



Simultaneous self-other integration and segregation support real-time interpersonal coordination in a musical joint action task

Hannah Liebermann-Jordanidis^a, Giacomo Novembre^b, Iring Koch^c, Peter E. Keller^{d,*}

^a Department of Medical Psychology|Neuropsychology and Gender Studies & Center for Neuropsychological Diagnostics and Intervention (CeNDI), Faculty of Medicine and University Hospital Cologne, Germany

^b Italian Institute of Technology (IIT), Rome, Italy

^c Cognitive and Experimental Psychology, Rheinisch-Westfälische Technische Hochschule Aachen, Germany

^d MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Australia

ARTICLE INFO

Keywords:

Joint action
Self-other distinction
Interpersonal coordination
Music
Movement

ABSTRACT

The ability to distinguish between an individual's own actions and those of another person is a requirement for successful joint action, particularly in domains such as group music making where precise interpersonal coordination ensures perceptual overlap in the effects of co-performers' actions. We tested the hypothesis that such coordination benefits from simultaneous integration and segregation of information about 'self' and 'other' in an experiment using a musical joint action paradigm. Sixteen pairs of individuals with little or no musical training performed a dyadic synchronization task on a pair of electronic music boxes. The relationship between pitches produced by paired participants (same vs. different) and the relationship between movement frequencies required to trigger synchronous tones (congruent vs. incongruent) were varied in a repeated measures design. The results indicate that interpersonal coordination was most accurate when sounds were different in pitch but movement frequency was congruent. Under other conditions, participants often drifted apart, resulting in poor coordination, especially with same sounds and incongruent movements across co-performers. These findings suggest that interpersonal coordination was facilitated when simultaneous self-other integration and segregation occurred across sensory modalities in an asymmetrical manner where pitch relations favoured segregation via auditory streaming while movement congruence favoured integration via visuo-motor coupling. Such self-other representational balance may enable co-performers to maintain autonomous control while attending, anticipating, and adapting to each other's timing when joint action requires precise temporal coordination.

1. Introduction

Many forms of joint action—social interactions executed by pairs or groups of people pursuing a common goal—require the precise coordination of each individual's actions in space and time (Sebanz et al., 2006). Everyday life instances of such joint activities range from playing football in a team to making music in a duet. Irrespective of the number of performers involved, the execution of joint actions involves action planning and action control at intrapersonal and interpersonal levels. Intrapersonal coordination is necessary to produce one's own actions with correct sequencing and timing (Keller & Koch, 2006, 2008), while interpersonal coordination is a matter of aligning one's own and another person's actions in order to achieve the collective goal (Coey et al., 2012; Heggli et al., 2019; Keller & Repp, 2008). In musical ensembles, for

example, individual performers typically produce sequences of tones with specific rhythmic patterning while synchronizing their output across individuals to produce a cohesive collective sound. Even in the general population, music is a useful domain for studying the mechanisms supporting real-time joint action to the extent that musical conventions allow aspects of the interaction to be readily controlled while maintaining ecological naturalness (D'Ausilio et al., 2015).

Musical joint action typically involves precise yet flexible interpersonal coordination between two or more performers playing separate parts (Keller, 2008). Such rhythmic interpersonal coordination requires co-performers to anticipate, attend, and adapt to the timing of each other's actions (Keller, 2014; Keller et al., 2014). It has been argued that these processes function effectively to the extent that information related to one's own part, others' parts, and the joint action outcome can

* Corresponding author at: MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Locked Bag 1797, Penrith, NSW 2751, Australia.
E-mail address: p.keller@westernsydney.edu.au (P.E. Keller).

<https://doi.org/10.1016/j.actpsy.2021.103348>

Received 25 February 2020; Received in revised form 21 April 2021; Accepted 24 May 2021

Available online 28 May 2021

0001-6918/© 2021 The Authors.

Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

be integrated while maintaining a distinction between self and other (Keller et al., 2016). Self-other integration enables joint action outcomes to be monitored and evaluated against shared performance goals (Loehr et al., 2013; Sebanz et al., 2006) while self-other segregation allows agency attribution and autonomous intrapersonal action control (De Jaegher & Di Paolo, 2007; Kahl & Kopp, 2018; Pacherie, 2012). Regulating the balance of self-other integration and segregation therefore allows goals at the individual (intrapersonal) and group (interpersonal) levels to be pursued simultaneously (Heggli et al., 2019; Keller et al., 2016; MacRitchie et al., 2018).

It has been proposed that such regulation is underpinned by a 'joint internal model' that controls the proportion of attention allocated to psychological representations of 'self' and 'other' (Harry & Keller, 2019; van der Steen & Keller, 2013). Studies with musical experts suggest that this weighting process is influenced by intrinsic factors such as co-performer knowledge and the compatibility of expressive goals (MacRitchie et al., 2018; Novembre et al., 2016). The present study extends the investigation of self-other integration and segregation in musical joint action by testing untrained individuals and examining the role of two extrinsic factors that potentially affect self-other representational balance: auditory pitch relations and movement congruency.

1.1. Auditory pitch relations

Pitch relations between separate parts in group music-making span situations where co-performers play the same pitch (unison) to scenarios involving different pitches (e.g., playing in harmony or in octaves). Research on auditory scene analysis indicates that pitch distance affects whether sounds are perceived as coming from the same or different sources, with large pitch separation encouraging the perception of sound sequences as segregated auditory streams (Bregman, 1990). While auditory streaming effects can occur automatically in a bottom-up fashion (Müller et al., 2005), the ability to hear sequences as segregated or integrated is nevertheless subject to some degree of top-down control through the allocation of attention (Billig & Carlyon, 2016).

Interpersonal coordination during musical joint action may therefore be influenced by how conducive the pitch relationship between parts is to the process of dividing attention between one's own sounds and the overall sonic texture that results when all parts are integrated (Keller, 2001). Accordingly, Keller and Repp (2008) found that participants were better able to perform a challenging multilevel coordination task—finger tapping in antiphase with metronomic tone sequences (inter-agent coordination) while alternating between the two hands (intra-personal coordination)—when feedback tones triggered by taps were close in pitch to, but nevertheless distinct from, the metronome. This suggests that self-other representational balance was best with pitch relations that were readily integrated while being distinct enough to be perceived as separate auditory streams.

