticipate upcoming sounds and the ability of those musicians to coordinate with one another in ensembles. Previous work on the role of anticipatory auditory imagery in the production of music-like action sequences has employed behavioral tasks designed to measure the degree to which images of one's own upcoming sounds influence the manner in which subsequent movements are carried out (Keller & Koch, 2006, 2008; Keller, Dalla Bella, & Koch, 2010). The results of these studies indicate that auditory imagery facilitates efficiency in movement planning (e.g., rapidly selecting which keys of a musical instrument to act upon) and biomechanical economy in movement execution. For example, Keller et al. (2010) used a motion capture system to record the finger movements of musicians as they responded to metronomic pacing signals by producing sequences of three unpaced taps on three vertically aligned keys at the given tempo. Successive taps triggered tones of differing pitch in some blocks of trials (where key-to-tone mappings were fixed and, hence, tones were predictable), whereas taps in other blocks did not produce tones. Keller et al. (2010) found that movements were less forceful, i.e., acceleration prior to impact was lowest, when tones were present than when they were absent. This effect can be seen as an index of anticipatory auditory imagery because it was already present at the first tap, i.e., *before* the onset of the first tone in blocks with auditory feedback.

The task used by Keller et al. (2010) was employed in the present study to assess anticipatory auditory imagery in individual pianists from a sample of piano duos. We did so under the assumption that individuals who form

[1]Anticipatory imagery of body movements—in addition to related focal and peripheral visual cues—may play an important role in establishing ensemble coordination at points when instruments are not yet sounding (e.g., at the start of a piece and between musical phrases).

vivid anticipatory auditory images for their own sounds can potentially form vivid images for others' sounds due to overlap in the neural machinery that drives both forms of action simulation (cf. Bekkering et al., 2009; Jeannerod, 2003; Sebanz, Bekkering, & Knoblich, 2006). To examine the hypothesized relationship between the anticipatory auditory imagery and ensemble coordination, measures obtained from the imagery task were compared with measures of the pianists' abilities to coordinate with one another in duos.

Ensemble coordination was quantified by: (1) computing the asynchrony between key strokes for nominally synchronous notes played by the two pianists within a duo (cf. Keller et al., 2007; Shaffer, 1984); and (2) by estimating the cross-correlation between their body sway in the anterior-posterior plane. The use of the latter measure was inspired by a motion capture study that used body sway coordination as an index of interpersonal synchrony during conversation (Shockley, Santana, & Fowler, 2003). The application of this measure to piano duos seemed reasonable in light of qualitative observations that the synchronization of interpersonal body sway occurs in such ensembles (Williamon & Davidson, 2002) and, more generally, that body sway plays a time-keeping role during music performance (Clarke & Davidson, 1998; King, 2006). A recent study by Goebl and Palmer (2009), which investigated the effects of auditory feedback on leader/follower relations during piano duet performance, illustrates the breadth of information that can be gained by examining both key stroke asynchrony and interpersonal movement kinematics. The analysis of the relative timing of key strokes in their study revealed that—under ideal circumstances (i.e., full auditory feedback)—pianists adopted a cooperative performance style, with bidirectional adjustments, irrespective of whether they were assigned to the role of leader or follower. Analysis of the cross-correlation of head movement acceleration time series indicated that visual cues were of greatest importance to coordination when auditory information was absent.

In the present study, the body movements of seven pairs of pianists were recorded using a motion capture system while they performed two contrasting duets on a pair of MIDI pianos. For each piece, the pianists within a duo took turns at playing the *primo* (i.e., the higher voice, which contained the bulk of the melodic material) and *secondo* (the lower voice, which provided accompaniment and occasional complementary thematic material). Key stroke asynchrony was computed based on MIDI data. Body sway coordination was quantified based on kinematic variables extracted from motion capture recordings. Specifically, the strength of the cross-correlation between anterior-posterior movements of the two pianists within each duo was estimated at a series of lags for the variables position, velocity (i.e., rate of change of position), and acceleration (i.e., rate of change of velocity). To gain insight into the movement parameters that ensemble musicians control to obtain optimal spatiotemporal relations during performance, we sought the kinematic variable that is the best predictor of sound synchrony (which is inversely related to key stroke asynchrony). Previous work addressing kinematic cues that are relevant to temporal coordination in musical contexts suggests that acceleration may be informative in this regard (Goebl & Palmer, 2009; Luck & Sloboda, 2009; Luck & Toiviainen, 2006).

Several months after recording the duets, the same pianists were invited back individually to complete the anticipatory auditory imagery task. This task yielded an index that was assumed to reflect the vividness of images for upcoming sounds. To address the main aim of the study, the degree to which the combined imagery indices for pianists within a duo predicted their body sway coordination was then examined. A correlation between these measures would be consistent with the hypothesized role of anticipatory auditory imagery in ensemble coordination.

The present study had two subsidiary aims. The first of these was to investigate how visual information about a co-performer's movements affects ensemble coordination. To this end, visual contact was manipulated by placing an occluder between the pianos for half of the recordings. Previous work has shown that visual cues provided by solo performers' body movements are important in communicating expressive intentions (e.g., Castellano, Mortillaro, Camurri, Volpe, & Scherer, 2008; Dahl & Friberg, 2007; Davidson, 1994; Thompson, Graham, & Russo, 2005; Vines, Krumhansl, Wanderley, & Levitin, 2006). Moreover, qualitative observations have revealed that eye contact, body sway, and hand and head gestures facilitate temporal coordination in musical ensembles (Clayton, 2007; Davidson & Good, 2002; Williamon & Davidson, 2002; see also Davidson, 2009) and studies of unintentional interpersonal coordination in nonmusical contexts typically yield marked effects of visual contact on interpersonal entrainment (e.g., Richardson, Marsh, & Schmidt, 2005; Oullier, de Guzman, Jantzen, Lagarde, & Kelso, 2008). Whether visual contact would have such effects on basic temporal coordination in the current study was an open question: Pianists harbored the intention to synchronize and the musical pieces were rich in auditory information and had regular metric structures with no large-scale tempo changes or silent pauses. We were interested in whether—if visual contact nevertheless affects ensemble coordination—the hypothesized benefits of vivid auditory imagery are most evident in the absence of visual contact.

The second subsidiary aim was to explore leader/follower relations. It was assumed that the pianist playing the primo part in a given performance would function temporarily as the leader of the duo. Rasch (1988) reported that the main melody instruments in string and wind trios sounded about 5–10 ms earlier than the other instruments. We sought to test (albeit in another type of ensemble) whether related leadership effects are evident in body sway. Specifically, we wondered whether the pianist serving as leader would display a greater range of anterior-posterior motion than, and sway ahead of, the follower. Goebl and Palmer (2009) found that the assigned leader raised the fingers higher and produced head movements that preceded those of the follower. Analogous findings for body sway (we did not record finger and head movements) would suggest that the leader uses his or her whole body to provide visual cues that communicate musical intentions. If so, then is this the case even in the absence of visual contact?

Method

Participants

Fourteen pianists (10 male and 4 female; median age = 26 years; range = 20–29 years) were invited to participate in the study. The majority were music students and all had played the piano for 12 or more years (median piano experience = 18.5 years; range = 12–24 years). Although none of the pianists reported regular experience with piano duets or duos, all of them played the piano or other instruments in mixed chamber groups, bands, or orchestras, and some sang in choirs (median ensemble experience = 10 years; range = 5–17 years). The pianists were randomly allocated into five male and two female pairs. The members within each pair did not know one another prior to the study. Pianists received financial reward in return for participation.[2]

Session 1: Duo Performance

MATERIALS

The musical materials were excerpts constituting the early sections of two piano duets by Carl Maria von Weber: No. 1 (Moderato) in D major (Bars 1–35) from *Huit Pièces* (Eight Pieces), Op. 60, and No. 3 (Andante con Variazioni) in G major (Bars 1–32) from *Six Pièces*, Op. 10 (see Appendices). They will be referred to in the following as W1 and W2, respectively. W1 was also used by Keller et al. (2007). It is in triple meter with dynamics ranging from *piano* (soft) to *forte* (loud), and keeps the pianists relatively busy with scale passages and several exchanges of melodic material between *primo* (upper) and *secondo* (lower) parts. W2 is in quadruple meter, to be played softly in a *dolce* (gentle and sweet) manner, and is less technically demanding than W1, although some pianists found the eighth note scale passages in the second half (bars 17–32) challenging. The primo plays the theme in the first half of W2, whereas both parts share thematic material in the second half.

The instruments employed were two Yamaha Clavinova CLP 150 digital pianos. These were positioned back-to-back, as shown in Figure 1. Each piano was set to the default sound (‘Grand Piano 1’) and its volume controller was fixed at the 12 o’clock position. The MIDI output jack of each piano was connected via an M-Audio MIDI interface to a Windows computer running MAX 4.5.7 software, which was used to record the performances in MIDI format.

An Optotrak Certus motion capture system (Northern Digital Inc.) was used to track the three-dimensional position of infrared-light emitting diodes (‘markers’) attached to small plastic tabs that protruded from harnesses worn by the pianists. The harnesses, which were created by making slight alterations to elastic trouser braces, did not hinder pianists’ movements. Each pianist had one marker located on his/her back between the scapulae (shoulder blades) at the thoracic spine; the other marker was placed at the lumbar spine about 30 cm below this point. Four reference markers were fixed to the pianos as shown in Figure 1, with 13.3 cm separating markers in the vertical, horizontal, and depth planes. Information about marker position was relayed by the Optotrak System Control Unit to a second computer for storage. The sampling frequency was 100 Hz. Motion data and MIDI data were synchronized offline based on digital triggers that issued from the serial port of the MAX computer to the System Control Unit (SCU) via an Optotrak Data Acquisition Unit II.

PROCEDURE AND DESIGN

Each pair of pianists came to the laboratory for a single recording session. Prior to this session, pianists had been told over the telephone that they would be required to record piano duets with another pianist, and scanned versions of the musical scores for primo and secondo parts from W1 and W2 were sent to them via email. They

[2]An eighth pair was tested, but their recording session was not completed because one pianist (with seven years piano experience) could not perform the pieces at the correct tempo without major errors, missing notes, and breakdowns.

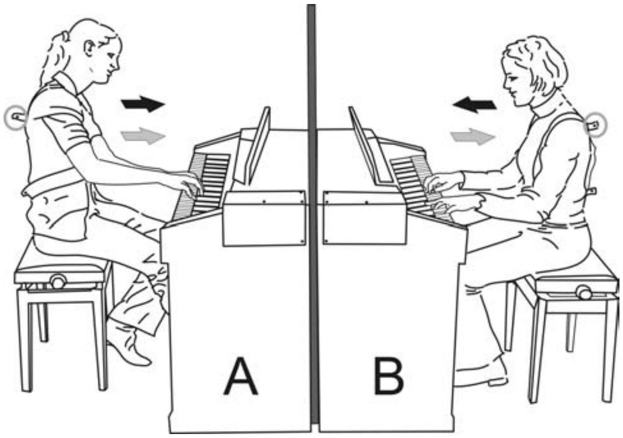

FIGURE 1. The experimental setup from the point of view of the motion capture system. The screen employed to manipulate visual contact is shown between the two pianos, which are labelled A and B. The small dots on the tabs protruding from pianists' backs and the front end of the pianos represent markers that were tracked by the system. The two circled dots correspond to those used in the analysis of body sway. Black arrows indicate movement in the in-phase coordination mode, where both pianists move forward toward (and backward away from) their respective pianos together; grey arrows indicate antiphase coordination, where one pianist moves forward as the other moves backward. Musical notation was used during performance but is not visible in the diagram.

were instructed to learn to perform all of these parts at the same moderate tempo of 100 beats per minute (which corresponds to a 600 ms interbeat interval).

Once both members of a pair had arrived at the laboratory, each pianist was randomly assigned to one of the two pianos. For convenience, the instrument on the left will be referred to as piano A and the instrument on the right will be designated as piano B. Pianists were given the opportunity to warm up and to rehearse the pieces first alone (wearing Sennheiser HD 270 headphones) and then together (without headphones). They were instructed to prepare two versions of each piece, one with the pianist at piano A playing the primo part and the other with piano B playing the primo. Pianists were able to see one another while rehearsing, and they were encouraged to take as much time as required to prepare the pieces. Once both members of the pair indicated to the experimenter that they were satisfied with their ability to perform the pieces accurately and with appropriate expression, the harnesses were fitted and recording commenced.

Three 'takes' that were judged by the pianists and the (musically trained) experimenter to be acceptable performances of each piece were recorded under four different conditions. These conditions resulted from the factorial combination of the variables *visual contact* (present; absent) and *part assignment* (piano A primo; piano B primo). Thus, there were a total of eight conditions in the 2 (visual contact) × 2 (piece) × 2 (part assignment) experimental design. The order in which these conditions were run was counterbalanced across pairs. Pianists were allowed a rest break upon completion of their first four conditions. An opaque screen (150 cm high × 122 cm wide) was placed between the pianos during this break to eliminate visual contact for three pairs of pianists. For the remaining four pairs, who started without visual contact, the screen was removed.

Each take began with a two-bar metronome lead-in; that is, a total of six beats in the case of W1 and eight beats for W2. The metronome consisted of identical piano tones (MIDI note number 98, corresponding to pitch D7) that a MAX program triggered at 600 ms intervals. An additional (silent) trigger that accompanied the final metronome tone was sent by MAX to the SCU, indicating the start of the take. The end of the take was marked by another silent trigger that MAX sent to the SCU when the primo pianist struck a specific key (note number 105, or pitch A7 for W1; note number 103, or pitch G7 for W2) on the first beat of an otherwise silent bar that was appended to the excerpt. The duration of each take was around 63.6 s for W1 and 77.4 s for W2 (which was played without the repeats marked in the original score) if pianists played at the indicated tempo. MIDI data were saved as text files on the MAX computer after each take. Motion data from the first and second halves of the recording session were stored in separate NDI formatted files on the second computer. At the conclusion of the session, pianists were unharnessed and each filled out a questionnaire addressing his or her music experience.

DATA PROCESSING

Key stroke asynchronies were calculated for the best one of the three takes—as judged by the authors—that each pair recorded in each of the eight conditions. This was the third take in all but two out of 56 instances. In the selected takes, asynchronies were computed between notes in the primo and secondo parts that were nominally simultaneous according to the musical score of each piece. If there were multiple simultaneous notes in the same part, only the one with the highest pitch was considered. An exception to this rule was made in the case of the first half of W2 (bars 1–16), where asynchronies were computed between all notes that were paired across the left hand of the primo and the right hand of the secondo. The calculation of asynchronies involved extracting the onset times of these nominally simultaneous notes from the MIDI data, and then subtracting the onset times of notes in the secondo part from the corresponding notes in the primo. This method generated a series of asynchronies for each recording, with negative values indicating that the primo pianist played ahead of the secondo pianist. The total possible number of asynchronies in each series was 114 for W1 and 210 for W2. Four measures of key stroke asynchrony were derived from these series: (1) median *signed asynchrony*; (2) median *unsigned asynchrony*; i.e., the median of the absolute values of asynchronies; (3) the coefficient of variation (*CV*) of signed asynchronies; i.e., the standard deviation of signed asynchronies was divided by the mean inter-beat interval, which was estimated by dividing the time between the first and final (notated) key stroke in the primo by the total number of intervening beats; (4) *synchronization failures*; i.e., the percentage of unsigned asynchronies that were greater than 150 ms (which corresponds to the duration of a sixteenth note at the instructed tempo; we assumed that asynchronies larger than this were unlikely to be intentional) or that could not be computed due to omitted or incorrect notes.

The first step in processing the motion capture data was to convert the NDI formatted files into text format and split them (based on the MAX triggers) to yield a separate file for each take. Prior to splitting, data were transformed in such a way that the markers associated with each pianist were oriented similarly in three-dimensional space, as defined by the Optotrak coordinate system, with a virtual marker at the center of the reference marker cluster on piano A serving as the origin. Forward motion towards the piano resulted in a rightward shift in position on the x-axis for both pianists after the transformation. Since we were interested in anterior-posterior body sway, subsequent processing was performed (using Matlab) only on data pertaining to position on the x-axis of the marker placed between each pianist's shoulder blades. Position time series ranging from one beat before the first key stroke until one beat following the final (notated) key stroke were extracted from the same takes that were used for the analysis of asynchronies. Beat duration was estimated as described above, and is thus a measure of mean beat duration rather than the duration of a particular beat. The extracted position time series were smoothed using a 5th order low-pass Butterworth filter with a cutoff frequency of 10 Hz.

Two sets of analyses were performed on the filtered motion capture data. One set addressed the range of movement in the anterior-posterior plane by each of the two pianists within a duo. The other set examined cross-correlations between their body sway. A z-transformation was applied to the position time series prior to this set of analyses. These normalized position profiles were differentiated once to derive velocity profiles and a second time to derive acceleration profiles for each pianist. Windowed cross-correlation analyses were then conducted on the paired profiles separately for position, velocity, and acceleration. For each analysis, cross-correlations were computed in a window spanning one bar that moved through the take in steps of one beat. Bar durations were computed based on estimates of average beat duration and thus represent the average bar duration for a take. (Note that although average bar and beat durations were used to choose window and step sizes, respectively, the actual cross-correlation analyses were carried out using milliseconds as the temporal unit.) At each step, cross-correlation coefficients were calculated for lags up to ±1 beat with a base lag interval of 1 frame (i.e., 10 ms). This was done in such a way that negative lags entailed shifting the secondo pianist's data series backward in time relative to the primo pianist's series. The maxima and minima of the resultant cross-correlation function at each step, and the lags at which these occurred, were stored in a data array. The following statistics were then computed from this data array: (1) the mean of maximal cross-correlation coefficients across steps (*maximum CC*); (2) the mean lag at which these maxima occurred (*signed lag of maximum CC*); (3) the absolute value of the mean lag at which maxima occurred (*unsigned lag of maximum CC*); (4) mean minimal cross correlation coefficients (*minimum CC*); (5) the mean lag at which these minima occurred (*signed lag of minimum CC*); (6) the absolute value of the mean lag at which minima occurred (*unsigned lag of minimum CC*). Examining absolute lags for maximum and minimum cross-correlations provides a means to address the cohesion of body sway coordination in a manner that is independent of the directionality of leader/follower relations.

Session 2: Anticipatory Auditory Imagery Task

Anticipatory auditory imagery was assessed in the 14 pianists using the task employed by Keller et al. (2010). As the experimental design, materials, and procedure were described in detail in this earlier published study, only a brief synopsis is given below.