1.2. Movement congruency

While auditory information is usually paramount in music performance, co-performers' body movements provide visual cues that can have functional benefits for interpersonal coordination in ensembles (Bishop et al., 2019; Chang et al., 2019; Davidson & Broughton, 2016; Glowinski et al., 2013; Hilt et al., 2019; Kawase, 2014). For instance, watching the continuous trajectories of musicians' movements as they play (e.g., motion of a pianist's fingers or a violinist's bowing arm) can assist one performer to predict the timing of a co-performer's sounds (Bishop & Goebel, 2015; Wöllner & Cañal-Bruland, 2010). Such facilitation may arise through a general mechanism whereby action observation evokes an internal process of motor simulation (i.e., covert activation of brain regions that would be involved in executing the action) that helps in generating predictions about the action's outcome (i.e., tones) (Jeannerod, 2001; Kilner et al., 2004; Schubotz, 2007; Wilson & Knoblich, 2005).

To the extent that motor simulation plays a role in planning one's own actions as well as predicting other's actions (Jeannerod, 2001), it requires simultaneous self-other integration and segregation in the sensory-motor system (Knoblich & Sebanz, 2008; Schütz-Bosbach et al., 2006; Sebanz & Knoblich, 2009). Self-other representational balance at this sensory-motor level is presumably influenced by the congruency of co-performers' movements during visually mediated interaction. Consistent with this assumption, previous research has demonstrated that observing another person performing an action can facilitate the imitative production of the same action (Blakemore & Frith, 2005; Brass & Heyes, 2005; Sebanz & Knoblich, 2009) but interfere with the concurrent production of different actions (Kilner et al., 2003). This finding holds across a range of task domains and is generally taken as indicative of tight links between perception and action that drive motor simulation (Cracco et al., 2018; Knoblich et al., 2011; Novembre & Keller, 2014; Prinz, 1997; Su & Keller, 2020).

In rhythmical joint action tasks, perception-action links facilitate the coupling of co-actors' periodic movements. Studies of visuo-motor coordination show that such interpersonal entrainment is most stable when the movements of co-actors are congruent in terms of their spatio-temporal dynamics (see Coey et al., 2012; Schmidt & Richardson, 2008). Tasks that require paired individuals to swing handheld pendulums or rock in chairs have found that in-phase relations, where the relative phase between movement cycles is close to zero across individuals, are more stable than other types of phase relation (Richardson et al., 2005; Schmidt et al., 1990; Schmidt et al., 1998). When co-actors have different preferred movement frequencies (due to natural individual variation or experimental manipulations), visual-motor coupling is weaker and temporal leader-follower relations where the faster individual leads the slower individual can emerge (Richardson et al., 2007; Schmidt et al., 1998). This suggests that self-other integration and segregation can be influenced by visuo-motor coupling strength, which varies as a function of movement congruence between co-performers.

1.3. Study overview & hypotheses

In the current study, we conducted an experiment to investigate simultaneous self-other integration and segregation in auditory and visual modalities during real-time interpersonal coordination. We employed a musical joint action task that required pairs of individuals to play a melody together on custom-made 'E-music boxes'—electromechanical musical instruments that transform rotatory movements into a pre-programmed melody (see Novembre et al., 2015). Each of the paired participants rotated the handle of an E-music box, with timing (but not tone sequencing) under the individual's control in the sense that performance tempo depended on the speed of revolution. This task allowed us to test individuals without musical training to avoid population-specific effects that could limit the generalizability of findings on self-other representational balance.

The relationship between co-performers' parts in terms of sound similarity—same pitches (unison) or different pitches (two octaves apart)—and movement similarity—same rotation frequency or different rotation frequencies—was varied in order to manipulate auditory streaming and visual movement congruence.

We assumed that having the same pitches across individuals leads to integration while different pitches encourage segregation due to auditory stream segregation. Importantly, when paired participants produce identical sounds, pitch information does not provide cues as to who is producing each sound, but timing does. Complete integration in this case could lead to confusion of agency, making it difficult to track which individual is producing which part (Jeannerod & Pacherie, 2004; Keller et al., 2016; Pacherie, 2012).

With regard to movement similarity, we expected that interpersonal coordination would be better when paired individuals used the same movement frequency than when they were required to move at different frequencies (in a 3:2 polyrhythm) in order to produce synchronous

sounds. Studies of intrapersonal (bimanual) polyrhythm production show that relatively complex ratios are unstable and (at tempos typical in music) tend to devolve to simpler ratios, ultimately 1:1 (Peper et al., 1995). Therefore, assuming the equivalence of intrapersonal and interpersonal coordination dynamics (Coey et al., 2012; Schmidt & Richardson, 2008), we expected that the condition requiring incongruent movement frequencies would be challenging to the extent that participants need to resist the tendency to produce congruent movements. Entrainment due to visual contact between paired participants was thus expected to encourage self-other integration even when the task goal of producing tones in synchrony proscribes it in the incongruent condition.

In line with the proposal that self-other representational balance characterized by simultaneous integration and segregation is beneficial, it was hypothesized that interpersonal coordination would be best with auditory segregation (different pitch streams) and visual integration (congruent motion). This question was addressed by analyzing asynchronies between tones produced by co-performers, as well as performance tempo, across conditions.

In order to have a means of checking whether participants paid attention to their partners' movements, we recorded data in a subset of trials while participants did the task without being able to see each other's actions. We expected the effect of movement similarity (mediated by visual observation of the partner), but not sound similarity, to be abolished or reduced in these trials.

2. Method

2.1. Participants

The 32 participants, forming 16 pairs, included 25 women and 7 men. Their ages ranged from 17 to 49 years with a mean age of 20.78 ($SD = 5.49$). All were undergraduate students of Western Sydney University, Australia, and participated as unpaid volunteers, completing the study for course credit. The participants were tested in pairs and knew each other before the experimental session. They were all self-declared right handers and had less than five years of musical training (apart from one participant with 7 years of training, who was inadvertently tested). The study was approved by the Human Research Ethics Committee at Western Sydney University (protocol number H10487), and all participants provided informed written consent.

2.2. Apparatus and stimuli

The experiment took place in a soundproof booth, where each pair of participants sat on either side of a table facing each other. The experiment was controlled using custom-built code written for Presentation

software (neurobehavioral systems, v16). Presentation was programmed to collect the inputs received from the two *E*-music boxes, and to present the mixed musical outputs to both participants through two headphone sets (Sennheiser HD 280 pro 64 Ω). Presentation also controlled the delivery of instructions to the participants via two monitors (BenQ HDMI, 16:9 Wide, Full HD 1080 P) (Fig. 1).

Two identical *E*-music boxes were placed on the table in front of the participants. The *E*-music box is an electromechanical musical device, which was developed to conduct empirical investigations exploring the human predisposition to make music and to interact with others through music, irrespective of previous musical training (Novembre et al., 2015). Its function is similar to a traditional barrel organ. The *E*-music box thus transforms rotary movements into a pre-programmed musical melody, whose tempo changes as a function of rotation velocity. The sequential order in which notes appear is fixed, meaning that participants cannot play wrong notes or notes in the wrong order. Only temporal aspects of performance—rhythm and tempo—are under participant control. The melody can be played with correct rhythm and stable tempo simply by rotating the handle with constant velocity. However, failure to rotate the handle evenly at the correct rate can lead to rhythmic distortions and tempo drift. A detailed description of software and hardware specifications of the *E*-music box is provided by Novembre et al. (2015).

The musical melody used in this experiment was an excerpt consisting of 23 tones from the song 'Somewhere over the rainbow' (see score in Fig. 1). All tones were played with piano as the musical instrument sound and had equal intensity (loudness). Each participant controlled one of two instrumental parts that could be relatively high or low in pitch (with the tone and register for the initial notes being C5 (higher-pitch part) and C3 (lower-pitch part) respectively). The song 'Somewhere over the rainbow' was chosen as musical material due to its popularity, which was assumed to facilitate the task. Familiarity was assessed using a 5-points Likert scale (1 = not familiar to 5 = very familiar) on which the melody received an average rating of 3.94 ($SD = 1.29$). Prior to the experiment, participants were asked if they knew the melody. Subjects who reported being unfamiliar with the melody ($n = 4$) were presented with the original recording of the song that was written for the movie 'The Wizard of Oz' in 1939 (Harburg & Arlen, 1939).

2.3. Design

A 2×2 repeated-measures design was employed, with the within-participant independent variables of *sound similarity* (same or different tone pitches across paired participants) and *movement similarity* (same or different rotation frequencies across participants) (Fig. 1). An extra qualitative control condition was included—separately from the main design—in which visual contact between participants was prevented (see Procedure section). The main dependent variables were indices of

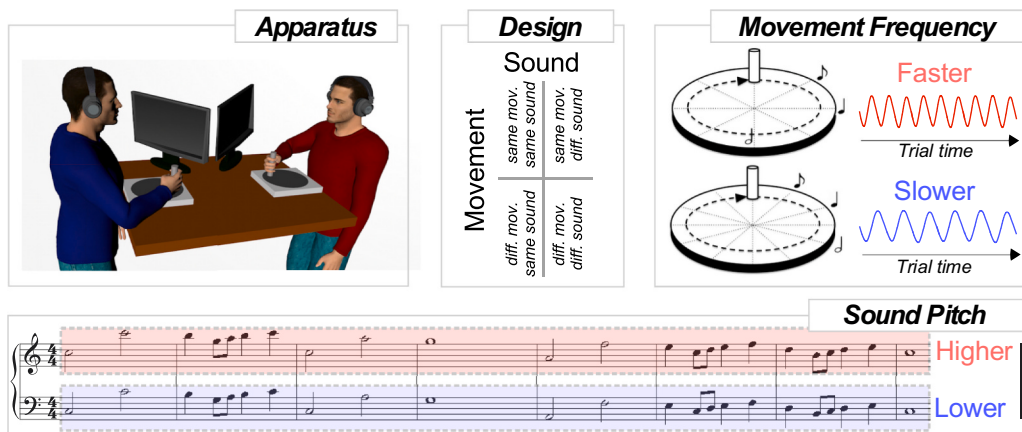


Fig. 1. Pairs of participants performed a musical melody together using *E*-music boxes in a 'dualing DJs' task. Across trials, each participant was assigned to either a relatively high or low musical pitch, leading to the two participants producing the same or a different musical output across conditions. The movement frequency necessary to perform the music at the same tempo across participants was also manipulated, leading to the two participants moving at the same or a different rotational pace across conditions.

how accurately participants synchronized the tones that they produced (interpersonal (a)synchrony), performance tempo, and tempo variability.

2.4. Procedure

An experimental trial involved a complete performance of the musical melody by the paired participants simultaneously producing their respective instrumental parts (see above). Each trial was structured as follows: First, a curved arrow presented on the monitor (3000 ms) instructed each participant whether to play the *E*-music box through clockwise or anticlockwise movement. Next, a metronome sounding four beats (750 ms inter-beat intervals) was presented while the monitor displayed an 'ear symbol' with an instruction stating 'listen'. This was intended to provide a temporal cue helping the participants to establish the correct tempo and to start in synchrony. When the metronome ceased, a 'go sign' was displayed to signal that the participants should start rotating their *E*-music box handles using the right hand and to synchronize their musical outputs as accurately as possible. Because the sequential order of tones was pre-programmed in the *E*-music boxes, participants only had to rotate their handle evenly at the prescribed tempo to perform the musical melody in synchrony. When the melody had been played entirely, an image was shown on the monitor to remind the participants to move the handle to the same start position (which corresponded to 0° or 12 o'clock on the *E*-music box). At that point, the next trial started.

Each pair completed 80 trials, which were grouped into four blocks in accordance with the experimental design: Two blocks entailed the performance of identical auditory outputs ('Same Sound' condition, both participants playing high pitch or low pitch tones) and two blocks entailed the performance of different auditory outputs ('Different Sound' condition, one participant playing high pitch tones and the other low pitch tones in one block, and the reverse arrangement in the other block). Each block was further split into four miniblocks: Two miniblocks required the participants to rotate the *E*-music boxes at the same frequency in order to produce synchronous music ('Same Movement Frequency' condition) and two miniblocks required the participants to turn the *E*-music boxes at different movement frequencies ('Different Movement Frequency' condition). Here, the relatively slower and faster movement frequencies resulted in the shortest musical note duration (an eighth note) sounding after a 30° or a 45° rotation, respectively (Fig. 1). The Same Movement Frequency condition thus entailed both participants moving at 30° rotations per eighth note (in one miniblock) or at 45° rotation per eighth note (in the other miniblock), whereas the Different Movement Frequency condition entailed one participant moving at 30° rotations per eighth note while the other moved at 45° rotation per eighth note (in one miniblock), or the reverse arrangement (in the other miniblock). Each miniblock comprised four trials, which included clockwise and anti-clockwise movement of one or both participants with equal probabilities (in order to vary motoric requirements of the task). The order of the four blocks and the four miniblocks within them was counterbalanced across participants, while the trials within a miniblock were randomized.