Each pianist returned to the laboratory individually 1–10 months after his or her duo recording session. He or she was not told that the task was designed to assess auditory imagery. The task required the participant to respond to an isochronous pacing signal by tapping three response keys in one of four prescribed orders at the tempo set by the pacing signal. The metal keys were embedded in the surface of a wooden box that was angled upwards away from the participant.[3] In response to a fixation cross that

[3]A novel, custom-built instrument was employed (rather than a piano) because in future studies we plan to compare imagery across different instrumentalists.

appeared on a computer monitor at the start of each trial, the participant rested his or her right index finger—which had an Optotrak marker attached to its fingernail—on the middle response key. A pink, yellow, green, or blue color patch immediately appeared and flashed three times with a 600 ms interonset interval, including 200 ms of color plus 400 ms of blank screen. The numbers 1, 2, and 3 then appeared in succession at the same tempo, accompanied by a medium pitched marimba tone (G4; 392 Hz). The participant was required to continue this tempo by tapping the three keys in different orders depending on the color: *pink* = top → middle → bottom; *yellow* = bottom → middle → top; *green* = top → bottom → middle; *blue* = bottom → top → middle.

This task was performed under three auditory feedback conditions in a repeated measures design. Taps triggered high (A4; 440 Hz), medium (G4; 392 Hz), and low (F4; 349 Hz) pitched marimba tones in two of these conditions, while the remaining condition was silent. In one of the conditions with feedback, a tap on the top key triggered the high-pitched tone, a tap on the middle key triggered the medium-pitched tone, and a tap on the bottom key triggered the low tone. The key-to-tone mapping in this condition is 'compatible' in terms of the correspondence between pitch and spatial height. These key-to-tone mappings were reversed in an 'incompatible' condition. Following a block of 16 practice trials (without auditory feedback), six blocks of 44 test trials were run. Each feedback condition appeared once across the first three test blocks, with presentation order counterbalanced. The three conditions were repeated across the final three test blocks. The order of trials with pink, yellow, green, and blue pacing stimuli was randomized within blocks. The first eight trials of each block gave the participant the opportunity to adapt to the current feedback condition, and were not analyzed.

Participants' finger movements were recorded at a sampling rate of 250 Hz by the same motion capture system as was used in session 1. Movement timing and kinematics were analyzed using methods described by Keller et al. (2010). The present study focuses exclusively on kinematic data; specifically, finger acceleration just prior to arrival at each target key.

Results

Session 1: Key stroke asynchrony

Data for median signed asynchrony, CV of asynchronies, median unsigned asynchrony, and synchronization failures for the two pieces under the two visual contact conditions are shown—averaged across duos—in Figure 2. The fact that signed asynchrony was generally negative indicates that the key strokes of the primo player typically preceded those of the secondo player (see Figure 2, top left panel). A *t*-test comparing grand averaged signed asynchronies (i.e., median asynchronies were averaged across all conditions separately for each duo) against zero confirmed that this effect was statistically significant, $t(6) = -8.74$, $p < .001$.

Separate 2 × 2 ANOVAs were run on signed asynchrony, CV, unsigned asynchrony, and synchronization failure data to examine the effects of visual contact (present vs. absent) and piece (W1 vs. W2) on these measures. Data were collapsed across part assignment in these analyses. The ANOVA on signed asynchrony data revealed no statistically significant effects: visual contact, $F(1, 6) = 2.34$, $p = .177$; piece, $F(1, 6) = 0.15$, $p = .713$; visual contact × piece interaction, $F(1, 6) = 1.37$, $p = .287$. Thus, leader/follower relations in key stroke asynchrony were not modulated by visual contact or piece. The ANOVA on CV data, however, yielded a significant main effect of visual contact, $F(1, 6) = 9.23$, $p = .023$. As can be seen in Figure 2 (top right panel), the variability of asynchronies was *higher* when visual contact was present than when it was absent (for W1, at least). The main effect of piece approached significance, $F(1, 6) = 5.40$, $p = .059$, reflecting a tendency for variability to be higher for W1 than W2. The visual contact × piece interaction was not significant, $F(1, 6) = 1.50$, $p = .267$.

The ANOVAs on unsigned asynchrony and synchronization failures both revealed significant effects of piece: $F(1, 6) = 12.60$, $p = .012$; $F(1, 6) = 8.72$, $p = .026$, respectively. As can be seen in the bottom panels of Figure 2, unsigned asynchrony was larger and synchronization failures were more frequent for W1 than W2.[4] These findings, together with the results for CV, corroborate pianists' subjective impressions that W1 is more demanding than W2. The main effect of visual contact and the interaction between visual contact and piece were not significant for either measure, *p*s > .28. Despite this qualitative correspondence, unsigned asynchrony and

[4]The average synchronization failure rate in the current study (13.6%) is higher than the 8.4% failure rate observed by Keller et al. (2007), despite the fact that a narrower tolerance region for unsigned asynchronies was adopted in the earlier study (100 vs. 150 ms). The opportunity for synchronization failures may be greater in duos comprised of two live pianists than for duos comprising one pianist and a recording. Although the percentage of synchronization failures in the current study may appear alarmingly high, this is mainly due to results for W1, where key stroke asynchrony was computed based on only a subset of notes. As noted above, all analyzed performances sounded musically acceptable to the authors and to the pianists themselves.

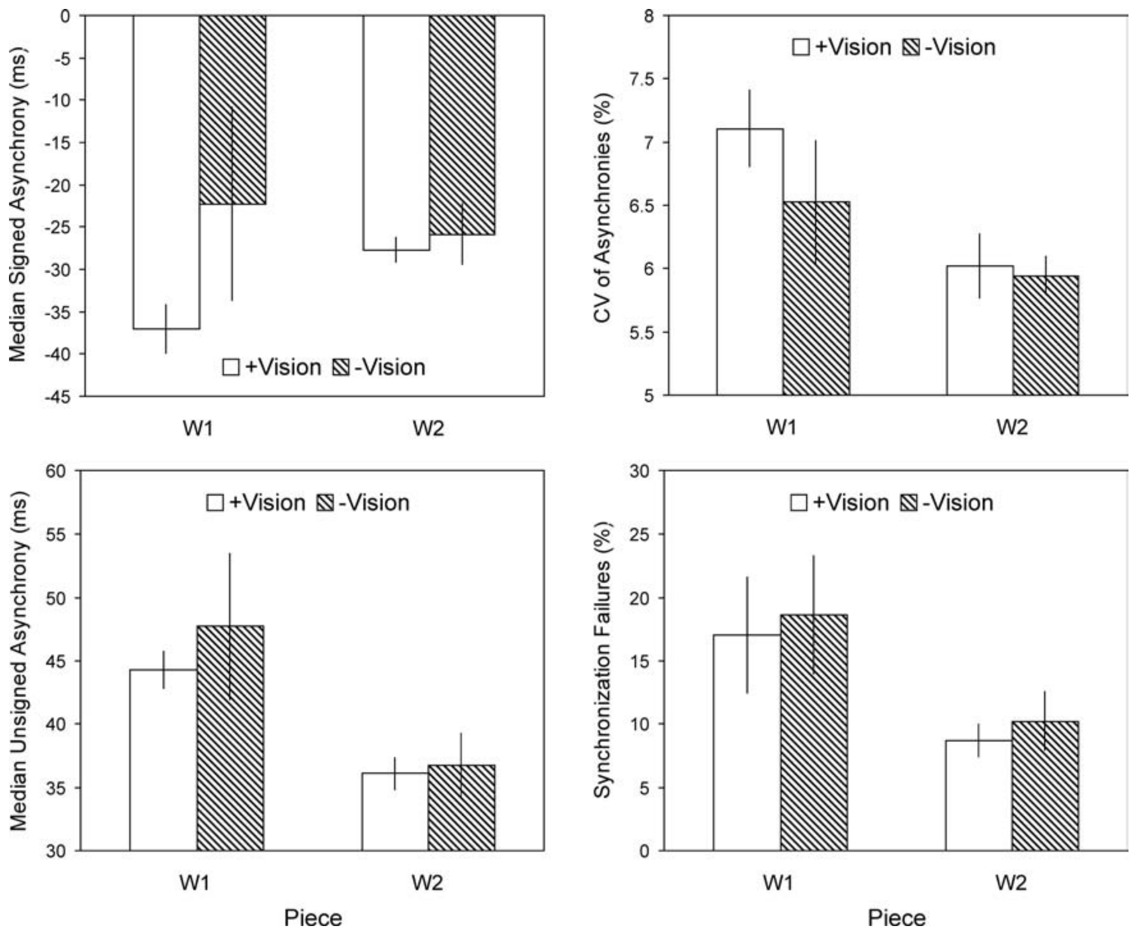

FIGURE 2. Average data for median signed asynchrony (top left, where negative values indicate primo lead), CV of asynchronies (top right), median unsigned asynchrony (bottom left), and synchronization failures (bottom right) for the two pieces (W1; W2) under the two visual contact conditions. "+Vision" = visual contact present; "–Vision" = visual contact absent. Error bars indicate the standard error of the mean.

synchronization failure data—each averaged across visual contact conditions and pieces—were not correlated with one another across duos: $r(5) = -.32$, $p = .484$ [$r_s = -.43$, $p = .337$]. CV data were likewise not correlated with either of these measures: ps > .256. This suggests that the three measures may reflect independent aspects of sound synchrony. Therefore, a composite asynchrony index was formed by combining the CV, unsigned asynchrony, and synchronization failure measures for use in further analyses. The first step in computing this composite was to standardize—via z-transformation—the CV, unsigned asynchrony, and synchronization failure data for each pair of pianists in each of the eight performance conditions. *Z*-scores were then averaged across performance conditions separately for each of the three measures. Next, the resultant mean *z*-scores were averaged across measures, yielding an array of seven grand averaged scores (one per pair). In a final step, these grand averages were z-transformed to yield a standardized composite asynchrony index.