Within each block, we also recorded data from an extra qualitative control miniblock during which participants were not able to see each other (i.e. a screen was placed in between the two participants). These control miniblocks always took place before the other four miniblocks within a block. The movement frequency in these control miniblocks (one movement frequency condition per miniblock) was counterbalanced across the four blocks. Before each experimental block, participants performed two practice trials in order to familiarize themselves with the assigned pitches and the task. The experimenter remained in the booth for the entire experiment to monitor the participants' performance and ensure that they followed instructions.

2.5. Data analysis

Interpersonal coordination was examined by computing indices of asynchrony between paired participants, while performance tempo was also analyzed to aid interpretation of observed effects on coordination measures.

For interpersonal (a)synchrony, we calculated the absolute asynchronies between corresponding (i.e. complementary) tones performed by the two participants in each pair (i.e. 23 absolute asynchronies per trial). The mean of the absolute asynchronies (computed within each trial) was taken as an index of synchronization (in)accuracy (Keller et al., 2007; Novembre et al., 2012). Note that a small mean absolute asynchrony indicates accurate interpersonal coordination: the smaller this value, the more accurate is the ability of the paired participants to synchronize melodic tones with each other. To examine the time course of synchronization (in)accuracy as the melody progresses throughout individual trials, we extracted information about movement timing for consecutive eighth-note rotations of each participant's *E*-music box (45° or 30°, depending on Movement Frequency assignment) and then computed the absolute difference, giving a measure of instantaneous movement asynchrony (cf. Novembre et al., 2019). This was done for visualization purposes rather than formal analysis.

The mean absolute asynchrony results were investigated further by an additional analysis conducted on permuted data to test whether observed effects of *sound similarity* and *movement similarity* are attributable to interpersonal coupling. Pseudo absolute asynchronies were computed for every participant paired with all other participants apart from their actual partner. This was done at the trial level, that is, for condition-matched trials from participants who did not really interact one another. We assume that effects can be attributed to interpersonal coupling to the extent that they are significant for real pairs of participants but not for permuted pseudo pairs.

Next, we conducted analyses of mean tempo and tempo variability to test the degree to which participants performed the task at the instructed tempo. These analyses involved extracting information about movement timing for consecutive eighth-note rotations of each participant's *E*-music box (45° or 30°), and then first-order differencing successive time points to yield series of eighth-note rotation interval durations. Based on these time interval data, mean tempo for each trial was indexed by the average of all intervals, and tempo variability was indexed by the coefficient of variation (CV), computed by dividing the standard deviation of eighth note rotation intervals by the mean eighth-note rotation interval for each trial. Mean tempo data and tempo CV data for individual participants were averaged across pair members to yield a single mean tempo measure and a single tempo CV measure for each pair in each condition.

Finally, we examined differences in mean tempo (as computed above) between members of a pair. Differences in mean tempo can contribute to asynchronies independently of interpersonal coupling because two individuals playing a melody at different tempi will drift apart, and hence asynchronies will grow cumulatively throughout a trial, whether or not the individuals interact with one another. Note that these tempo analyses are based on eighth-note rotation interval durations, which should be the same (375 ms, given the target 750 ms inter-beat interval) irrespective of whether a participant is making 45° or 30° rotations per eighth note to produce tones at the given tempo (i.e., different distances need to be covered in the same amount of time). To test for differences in mean tempo between pair members, we computed the absolute value of the difference between participant 1 and participant 2's mean rotation interval durations for each pair in each experimental condition.

Preliminary analyses of variance (ANOVAs) demonstrated that there were no statistical differences between high pitched and low pitched sounds or faster and slower movements. Likewise, whether or not the participants were both moving clockwise or anti-clockwise did not interact with our experimental manipulations. This finding allowed us to

collapse the data into four conditions referred to as ‘Same Sound’, ‘Same Movement Frequency’, ‘Different Sound’ and ‘Different Movement Frequency’.

Before averaging single-trial values as a function of these conditions, we excluded trials with outlying values, i.e. those deviating by more than 2 SDs from the mean of each condition (4.38% of trials for absolute asynchronies, 2.50% of trials for mean tempo indices, 3.75% of trials for tempo variability indices, and 3.68% of trials for absolute value of the difference in the mean tempo). Next, the data were entered into repeated measures ANOVAs (one for each dependent variable) testing for effects of *sound similarity* and *movement similarity*. The additional factor *pair authenticity* was included in the analysis comparing mean absolute asynchrony data for real pairs of participants and permuted pseudo pairs. Violations of the sphericity assumption were corrected using the Greenhouse–Geisser correction.

For the plots of asynchrony time courses included for visualization purposes, we excluded trials where at least one data point was 3 SDs away from the mean (irrespective of condition). A different criterion from that used in the above formal analyses was employed because otherwise, with so many data points, many trials would be lost in these visualizations.

The absolute asynchronies data recorded in the absence of visual feedback in the control miniblocks (“Feedback Absent”; constituting only 20% of the trials) were analyzed separately following the same criteria as described above (“Feedback Present”). Data from these miniblocks were not compared directly against data from experimental blocks due to the smaller number of trials and the fact that control trials always constituted the first miniblock of a main experimental condition block (risking strong order effects). While the lack of a direct comparison might limit the strength of some potential conclusions, these do not bear on our main hypotheses.

For all analyses, partial eta squared (η_p^2) was calculated as a measure

of effect size.

3. Results

3.1. Interpersonal (a)synchrony

Fig. 2 shows interpersonal (a)synchrony (mean absolute asynchronies between tone onsets or rotation movements) computed when participants were able to see one another (Fig. 2A) and when they were not able to see each other (Fig. 2B). The bottom panel (Fig. 2C) shows time course plots of asynchrony throughout trials (with visual feedback). It can be seen in these plots that interpersonal coordination varied considerably across experimental conditions and participant pairs. Results for permuted asynchrony data computed for pseudo pairs are shown in the inset to Fig. 2A.

3.1.1. Performance with visual feedback

The ANOVA on mean absolute asynchronies in miniblocks with visual feedback revealed statistically significant main effects of *sound similarity*, $F(1,15) = 18.188$, $p < .001$, $\eta_p^2 = 0.548$, and *movement similarity*, $F(1,15) = 44.956$, $p < .001$, $\eta_p^2 = 0.750$. These results indicate that interpersonal coordination was more accurate when the paired participants produced sounds with different pitches than the same pitch, as well as when they performed with the same movement frequency than different movement frequencies. The interaction between *sound similarity* and *movement similarity* was also significant, $F(1,15) = 10.425$, $p = .006$, $\eta_p^2 = 0.410$. As can be seen in Fig. 2A, interpersonal coordination was most accurate when pairs of participants produced sounds of different pitch using the same movement frequency, and was least accurate (and in fact very poor) when they produced same-pitched sounds using different movement frequencies. It can also be seen that interpersonal coordination was impaired to a greater degree by different

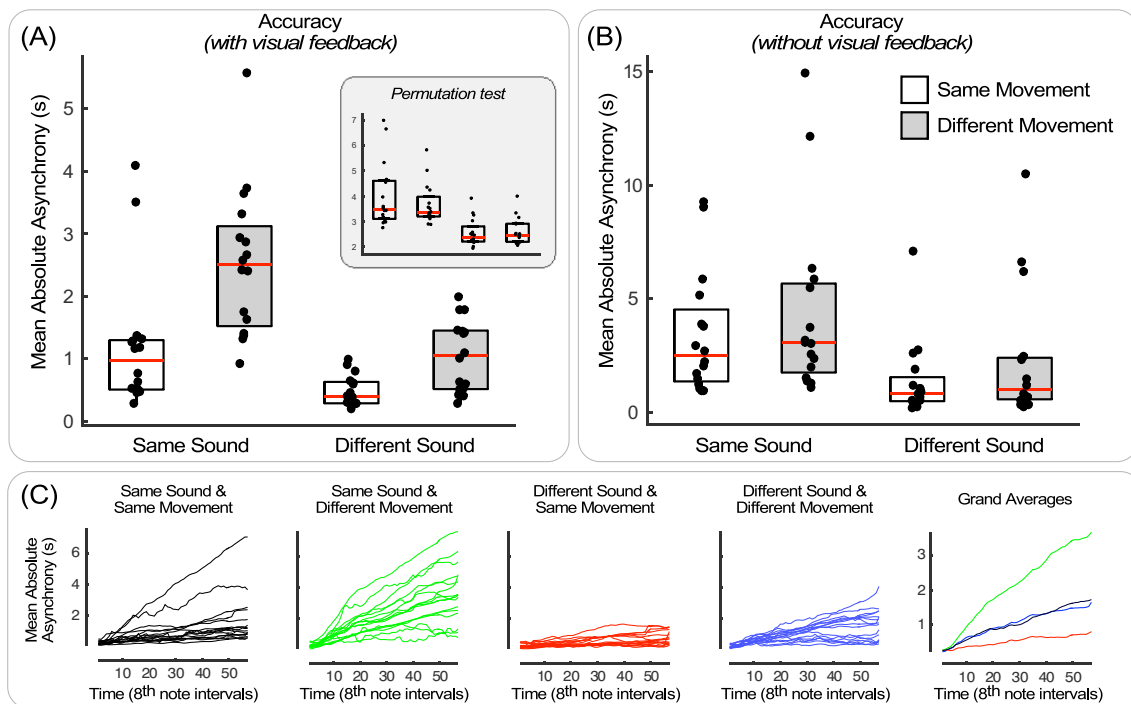


Fig. 2. Interpersonal (a)synchrony (mean absolute tone-onset or movement time difference) computed while paired participants produced either the same or different movements as well as the same or different sounds. (A) Participants were able to see each other's movements in all conditions (with visual feedback). The inset shows results for permuted tone-onset asynchrony data computed for pseudo pairs. (B) In a subset of control trials, participants were not able to see each other (without visual feedback). Black dots indicate within-pair averages. Red horizontal lines indicate medians across pairs. Confidence intervals are represented by the lower and upper edges of the boxes (25 to 75 centiles). (C) Time course of absolute asynchronies between participants' rotation movements across eighth-note intervals throughout a trial. The four left panels show data from individual pairs in each experimental condition, and the right panel shows data averaged across pairs. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

movements (relative to same movements) in the same sound condition than the different sound condition. To confirm this, we subtracted the mean absolute asynchronies (of each pair) associated with the Same Movement Frequency conditions from those associated with the Different Movement Frequency conditions, separately for the Same Sound and Different Sound conditions. The computed difference values, indexing the detrimental effect of different movement as a function of sound similarity, were compared one another using a non-parametric test (Wilcoxon Signed Rank Test). This test yielded a significant result, $Z = 2.896$, $p = .004$, indicating reliably larger disruption by a difference in movement frequencies when sounds were the same than when sounds were different.

To check that the observed effects of *sound similarity* and *movement similarity* are attributable to interpersonal interaction, we conducted an ANOVA that included mean absolute asynchrony data for real pairs of participants and permuted pseudo pairs. This ANOVA yielded statistically significant main effects of *sound similarity*, $F(1,15) = 19.575$, $p < .001$, $\eta^2_p = 0.566$, *movement similarity*, $F(1,15) = 42.314$, $p < .001$, $\eta^2_p = 0.738$, and *pair authenticity* (real pairs vs pseudo pairs), $F(1,15) = 53.501$, $p < .001$, $\eta^2_p = 0.781$. All interactions between these factors were likewise statistically significant: *sound similarity* $\times$ *movement similarity*, $F(1,15) = 9.632$, $p = .007$, $\eta^2_p = 0.391$; *sound similarity* $\times$ *pair authenticity*, $F(1,15) = 16.569$, $p = .001$, $\eta^2_p = 0.525$; *movement similarity* $\times$ *pair authenticity*, $F(1,15) = 47.331$, $p < .001$, $\eta^2_p = 0.759$; *sound similarity* $\times$ *movement similarity* $\times$ *pair authenticity*, $F(1,15) = 11.079$, $p = .005$, $\eta^2_p = 0.425$. The significant interactions involving *pair authenticity* indicate differential effects for real pairs and pseudo pairs that are presumably attributable to variations in the degree and success of interpersonal interaction in the case of real pairs.

To examine this further, an ANOVA was conducted on the permuted absolute asynchronies alone (i.e., pseudo pairs only). This ANOVA yielded a significant main effect of *sound similarity* ($F(1,15) = 32.074$, $p < .001$, $\eta^2_p = 0.681$), indicating that pairing the datasets of participants who did not perform the task together resulted in relatively larger pseudo asynchronies associated with the same sound condition. The main effect of *movement similarity* ($F(1,15) = 2.368$, $p = .145$, $\eta^2_p = 0.136$) and the interaction ($F(1,15) = 4.205$, $p = .058$, $\eta^2_p = 0.219$) were not statistically significant. These results suggest that the effect of *sound similarity* in the main analysis (of real pairs) may not be entirely attributable to differences in interpersonal coupling, while effect of *movement similarity* and the interaction between *sound similarity* and *movement similarity* are attributable to differences in interpersonal coupling.