Session 1: Body Sway

Figure 3 shows data for maximum and minimum cross-correlations between pianists' position, velocity, and acceleration profiles for the two pieces under the two visual contact conditions averaged across duos. Maximum CC is a measure of the strongest *positive* correlation observed between the two pianists' profiles, on average, across successive cross-correlation windows. Assuming that anterior-posterior body sway approximates oscillation around an equilibrium point, maximum CC reflects the degree to which the body sway movements of the two pianists were in-phase with one another; i.e., both moved forward toward their respective keyboards together and backward together. Minimum CC is a measure of the

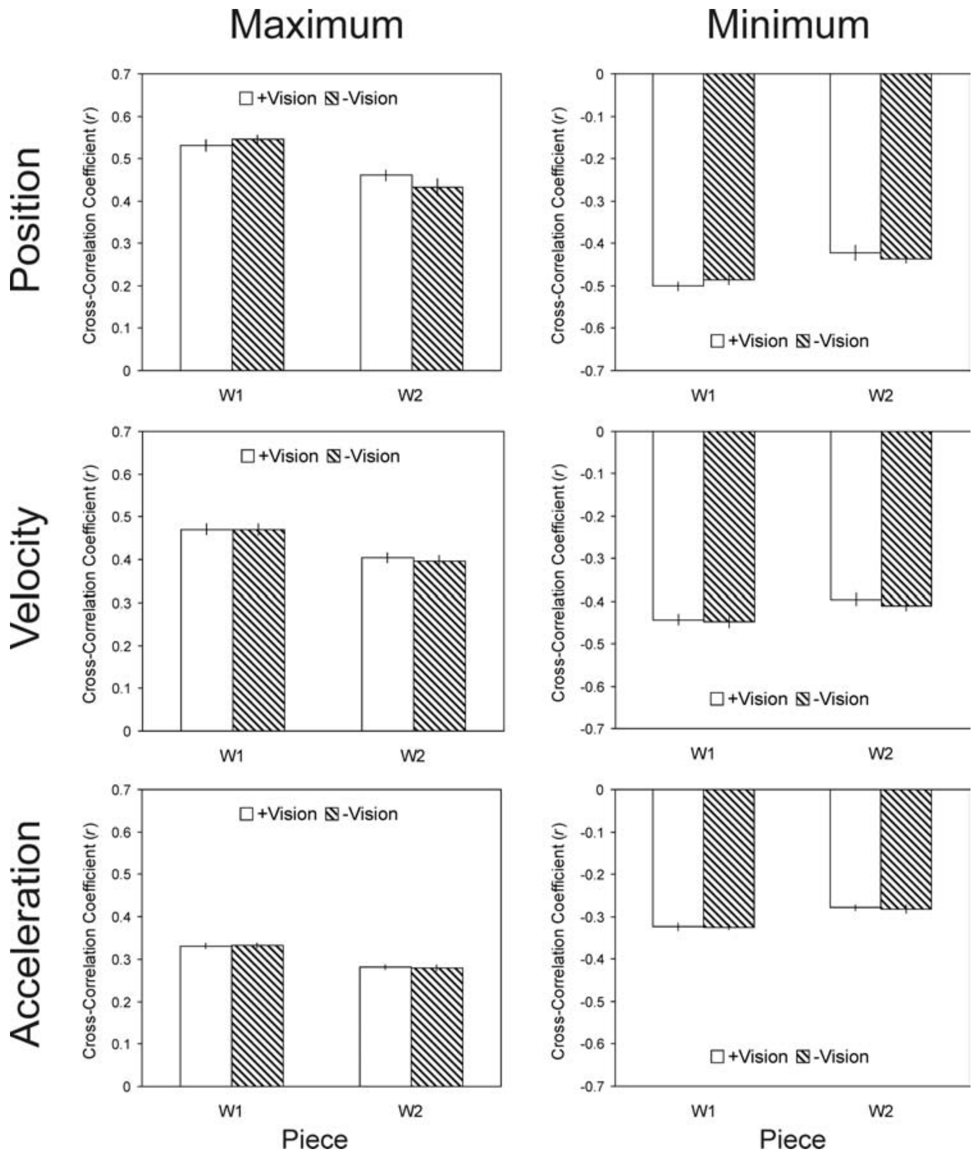

FIGURE 3. Average maximum and minimum cross-correlation coefficients representing the strength of the relationship between paired pianists' position, velocity, and acceleration profiles for the two pieces (W1; W2) under the two visual contact conditions. "+Vision" = visual contact present; "−Vision" = visual contact absent. Error bars indicate the standard error of the mean.

strongest *negative* correlation observed between the pianists' profiles across successive windows. It reflects the degree to which pianists swayed in an antiphase relation to one another; i.e., one moved forward as the other moved backward, and vice versa. Thus, maximum CC and minimum CC provide indices of two basic forms of body sway coordination—in-phase and antiphase—that correspond to the two most stable coordination modes generally observed in human motor control (see Kelso, 1995; Marsh, Richardson, & Schmidt, 2009; Oullier & Kelso, 2009; Turvey, 1990).

Separate 2 × 2 ANOVAs were run on maximum CC and minimum CC data for position, velocity, and acceleration to examine the effects of visual contact (present; absent) and piece (W1; W2). (Correlation coefficients were converted to Fisher *z* scores prior to analysis). Neither the main effect of visual contact nor the interaction between visual contact and piece was significant in any of these analyses (all ps > .05). Although a significant main effect of piece was found in each of the six analyses, where $F(1, 6)$ values ranged from 12.18 to 72.12 (ps < .05), this was most likely a statistical artifact of the difference in the size of the

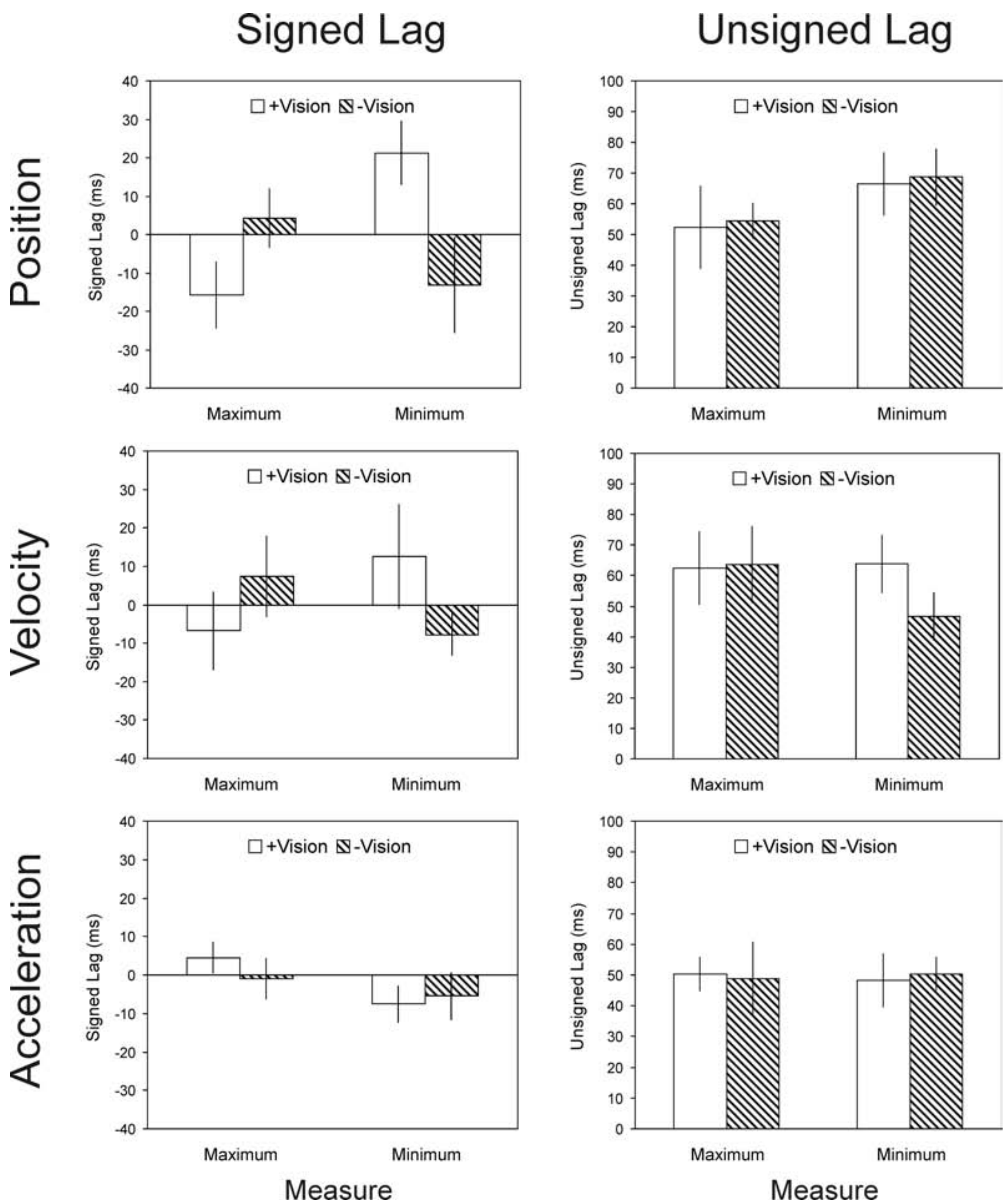

FIGURE 4. Average signed and unsigned lags (in ms) for maximum and minimum cross-correlation measures representing the strength of the relationship between paired pianists' position, velocity, and acceleration profiles, collapsed across pieces, under the two visual contact conditions. "+Vision" = visual contact present; "–Vision" = visual contact absent. Error bars indicate the standard error of the mean.

moving window used in the analysis of each piece: 3 beats for W1 and 4 beats for W2. A reanalysis of the data with a 6-beat (i.e., 2-bar) window for W1 and a 4-beat window for W2 confirmed that cross-correlations were generally weaker (i.e., maximum CC was lower and minimum CC was higher) with larger windows. This suggests that body sway coordination may have alternated frequently between in-phase and antiphase modes throughout takes.