The time course of synchronization accuracy throughout trials in the different conditions (Fig. 2C) is potentially informative about these differences in interpersonal coupling. Asynchrony time series have positive slopes in many cases (i.e., asynchronies increase over time), indicating that participants drifted apart, presumably due to weak or absent coupling. Such drift is clearest in the Same Sound/Different Movement condition (with large individual differences), intermediate in the Same Sound/Same Movement and Different Sound/Different Movement conditions, and least evident in the Different Sound/Same Movement condition. This suggests that participants struggled, and sometimes failed, to maintain interpersonal coupling in all but the latter condition.

We conducted two additional analyses to assess the impact of the observed drift on the main results. The first of these ‘drift-control’ analyses took the approach of excluding data from participants who showed the greatest propensity to drift, while the second analysis retained all participants but removed trials that displayed greatest drift. Neither analysis changed the overall pattern of results and conclusions (see Appendix A).

3.1.2. Performance without visual feedback

The analysis of variance on mean absolute asynchronies in control miniblocks without visual feedback (Fig. 2B) showed a significant main

effect of *sound similarity*, $F(1,15) = 14.612$, $p = .002$, $\eta^2_p = 0.493$, indicating that the interpersonal coordination was more accurate when the paired participants played different sounds than the same sound. However, we did not find a significant main effect of *movement similarity* ($F(1,15) = 0.966$, $p = .341$, $\eta^2_p = 0.061$) or an interaction between *sound similarity* and *movement similarity* ($F(1,15) = 0.001$, $p = .972$, $\eta^2_p < 0.001$). The qualitative correspondence between this result and the permutation analysis on data with visual feedback suggests that visual coupling may be a key determinant of interpersonal (a)synchrony on this task.

3.2. Performance tempo

Fig. 3 shows data for mean tempo (mean eight-note rotation interval duration, Fig. 3A) and tempo variability (CV of eight-note rotations, Fig. 3B) from miniblocks with visual feedback. It can be seen that the target tempo of 375 ms per eighth note was generally undershot (Fig. 3A), indicating that participants performed faster than instructed.

The ANOVA on mean tempo, averaged across players within each pair, yielded a significant main effect of *sound similarity*, $F(1,15) = 11.581$, $p = .004$, $\eta^2_p = 0.436$, indicating that the overall tempo was relatively slower when the two players produced the same pitched sounds (Fig. 3A). The main effect of *movement similarity* ($F(1,15) = 0.001$, $p = .974$, $\eta^2_p = 0.000$) and the interaction between *sound similarity* and *movement similarity* were not significant ($F(1,15) = 1.567$, $p = .230$, $\eta^2_p = 0.095$). Participants thus slowed down when the task was made difficult by identical pitch assignments, but not due to challenges associated with different movement frequencies.

The ANOVA on tempo CV data, averaged across players within each pair, yielded a significant main effect of *sound similarity*, $F(1,15) = 4.877$, $p = .043$, $\eta^2_p = 0.245$, and *movement similarity*, $F(1,15) = 16.960$, $p = .001$, $\eta^2_p = 0.531$, while the interaction was not significant ($F(1,15) = 0.059$, $p = .811$, $\eta^2_p = 0.004$) (Fig. 3B). This indicates that the tempo was more variable when the participants were producing same-pitched sounds and when they were performing different frequency movements. Challenging pitch and movement relations were thus both associated with unstable tempo.

Differences in mean tempo between participants within each pair were analyzed (in miniblocks with visual feedback), as these differences can contribute to asynchronies independently of interpersonal coupling. The ANOVA on the absolute difference in tempo yielded significant main effects of *sound similarity*, $F(1,15) = 14.660$, $p < .001$, $\eta^2_p = 0.699$, and *movement similarity*, $F(1,15) = 34.822$, $p = .002$, $\eta^2_p = 0.494$, as well as a significant interaction, $F(1,15) = 13.027$, $p = .003$, $\eta^2_p = 0.465$ (Fig. 3C). These results indicate that the tempi produced by the two players were generally more distinctive in the Same Sound condition than the Different Sound condition, and in the Different Movement Frequency condition than the Same Movement Frequency condition. The interaction further indicated that the effect of *movement similarity* was stronger when associated with the Same Sound condition than the Different Sound condition. This was confirmed by a non-parametric test (Wilcoxon Signed Rank Test) comparing the difference of the *movement similarity* conditions (same movement minus different movement) across *sound similarity* conditions ($Z = 2.947$, $p = .003$).

4. Discussion

The present study investigated simultaneous self-other integration and segregation during real-time interpersonal coordination in a musical joint action task. To this end, pairs of participants with little or no musical training performed a melody on E-music boxes in a duo setting while relations between their parts were varied in terms of auditory pitch and the frequency of rotational movements that produced tones. As hypothesized, it was found that interpersonal coordination was most accurate (i.e., asynchronies between tones were smallest) when sounds were different in pitch but movement frequency was the same across co-

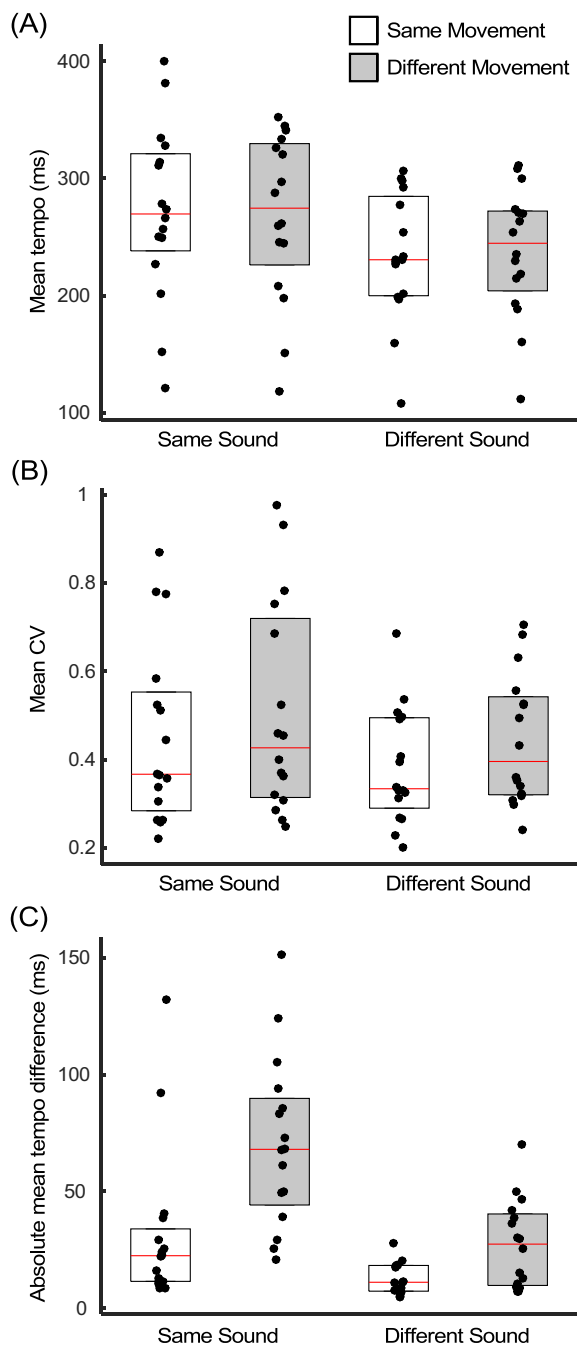


Fig. 3. Performance tempo and tempo variability. (A) Mean tempo indexed by mean eighth-note rotation interval duration averaged across paired participants; (B) tempo variability indexed by the CV of eighth-note rotation interval durations averaged across paired participants; (C) absolute value of the difference in mean tempo (eighth-note rotation interval) between paired participants. Black dots indicate within-pair averages (in A & B) or differences (in C). Red horizontal lines indicate medians across pairs. Confidence intervals are represented by the lower and upper edges of the boxes (25 and 75 centiles). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

performers. Coordination was poorest with same sounds and different movements across co-performers, who generally drifted apart, and intermediate in the same sound/same movement and the different sound/different movement conditions. These findings indicate that interpersonal coordination was facilitated under conditions that were conducive to simultaneous self-other integration and segregation across sensory

modalities, albeit in an asymmetrical manner where pitch relations favoured segregation in the auditory modality while movement congruence favoured integration in the visual modality.

4.1. Auditory segregation & visual integration rule

Our findings are consistent with the proposal that auditory segregation assists interpersonal coordination by helping paired participants to keep track of who is who during musical joint action (Keller et al., 2016). We assume that the condition in which co-performers played the melody together at different pitches (separated by two octaves) favoured self-other segregation to the extent that large pitch separation encourages auditory streaming (Bregman, 1990) where each individual's part is perceptually distinct. Nevertheless, the two streams are still highly correlated in terms of pitch transitions, and, therefore, in accordance with the principle of dimensional overlap (Kornblum et al., 1990), the two sequential structures can be readily mapped onto each other.

Self-other distinction in the auditory domain may be beneficial to a range of sensory-motor and cognitive processes supporting rhythmic interpersonal coordination during musical joint action. On this account, auditory segregation allows co-performers to divide attention between each other's parts as well as the overall group output, a process referred to as 'prioritized integrative attending' (Keller, 2001; Ragert et al., 2014). Furthermore, maintaining a perceptual distinction between parts enables a process of temporal anticipation by allowing the information about the timing of the other's performance—their successive inter-onset intervals—to be extracted in order to generate predictions about upcoming tone onsets (Pecenka & Keller, 2011). Finally, at a basic sensory-motor level, perceptually distinct auditory streams facilitate temporal adaptation processes that keep interpersonal asynchronies in check by implementing error correction (Repp & Keller, 2004). These error correction processes detect discrepancies in interpersonal timing and make compensatory adjustments to the phase or period of internal timekeepers that control action timing (Repp & Keller, 2008; Vorberg & Wing, 1996). To work effectively in reducing asynchronies, these mechanisms need access to information about whether an individual's action timing was early or late in order to program an appropriate corrective delay or advance in timing for the next action.

Playing the melody at the same pitch, by contrast, creates a situation that presumably encourages the perception of co-performers' parts as a single auditory stream, potentially making them difficult to distinguish due to acoustic masking, especially under conditions where interpersonal synchrony is the goal. This lack of distinction could lead to ambiguity of agency (Dumas et al., 2012; Jeannerod & Pacherie, 2004; Pacherie, 2012; van der Weiden et al., 2019; van der Wel, 2015; van der Wel et al., 2012), thereby interfering with prioritized integrative attending and temporal anticipation and adaptation due to uncertainty about which individual is producing which part.

The finding that participants adopted a slower performance tempo, which was also less stable, when they produced the same sounds than when producing different sounds suggests that difficulty in distinguishing each other's sounds might have induced a cautious approach (although performances were still overall faster than instructed). Whether this represents an attempt to trade speed for accuracy (see Plamondon & Alimi, 1997) is not known, and, if it was, the strategy was counterproductive, since movement timing variability generally increases with decreasing tempo (see Vorberg & Wing, 1996). It may therefore be the case that the effects of sound similarity were not exclusively attributable to differences in interpersonal coupling but also to a change in individual strategy. Consistent with this interpretation, the qualitative correspondence between results of the permutation (pseudo pair) analysis for blocks with visual feedback and results for blocks without visual feedback (i.e., a main effect of sound similarity in both analyses) speaks against an explanation based on the strength of interpersonal coupling. More generally, the observation that asynchronies increased steeply over the course of individual performances of

the melody—in other words, some participants drifted apart—in all conditions apart from the condition in which sounds were different and movement frequency was the same is suggestive of a breakdown in interpersonal coupling in the more challenging conditions. Participants were apparently ‘self’ focused in prioritizing the maintenance of a temporal distinction between self and other, even though interpersonal synchrony was sacrificed as a consequence (see Fairhurst et al., 2014).

As expected, the effects of movement congruence were complementary to those of sound similarity in terms of achieving self-other representational balance across modalities. Interpersonal coordination was facilitated by self-other integration in the visual modality when the same movement frequency was required to produce tones in synchrony relative to when the instructed movement frequency was different across paired participants. This finding is consistent with the assumption that observing another's actions can trigger a process of covert simulation due to links between the perceptual and motor processes in the brain (e.g., Jeannerod, 2001; Kilner et al., 2004; Schütz-Bosbach & Prinz, 2007). In the context of rhythmic interpersonal coordination, it has been claimed that such perception-action links can lead to the coupling of internal oscillations that control the timing of each individual's continuous movements (see Coey et al., 2012), with the strength of this interpersonal entrainment varying with similarity in movement frequency (e.g., Richardson et al., 2005). From this perspective, interpersonal coordination was better when participants were required to move in a 1:1 frequency ratio than when a 2:3 poly-rhythmic ratio was required because visuo-motor entrainment—hence self-other integration—was stronger and more stable for the simpler coupling ratio. Building on previous work showing that spatial incongruence is not necessarily disruptive (e.g., Clarke et al., 2019; Sacheli et al., 2018), our results demonstrate that temporal incongruence can be a potent source of interference in musical joint action. The finding that tempo was more variable when movement frequency differed across pair members supports this interference-based explanation.