The signed and unsigned lags at which maximum and minimum cross-correlations were observed are displayed in Figure 4. Signed lag of maximum CC is a measure of the average lag at which the tendency towards in-phase body sway was strongest and signed lag of minimum CC indicates the average lag at which the tendency towards antiphase swaying was strongest. While these signed measures are sensitive to the directionality of leader/follower relations (i.e., negative lags indicate that the primo pianist's movements were ahead of those of the secondo pianist), the unsigned lag of maximum and minimum CC gauge the magnitude of the lag but not its direction. For all measures, however, larger lags imply looser coordination.

Separate 2 × 2 × 2 ANOVAs on signed and unsigned lag measures for position, velocity, and acceleration failed to reveal any significant main effects of, or interactions between, visual contact, piece, and cross-correlation measure (maximum vs. minimum) (ps > .05). Nevertheless, two aspects of these data are noteworthy. First, the average unsigned lags of maximum and minimum CC for position, velocity, and acceleration are similar in magnitude to the median unsigned key stroke asynchrony (41 ms), with the correspondence being closest in the case of acceleration (where the average unsigned lag is 49 ms for both maximum and minimum CC). Indeed, unsigned asynchrony was correlated positively with the two acceleration measures across duos: unsigned lag of maximum CC $r(5) = .84$, $p = .017$ [$r_s = .75$, $p = .052$]; unsigned lag of minimum CC $r(5) = .90$, $p = .006$ [$r_s = .86$, $p = .014$]. Second, signed lags are generally much smaller in magnitude than unsigned lags for both maximum CC and minimum CC. This suggests that leader/follower relations in body sway coordination did not necessarily reflect the nominal assignment of pianists to primo and secondo: In some duos the primo pianist moved ahead of the secondo player, while in others the secondo player led. Indeed, signed lags of maximum and minimum CC for position, velocity, and acceleration in each visual contact condition and piece did not differ significantly from zero, as indicated by a series of one-sample t-tests (ps > .05).

Finally, range of movement (i.e., the difference between the maximum and minimum position values) in the anterior-posterior plane was examined as a function of whether pianists were playing the primo or secondo part of each piece under the two visual contact conditions. Data for this measure are shown, collapsed across piece, in Figure 5. A 2 × 2 × 2 ANOVA that tested for effects of visual contact (present; absent), piece (W1; W2), and assigned role (primo; secondo) revealed a significant main effect of visual contact, $F(1, 6) = 11.97$, $p = .013$, reflecting the fact that movements spanned a greater range of space in the anterior-posterior plane when visual contact was absent than when it was present. It may be the case that the amplitude of body sway was increased to compensate for the absence of visual coordination cues from the co-performer. Indeed, the effects of visual contact on range of movement and the composite asynchrony index (i.e., the

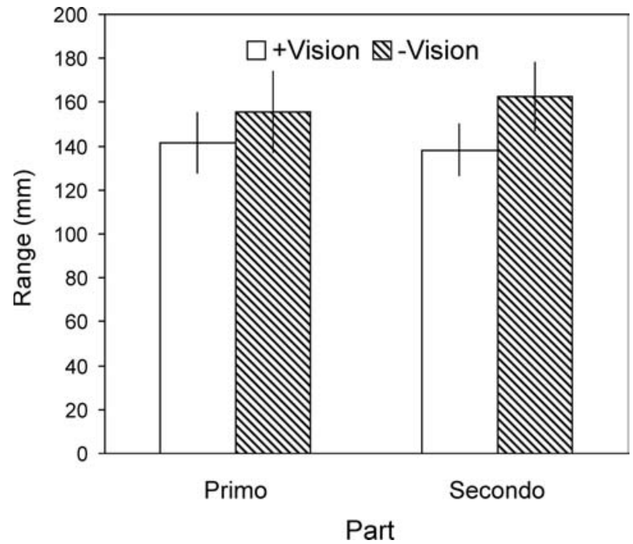

FIGURE 5. Average range of movement in the anterior-posterior plane for the two parts (primo; secondo) under the two visual contact conditions. "+Vision" = visual contact present; "–Vision" = visual contact absent. Error bars indicate the standard error of the mean.

differences between scores on each measure when visual contact was present vs. absent) were correlated positively with one another across duos: $r(5) = .70$, $p = .077$ [$r_s = .82$, $p = .023$] (Note that only the nonparametric correlation is significant at the two-tailed level). A greater range of movement may have facilitated synchronization, perhaps by enhancing the time-keeping role of body sway. The main effects of piece and role, and all interactions, fell well short of statistical significance in the ANOVA on range data (ps > .24).

Session 2: Anticipatory Auditory Imagery

Anticipatory auditory imagery was assessed by examining the acceleration profiles of finger movements in the sequence production task administered in Session 2. Average acceleration was computed for the final 5% of each movement trajectory time series. This measure of 'final acceleration' is informative about net force just prior to the finger's arrival at the target key. Final acceleration for the three taps in the three feedback conditions is shown, averaged across participants, in Figure 6. These results closely replicate the main findings of Keller et al. (2010). Final acceleration was generally negative, indicating that the finger increased in speed as it moved downward toward the target key. Moreover, final acceleration was greater (i.e., more negative) when auditory feedback was absent than when it was present

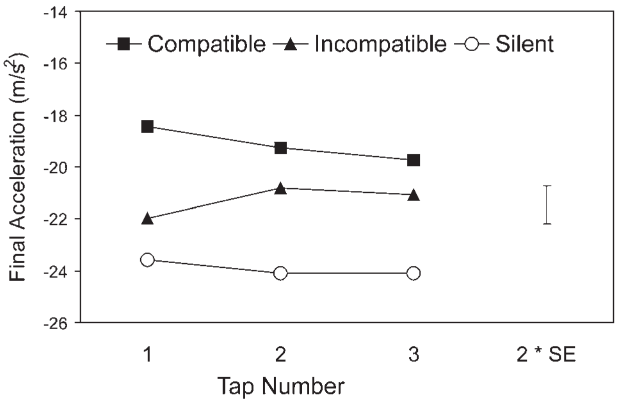

FIGURE 6. Average acceleration during the final 5% of each trajectory for the three taps in each feedback condition (compatible; incompatible; silent). The error bar on the right—labeled 2*SE—represents double the standard error of the mean.

and, in the latter conditions, when pitch contours were incompatible than when they were compatible with movement transitions on the height dimension. An ANOVA on these data revealed a significant main effect of feedback condition, $F(2, 26) = 8.01$, $p < .01$ (with the Greenhouse-Geisser correction). The main effect of tap number and the interaction between tap number and feedback condition were not significant ($ps > .47$). Thus, movements were generally characterized by less acceleration—hence, less force—when feedback was present, especially when it was compatible. The fact that the effects of feedback were observed at the first tap, $F(2, 26) = 3.94$, $p = .04$, suggests that they can be attributed to anticipatory auditory imagery.

An index of the vividness of auditory imagery for upcoming sounds was calculated based on these data for each individual participant in the following manner. First, 'general' and 'specific' effects of feedback were computed separately for average data associated with each tap number. General effects of feedback (presence versus absence) were estimated by calculating the difference in final acceleration between the silent condition and the combined conditions with tones. Specific effects of feedback (compatible versus incompatible) were estimated by calculating the difference in final acceleration between conditions in which melodic contours and movement transitions were compatible versus incompatible in terms of (pitch and spatial) height. Indices of general and specific anticipatory auditory imagery were then computed by dividing the effect of feedback at tap 1 by the mean effect of feedback at taps 2 and 3. Finally, these general and specific indices were summed to yield an aggregate index of anticipatory auditory imagery. This anticipatory auditory imagery index (AAII) is described by the equation

$$\mathrm{AAII} = \frac{\mathrm{I_c} + \mathrm{I_i} - 2\times\mathrm{I_s}}{\left(\mathrm{II_c} + \mathrm{II_i} - 2\times\mathrm{II_s}\right) + \left(\mathrm{III_c} + \mathrm{III_i} - 2\times\mathrm{III_s}\right)} + \frac{\mathrm{I_c} - \mathrm{I_i}}{\left(\mathrm{II_c} - \mathrm{II_i}\right) + \left(\mathrm{III_c} - \mathrm{III_i}\right)}$$

where tap number is represented by Roman numerals I, II, and III, and feedback condition is represented by subscripts c (compatible), i (incompatible), and s (silent). The rationale for expressing the AAII as a ratio between the effect of auditory feedback at tap 1 and the effect of feedback at taps 2 and 3 is as follows. While the effects of auditory feedback at tap 1 can be attributed to imagery, effects on the second and third taps most likely reflect the combined effects of imagery and actual auditory feedback from the preceding tap(s). The AAII, as a ratio, quantifies the potency of the effects of auditory imagery on movement force relative to the combined effects of imagery and perception on movement: the higher the value of the AAII, the more potent the effects of auditory imagery on performance.

Relations Between Asynchrony, Body Sway, and Imagery

Relations between key stroke asynchrony and body sway coordination across the seven duos were assessed by searching for the kinematic measure that was the best predictor of the composite asynchrony index. A forward linear regression analysis was run in SPSS with the composite asynchrony index as the dependent variable and maximum CC, signed lag of maximum CC, unsigned lag of maximum CC, minimum CC, signed lag of minimum CC, and unsigned lag of minimum CC for position, velocity, and acceleration (all averaged across performance conditions) as independent variables. Results indicated that signed lag of minimum CC for acceleration alone accounted for a significant proportion of the variance in key stroke asynchrony across the duos ($R^2 = .64$; adjusted $R^2 = .56$; standardized beta = .80; $p = .032$). Note that low (negative) values of signed lag of minimum CC indicate that the primo player moved ahead of the secondo player, while high (positive) values indicated that the primo lagged behind the secondo. Thus, the observed relationship between body sway and key stroke asynchrony indicates that the greater the tendency for antiphase sway movements to be characterized by the primo player lagging behind the secondo, the greater the overall asynchrony between sounds.

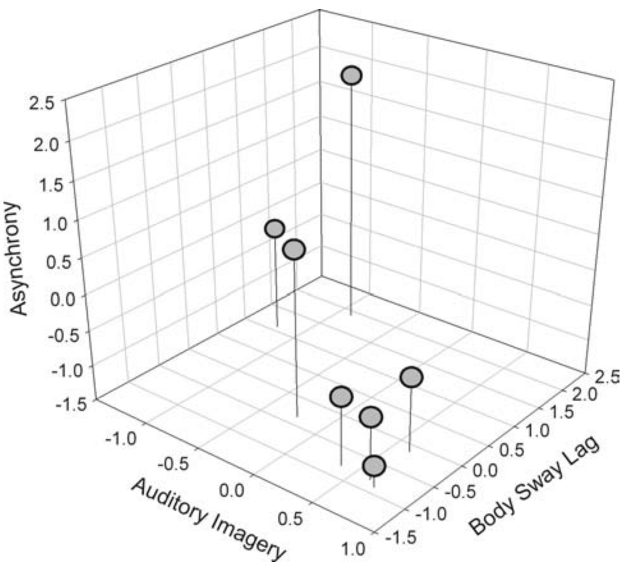

FIGURE 7. Scatter plot showing the relationship between z-transformed indices of key stroke asynchrony (ranging from low to high on the axis labelled "Asynchrony"), body sway coordination (ranging from primo lead to secondo lead on the axis labelled "Body Sway Lag"), and anticipatory auditory imagery (ranging from low to high on axis labelled "Auditory Imagery") for the seven pairs of pianists. For the AAII, each data point represents the mean score for a pair of pianists. All values are in Z units.

Finally, to address the main aim of the study, pairwise intercorrelations between the AAII (auditory imagery), the composite asynchrony index (asynchrony), and signed lag of minimum CC for acceleration (body sway lag) were examined. The AAII scores of the two pianists in each duo were averaged prior to this analysis. The relations between the three measures are depicted in Figure 7. In addition to the positive correlation between asynchrony and body sway lag detected by the regression analysis reported above—$r(5) = .80$, $p = .016$ [$r_s = .89$, $p = .003$]—duo mean AAII was found to be negatively correlated with asynchrony, $r(5) = -.721$, $p = .034$ [$r_s = -.86$, $p = .007$], and body sway lag, $r(5) = -.87$, $p = .006$ [$r_s = -.86$, $p = .007$] (all p-values are one-tailed).[5] Thus, duos comprised of pianists with relatively high anticipatory auditory imagery indices were relatively well coordinated in terms of key strokes and body sway. The need to test whether this apparent benefit of auditory imagery was most evident in the absence of visual contact was obviated by the fact that visual contact had little effect on our measures of ensemble coordination (apart from producing an increase in CVs).

[5]To test the specificity of the observed relationship between anticipatory auditory imagery and ensemble coordination, the correlation analysis was re-run with an alternative index reflecting the ratio between movement force at the first and subsequent taps irrespective of feedback condition (i.e., force at tap 1 was divided by average force at taps 2 and 3) instead of the AAII. This alternative index, which can be seen as a measure of tactile imagery, was not correlated significantly with asynchrony, $r(5) = .50$, $p = .128$ [$r_s = .32$, $p = .241$], or body sway lag, $r(5) = .54$, $p = .103$ [$r_s = .57$, $p = .090$], at the one-tailed level.

Discussion

The role of anticipatory auditory imagery in musical ensemble coordination was investigated in a sample of skilled pianists. The vividness of imagery for upcoming musical sounds was assessed in each individual by a task that required the production of rhythmic movement sequences either with or without auditory feedback tones. Ensemble coordination was assessed by examining temporal relations between body movements and sounds in duos comprised of pianists from the same sample. Sound synchrony—which was quantified using a composite measure that took into account the magnitude and variability of key stroke asynchronies—was correlated with anterior-posterior body sway coordination across duos in a manner that depended upon leader/follower relations. Moreover, these differences in ensemble coordination could be predicted on the basis of individual differences in anticipatory auditory imagery. Although ensemble coordination was not affected markedly by whether or not pianists were in visual contact, the variability of key stroke asynchronies was higher when visual contact was present than when it was absent and the overall range of body sway movement was greatest in the absence of visual contact. These findings will be discussed in turn.

Relations Between Sound Synchrony and Body Sway

Evidence for interdependencies between sound synchrony and body sway coordination was found when MIDI key stroke data and kinematic motion capture data were compared. Specifically, we observed a positive correlation between the composite measure of key stroke asynchrony and a kinematic measure reflecting the lag separating the acceleration profiles of primo and secondo pianists' antiphase anterior-posterior sway movements. Thus, temporal relations at the level of finger movements were systematically related to larger scale spatiotemporal relations between body movements, and acceleration proved to be the most informative kinematic measure in this regard (cf. Goebl & Palmer, 2009). The precise nature of this relationship suggests that the direction of leader/follower relations in body sway influences overall sound synchrony. Sound synchrony was found to be high (i.e., composite asynchrony scores were low) to the extent that the movements of the pianist assigned to the primo part preceded those of the secondo pianist.

This result is remarkable because, while the tendency toward primo lead in body sway varied between duos, evidence for primo lead was found in signed key stroke asynchronies without exception across duos. Leader/follower relations in key strokes were thus dissociated from leader/follower relations in body sway. It may be the case that optimal conditions for overall sound synchrony are characterized by the congruence of leader/follower relations in key strokes and body sway coordination. Such congruence may facilitate the maintenance of systematic deviations from perfect synchrony that are the hallmark of much ensemble music (see Hove, Keller, & Krumhansl, 2007; Iyer, 2002; Keil, 1995). Congruence between key strokes and body sway is not a trivial inevitability, however, because finger movements are primarily instrumental actions that are necessary for sound production on the piano, while body sway is to a greater degree ancillary and free to be guided by a performer's expressive intentions in a relatively unconstrained manner (see Cadoz & Wanderley, 2000; Nusseck & Wanderley, 2009).

At this point, some brief comments on the role of body sway movements in ensemble performance are in order. We assume that body sway may fulfill a dual function by facilitating both the temporal regulation of (individual and joint) actions and the communication of performers' expressive intentions. Each of these functions may be exercised with varying degrees of intermittency, success, and dependence upon the other function. According to these assumptions, signals generated by body sway movements—trajectories in the anterior-posterior plane, for example—carry varying amounts of convolved information about basic timing and musical expression. The type of information that is extracted from the signals via statistical analysis depends upon the type of technique that is applied to them. The windowed cross-correlation technique that we employed estimates the similarity between the time courses of two such anterior-posterior sway signals. This technique thus quantifies the degree of temporal coordination between paired individuals' swaying movements without distinguishing between information related to time-keeping and expressive communication. Partitioning such information presents a significant empirical challenge. One potentially applicable method involves manipulating the performers' expressive intentions via instructions (e.g., Davidson, 1994), and then using principal components analysis or related approaches (e.g., Leman & Naveda, 2010; Toiviainen, Luck, & Thompson, 2010) to decompose performers' body sway trajectories into components that change across levels of expression versus components that are relatively uniform (and therefore likely to be related to basic temporal coordination).

Anticipatory Auditory Imagery and Ensemble Coordination

The main finding of the current study concerns the relationship between individual differences in anticipatory auditory imagery and ensemble coordination. Evidence for such a relationship was found despite the fact that imagery and coordination were assessed independently using tasks that are superficially unrelated. We interpret the results to suggest that the degree to which images of one's own upcoming sounds affected motor control in the sequence production task was related to the use of imagery to anticipate a co-performer's sounds during interpersonal action coordination in piano duo performance.

This interpretation is consistent with the notion that accurate musical synchronization relies upon predictive mechanisms that are grounded in online action simulation and internal models instantiated in the central nervous system (see Keller, 2008). On this account, a performer simulates a co-performer's actions—in real time, but slightly ahead of action execution—using two types of internal models that represent sensorimotor transformations between bodily states and events in the immediate environment. *Forward models* represent the causal relationship between efferent motor signals and their ultimate effects on the body and the environment. *Inverse models* represent sensorimotor transformations from desired action outcomes (sounds, in the case of music) to the motor commands that give rise to these outcomes (Wolpert, Miall, & Kawato, 1998). Anticipatory auditory imagery in musical contexts involves bringing sounds to mind by running inverse models.

It has been proposed that when interpersonal coordination is required, as it is in ensembles, paired forward and inverse models that serve one's own actions are coupled with a second class of paired forward-inverse models that specialize in anticipating others' sounds (Keller, 2008; cf. Wolpert et al., 2003). Such coupling between 'own' and 'other' internal models at the level of covert action simulation facilitates the accurate coordination of overt actions by improving the efficiency of temporal error correction processes. Specifically, interpersonal action simulation allows corrections to be made on the basis of the anticipated relation between musical parts (because the proposed simulations run ahead of execution), rather than only in response to the perception of actual discrepancies between the timing of one's own and others' actions. Consistent with the notion that internal models require training to 'learn' the sensorimotor transformations that they embody (Wolpert et al., 1998, 2003), we assume that the reliability and efficiency of forward and inverse models of one's own and others' musical actions develop with ensemble playing. Individual differences in the operation of these internal models may therefore be related to differences in the quantity and quality of ensemble experience. Relevant aspects of such experience include familiarity with musical genres, specific musical pieces, particular instruments, and the stylistic idiosyncrasies of co-performers.

Recent studies that investigated interrelations between auditory imagery and basic sensorimotor synchronization abilities in relatively large samples of musicians provide further evidence for a link between anticipatory auditory imagery and temporal prediction in musical coordination (Pecenka & Keller, 2009a, 2009b). In these studies, adaptive threshold estimation procedures were employed to measure the acuity of single-tone pitch images and images of event timing in short auditory sequences that contained tempo changes. Basic sensorimotor synchronization skills were assessed in tasks that required finger tapping in time with longer sequences containing similar tempo changes. The degree to which individuals predicted versus tracked tempo changes was estimated based on lag-0 and lag-1 cross-correlations between intertap intervals and interonset intervals in the pacing sequences (cf. Repp, 2002). Results indicated that the accuracy and stability of sensorimotor synchronization depended on the degree of prediction, and that high prediction was associated (weakly) with high pitch imagery acuity and (more strongly) with high temporal imagery acuity. Although the focus of these studies and the present investigation has been on anticipatory auditory imagery, a complete account of predictive processes in ensemble performance ultimately will need to address how such top-down processes interact with bottom-up temporal expectancies generated on the basis of the perception of actual sounds (see Janata, 2001; Janata & Paroo, 2006; Vuust, Ostergaard, Pallesen, Bailey, & Roepstorff, 2009).

Effects of Visual Contact

Ensemble coordination was not affected markedly by whether pianists were in visual contact in the current study. The only hint of such an effect was seen in higher variability of asynchronies when visual contact was present than when it was absent. One possible (ad hoc) explanation of this unexpected finding is that pianists may have been encouraged to indulge in greater expressive timing variation when visual information was available about their partner's intentions concerning the future course of the music. The resultant local tempo modulations could cause an increase in the variability of asynchronies, as can be observed in sensorimotor synchronization studies that compare finger tapping with isochronous metronomes or quantized music to tapping with nonisochronous metronomes or expressively timed music (e.g., Pecenka & Keller, 2009a; Rankin, Large, & Fink, 2009).

The lack of beneficial effects of visual contact on basic ensemble coordination is perhaps not surprising. The requirement to perform in synchrony with invisible co-performers is not uncommon (in recording studios or via the internet, for example), and Keller et al. (2007) found that pianists were able to synchronize well with recordings of musical pieces that were similar to (or the same as, in one case) the pieces used in the current study. Visual contact may not be essential to basic temporal coordination in the context of these pieces due to their richness in auditory cues and the fact that they have regular metric structures devoid of large tempo changes. Previous research addressing the role of visual cues in ensemble coordination has emphasized their role in the communication of expressive intentions (e.g., Clayton, 2007; Davidson, 2009; Moran, 2007). A logical extension of the present study could therefore involve examining the effects of visual contact upon the degree to which expressive parameters of pianists' performances are matched within duos.

Although visual contact did not affect body sway coordination per se, the amplitude of anterior-posterior sway movements was generally greater when visual contact was absent than when it was present. This result finds a parallel in the classic observation that body sway increases when visual input is restricted in studies of postural control (e.g., Edwards, 1946; Hafström, Franson, Karlberg, Ledin, & Magnusson, 2002). The increase in movement amplitude observed in the present study may constitute an attempt to compensate for the absence of visual coordination cues from the co-performer by increasing kinesthetic feedback from one's own movement. Such compensation may facilitate coordination to the extent that increased kinesthetic feedback enhances the time-keeping function of body sway.

Conclusions

The current study identified systematic relations between sound synchrony and interpersonal body sway in piano duos and, furthermore, yielded evidence that differences in the quality of basic ensemble coordination are related to differences in the vividness of individuals' anticipatory images of upcoming sounds. The analysis of leader/follower relations revealed that—although key strokes produced by the pianist assigned to the primo part typically preceded nominally synchronous key strokes produced by the secondo player in each duo—anterior-posterior swaying movements of the primo player were not ahead of the secondo player's movements consistently across duos. The finding that overall sound synchrony was high to the extent that the primo player moved ahead of the secondo player suggests that congruency between leader/follower relations in key strokes and body sway may be optimal for basic ensemble coordination. Visual contact is apparently not necessary for achieving such optimal relations. Vivid anticipatory auditory imagery, however, may assist this process by enhancing the operation of internal models that simulate one's own and others' actions during interpersonal coordination.

The overarching goal of a broader research project linked to this study is to identify the cognitive/motor skills that mediate interpersonal interaction in musical ensembles, and to explore how these 'ensemble skills' mesh to determine the quality of musical coordination under naturalistic performance conditions. The present findings suggest that examining the relationship between individual differences in skills such as anticipatory auditory imagery and basic ensemble coordination is a promising approach to meeting this goal. However, to ascertain the true value of this approach, it will be necessary in future research to work with larger samples of musicians than the sample employed in the current study, and to investigate coordination in different types of ensembles playing different styles of music. The finding that body sway coordination provides a valid index of ensemble cohesion is therefore welcome not only for its theoretical implications concerning the notion of music as an embodied phenomenon (see Leman, 2007), but also on practical grounds. Body sway coordination is fairly neutral to the instruments being played, is readily applicable to vocal performance, and—because it is a continuous measure—can be used to sidestep difficulties in identifying which sounds are intended to be played synchronously in complex, interlocking rhythms and when no score is available (e.g., in improvised music). Nevertheless, issues such as the degree to which measures of body sway coordination correspond to observers' judgments of ensemble cohesion in auditory and audio-visual displays remain to be resolved. Only then can we hope to understand how temporal relations in the auditory domain and spatiotemporal relations in interpersonal movement kinematics coalesce to determine the quality of musical ensemble coordination.

Author Note

Mirjam Appel worked as an intern on this project during the period August-September 2006. She was responsible for collecting and pre-processing the piano duo data. Portions of the results of the current study have been reported at conferences (Keller, 2007; Keller & Appel, 2008) and in a book chapter (Keller, 2008). The authors thank Wenke Moehring, Janne Richter, Nadine Seeger, and Kerstin Traeger for assisting with data collection; Stephan Liebig and Regine Steinke for help in preparing the figures; and Jan Bergmann, Henrik Grunert, Rick Hegewald, and Andreas Romeyke for technical assistance. We are also grateful to Wolfgang Prinz for making available the resources required for the study, and for fostering the intellectual environment in which it took place. Finally, we thank Bruno Repp, Petri Toiviainen, Jessica Phillips-Silver, and two anonymous reviewers for helpful comments on an earlier version of this paper.

Correspondence concerning this article should be addressed to Peter Keller, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstr. 1a D-04103 Leipzig, Germany. E-MAIL: keller@cbs.mpg.de

References

Bekkering, H., De Bruijn, E., Cuijpers, R. H., Newman-Norlund, R., Van Schie, H. T., & Meulenbroek, R. G. J. (2009). Joint action: Cognitive and neural mechanisms supporting human interactions. *Topics in Cognitive Science, 1,* 340–352.

Blank, M., & Davidson, J. W. (2007). An exploration of the effects of musical and social factors in piano duo collaborations. *Psychology of Music, 35,* 213–230.

Cadoz, C., & Wanderley, M. M. (2000). Gesture—music. In M. M.Wanderley & M. Battier (Eds.), *Trends in gestural control of music* (pp. 71–94). Paris: IRCAM/Centre Pompidou.

Castellano, G., Camurri, A., Mortillaro, M., Scherer, K., & Volpe, G. (2008). Espressive gesture and music: Analysis of emotional behaviour in music performance. *Music Perception, 25,* 103–119.

Clarke, E. F., & Davidson, J. W. (1998). The body in music as mediator between knowledge and action. In W. Thomas (Ed.), *Composition, performance, reception: Studies in the creative process in music* (pp. 74–92). Oxford, UK: Oxford University Press.

Clayton, M. (2007). Time, gesture and attention in a khyal performance. *Asian Music, 38,* 71–96.

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

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

Davidson, J. W. (2009). Movement and collaboration in musical performance. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (pp. 364–376), Oxford, Oxford University Press.

Davidson, J. W., & Good, J. M. M. (2002). Social and musical co-ordination between members of a string quartet: An exploratory study. *Psychology of Music, 30,* 186–201.

Davidson, J. W., & King, E. C. (2004). Strategies for ensemble practice. In A. Williamon (Ed.), *Enhancing musical performance* (pp. 105–122). Oxford, UK: Oxford University Press.

Edwards, A. S. (1946). Body sway and vision. *Journal of Experimental Psychology, 36,* 526–535.

Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Sciences, 8,* 396–403.

Ginsborg, J., Chaffin, R., & Nicholson, G. (2006). Shared performance cues in singing and conducting: A content analysis of talk during practice. *Psychology of Music, 34,* 167–192.

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

Goodman, E. (2002). Ensemble performance. In J. Rink (Ed.), *Musical performance: A guide to understanding* (pp. 153–167). Cambridge, UK: Cambridge University Press.

Hafström, A., Fransson, P-A., Karlberg, M., Ledin, T., & Magnusson, M. (2002). Visual influence on postural control, with and without visual motion feedback. *Acta Oto-Laryngologica, 122,* 392–397.

Hove, M. J., Keller, P. E., & Krumhansl, C. L. (2007). Sensorimotor synchronization with chords containing tone-onset asynchronies: The role of P-centers. *Perception and Psychophysics, 69,* 699–708.

Iyer, V. (2002). Embodied mind, situated cognition, and expressive microtiming in African-American music. *Music Perception, 19,* 387–414.

Janata, P. (2001). Neurophysiological mechanisms underlying auditory image formation in music. In R. I. Godøy & H. Jørgensen (Eds.), *Elements of musical imagery* (pp. 27–42). Lisse, Netherlands: Swets & Zeitlinger Publishers.

Janata, P., & Paroo, K. (2006). Acuity of auditory images in pitch and time. *Perception and Psychophysics, 68,* 829–844.

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

Keil, C. (1995). The theory of participatory discrepancies: A progress report. *Ethnomusicology, 39,* 1–19.

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

Keller, P. E. (2007). Musical ensemble synchronisation. In E. Schubert, K. Buckley, R. Eliott, B. Koboroff, J. Chen, & C. Stevens (Eds.), *Proceedings of the International Conference on Music Communication Science* (pp. 80–83). Sydney, Australia.

Keller, P. E. (2008). Joint action in music performance. In F. Morganti, A. Carassa, & G. Riva (Eds.), *Enacting intersubjectivity: A cognitive and social perspective to the study of interactions* (pp. 205–221). Amsterdam: IOS Press.

Keller, P. E., & Appel, M. (2008, August). *Coordination of body movements and sounds in musical ensemble performance.* Paper presented at the 10th International Conference on Music Perception and Cognition. Hokkaido University, Sapporo, Japan.

Keller, P. E., & Burnham, D. K. (2005). Musical meter in attention to multipart rhythm. *Music Perception, 22,* 629–661.

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*, 508–513.

Keller, P. E., & Koch, I. (2006). The planning and execution of short auditory sequences. *Psychonomic Bulletin and Review, 13,* 711–716.

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

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

Kelso, J. A. S. (1995). *Dynamic patterns: The self-organization of brain and behavior.* Cambridge, MA: MIT Press.

King, E. (2006). The roles of student musicians in quartet rehearsals. *Psychology of Music, 34,* 263–283.

Knoblich, G., & Sebanz, N. (2008). Evolving intentions for social interaction: From entrainment to joint action. *Philosophical Transactions of the Royal Society B*, *363*, 2021–2031.

Leman, M. (2007). *Embodied music cognition and mediation technology.* Cambridge, MA: MIT Press.

Leman, M., & Naveda, L. (2010). Basic gestures as spatio-temporal reference frames for repetitive dance/music patterns in Samba and Charleston. *Music Perception, 28,* 71–91.

Lindenberger, U., Li, S-C., Gruber, W., & Müller, V. (2009). Brains swinging in concert: Cortical phase synchronization while playing guitar. *BMC Neuroscience, 10:22.* Retrieved from http://www.biomedcentral.com/1471–2202/10/22

Luck, G., & Sloboda, J. A. (2009). Spatio-temporal cues for visually mediated synchronization. *Music Perception, 26,* 465–473.

Luck, G., & Toiviainen, P. (2006). Ensemble musicians' synchronization with conductors' gestures: An automated feature-extraction analysis. *Music Perception, 24, 189*–200.

Maduell, M., & Wing, A. M. (2007). The dynamics of ensemble: The case for flamenco. *Psychology of Music, 35,* 591–627.

Marsh, K. L., Richardson, M. J., & Schmidt, R. C. (2009). Social connection through joint action and interpersonal coordination. *Topics in Cognitive Science*, *1*, 320–339.

Merker, B., Madison, G., & Eckerdal, P. (2009). On the role and origin of isochrony in human rhythmic entrainment. *Cortex, 45,* 4–17.

Moran, N. (2007). *Measuring musical interaction: Analysing communication in embodied musical behaviour* (Unpublished doctoral dissertation). Open University, United Kingdom.

Nusseck, M., & Wanderley, M. M. (2009). Music and motion—How music-related ancillary body movements contribute to the experience of music. *Music Perception*, *26*, 335–353.

Oullier, O., de Guzman, G. C., Jantzen, K. J., Lagarde, J., & Kelso, J. A. S. (2008). Social coordination dynamics: Measuring human bonding. *Social Neuroscience, 3*, 178–192.

Oullier, O., & Kelso, J. A. S. (2009). Coordination from the perspective of social coordination dynamics. In K. Nowak (Ed.), *Encyclopedia of complexity and systems science*. Berlin: Springer-Verlag.

Overy, K., & Molnar-Szakacs, I. (2009). Being together in time: Musical experience and the mirror neuron system. *Music Perception, 26,* 489–504.

Pecenka, N., & Keller, P. E. (2009a). Auditory pitch imagery and its relationship to musical synchronization. *Annals of the New York Academy of Sciences, 1169*, 282–286.

Pecenka, N., & Keller, P. E. (2009b). The relationship between auditory imagery and musical synchronization abilities in musicians. In J. Louhivuori, T. Eerola, S. Saarikallio, T. Himberg, & P.-S. Eerola (Eds.), *Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music* (pp. 409–414). Jyväskylä, Finland.

Rankin, S. K., Large, E. W., & Fink. P. W. (2009). Fractal tempo fluctuation and pulse prediction. *Music Perception, 26,* 401–413.

Rasch, R. A. (1988). Timing and synchronization in ensemble performance. In J. A. Sloboda (Ed.), *Generative processes in music: The psychology of performance, improvisation, and composition* (pp. 70–90). Oxford, UK: Clarendon Press.

Repp, B. H. (2002). The embodiment of musical structure: Effects of musical context on sensorimotor synchronization with complex timing patterns. In W. Prinz & B. Hommel (Eds.), *Common mechanisms in perception and action: Attention and Performance XIX* (pp. 245–265). Oxford, UK: Oxford University Press.

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

Repp, B. H. (2006). Musical synchronization. In E. Altenmüller, M. Wiesendanger, & J. Kesselring (Eds.), *Music, motor control, and the brain* (pp. 55–76). Oxford, UK: Oxford University Press.

Repp, B. H., & Keller, P. E. (2004). Adaptation to tempo changes in sensorimotor synchronization: Effects of intention, attention, and awareness. *Quarterly Journal of Experimental Psychology, 57A,* 499–521.

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

Richardson, M. J., Marsh, K. L., & Schmidt, R. C. (2005). Effects of visual and verbal information on unintentional interpersonal coordination. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 62–79.

Schögler, B. (1999–2000). Studying temporal co-ordination in jazz duets. *Musicae Scientiae, Special Issue 1999–2000,* 75–91.

Schögler, B. (2003). The pulse of communication in improvised music. In R. Kopiez, A. C. Lehmann, I. Wolther, & C. Wolf (Eds.), *Proceedings of the 5th Triennal ESCOM Conference* (pp. 574–577). Hannover, Germany.

Schubotz, R. I. (2007). Prediction of external events with our motor system: Towards a new framework. *Trends in Cognitive Sciences, 11,* 211–218.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences, 10,* 70–76.

Sebanz, N., & Knoblich, G. (2009). Prediction in joint action: What, when, and where. *Topics in Cognitive Science, 1,* 353–367.

Shaffer, L. H. (1984). Timing in solo and duet piano performances. *Quarterly Journal of Experimental Psychology, 36A,* 577–595.

Shockley, K., Santana, M., & Fowler, C. A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance, 29,* 326–332.

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

Toiviainen, P., Luck, G., & Thompson, M. (2010). Embodied meter: Hierarchical eigenmodes in *music-induced movement. Music Perception, 28,* 59–70.

Trusheim, W. H. (1993). Audiation and mental imagery: Implications for artistic performance. *The Quarterly Journal of Music Teaching and Learning, 2,* 139–147.

Turvey, M. T. (1990). Coordination. *American Psychologist, 45,* 938–953.

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

Vuust, P., Ostergaard, L., Pallesen, K. J., Bailey, C., & Roepstorff, A. (2009). Predictive coding of music—Brain responses to rhythmic incongruity. *Cortex, 45,* 80–92.

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

Wilson M., & Knoblich G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological Bulletin, 131,* 460–473.

Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society B, 358,* 593–602.

Wolpert, D. M., Miall, R. C., & Kawato, M. (1998). Internal models in the cerebellum. *Trends in Cognitive Sciences, 2,* 338–347.

Appendices

The piece referred to as W1: No. 1 (Moderato) in D major (Bars 1–35) from Huit Pièces (Eight Pieces), op. 60, by Carl Maria von Weber.

The piece referred to as W2: No. 3 (Andante con Variazioni) from Six Pièces, op. 10, by Carl Maria von Weber.