The present effects of movement congruence are remarkable because the explicit goal of the task was to produce tones in synchrony, while movement coordination was incidental. It turns out that our assumption that auditory information would be paramount was apparently not entirely valid, perhaps because our participants were not trained musicians. In any case, our results speak to the automaticity and potency of the movement congruence effect, consistent with a large body of dynamical systems research on the spontaneous emergence of interpersonal synchrony via visuo-motor entrainment (see Coey et al., 2012; Schmidt & Richardson, 2008; Tognoli & Kelso, 2015). Indirect evidence for a visuo-motor (rather than purely motor) locus of our effect comes from the finding that, when participants could not see one another, only effects of pitch relations were reliable. In other words, it was the influence of visual information about a co-performer's incongruent movements on one's own movements—rather than merely the requirement to produce different motion trajectory profiles across co-performers—that interfered with interpersonal coordination. Although we did not directly compare performance with and without visual contact (since our research questions related to effects of movement congruency on visuo-motor coupling, not whether such coupling occurs), this could be done in future work to test whether different levels movement congruency facilitate versus interfere with performance relative to a neutral baseline.

4.2. Conclusions

Our findings provide evidence that simultaneous self-other integration and segregation plays a role in determining the quality of rhythmic interpersonal coordination during musical joint action. This process exhibits modality specificity wherein self-other representational balance facilitates coordination when information about self and other is readily segregated in the auditory domain and integrated in the visual domain. In the absence of conditions that are conducive to self-other integration

and segregation, interpersonal coupling can be lost, as indicated by the large asynchronies and differences in performance tempi between paired individuals that we observed in some conditions.

Specifically, co-performers drifted apart when extrinsic factors affecting interpersonal pitch relations and movement dynamics did not favor self-other integration and segregation across sensory modalities. The potency of these external constraints is highlighted by their ability to cause breakdowns in interpersonal synchrony, at least in individuals without musical training. Musicians might be able to overcome such challenges, and thereby avoid drifting apart, through generally enhanced timing skills, more finely tuned joint internal models that simulate upcoming actions of self and other, or greater flexibility in balancing reliance on external information versus internalized representations of the joint task (see MacRitchie et al., 2017; Novembre et al., 2016; Paas et al., 2021).

More broadly, the present results suggest that self-other representational balance is regulated via the interaction of action control processes related to discrete auditory information and continuous body movements. The involvement of these different classes of control process highlights the need for a theoretical framework that combines discrete information processing and continuous coordination dynamics (see Harry & Keller, 2019; Knoblich et al., 2011; Knoblich & Sebanz, 2008; MacRitchie et al., 2017; van der Steen & Keller, 2013; Vesper et al., 2010; Vesper & Richardson, 2014). Such a hybrid theoretical framework may ultimately be the key to understanding the effects of self-other integration and segregation in musical joint action and real-time interpersonal coordination more generally.

CRedit authorship contribution statement

Hannah Liebermann-Jordanidis: Conceptualization; Formal analysis; Investigation; Methodology; Writing - original draft. **Giacomo Novembre:** Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Software; Writing - review & editing. **Iring Koch:** Conceptualization; Methodology; Supervision; Writing - review & editing. **Peter Keller:** Conceptualization; Formal analysis; Funding acquisition; Methodology; Project administration; Supervision; Writing - review & editing.

Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent

Informed consent was obtained from all individual participants included in the study.

Declaration of competing interest

Hannah Liebermann-Jordanidis has no conflicts of interest to declare.

Giacomo Novembre has no conflicts of interest to declare.

Iring Koch has no conflicts of interest to declare.

Peter E. Keller has no conflicts of interest to declare.

Acknowledgements

This research was supported by a Future Fellowship grant from the Australian Research Council (FT140101162) awarded to P.E.K.

Appendix A

The main analysis of the effects of sound similarity and movement similarity on mean absolute asynchronies indicated that participants sometimes drifted apart. We therefore conducted two additional analyses to assess the impact of such drift on the main results. The first of these 'drift-control' analyses took the approach of excluding data from participants who showed the greatest propensity to drift, while the second analysis retained all participants but removed trials that displayed greatest drift.

For the first drift-control analysis, it was noted that the individual differences evident in the time course plot for the Same Sound/Same Movement condition highlight two pairs who were particularly susceptible to drift in this condition. An ANOVA on mean absolute asynchronies excluding these two outliers returned significant main effects of *sound similarity*, $F(1,13) = 34.769$, $p < .001$, $\eta^2_p = 0.728$, and *movement similarity*, $F(1,13) = 44.068$, $p < .001$, $\eta^2_p = 0.772$, and a significant interaction between *sound similarity* and *movement similarity*, $F(1,13) = 9.303$, $p = .009$, $\eta^2_p = 0.417$. Furthermore, an ANOVA on the reduced dataset with the additional factor *pair authenticity* yielded main effects and interactions that, as in the corresponding full analysis above, were all statistically significant ($ps < 0.01$). These results suggest that the two pairs of participants who exhibited large degrees of drift in the Same Sound/Same Movement condition did not drive the results observed in the analyses of the full sample.

In the second drift-control analysis, we removed trials where the slope of asynchronies over time was more than 2 SDs above or below the mean slope of asynchrony series, calculated separately for each participant pair, pooled across experimental conditions. This resulted in a further 6.10% percent of trials being excluded from analysis (on top of the 4.38% of trials already removed due to outlying asynchronies). As in the main analysis, the ANOVA on these data yielded significant main effects of *sound similarity*, $F(1,15) = 17.008$, $p = .001$, $\eta^2_p = 0.531$, and *movement similarity*, $F(1,15) = 40.289$, $p < .001$, $\eta^2_p = 0.729$, and a significant interaction between *sound similarity* and *movement similarity*, $F(1,15) = 7.748$, $p = .014$, $\eta^2_p = 0.341$.

In sum, the two drift-control analyses reported here support the conclusions drawn from the analyses of mean absolute asynchrony reported in the main article.

References

- Billig, A. J., & Carlyon, R. P. (2016). Automaticity and primacy of auditory streaming: Concurrent subjective and objective measures. *Journal of Experimental Psychology: Human Perception and Performance*, 42(3), 339–353. <https://doi.org/10.1037/xhp0000146>
- Bishop, L., Cancino-Chacón, C., & Goebel, W. (2019). Moving to communicate, moving to interact: Patterns of body motion in musical duo performance. *Music Perception*, 37(1), 1–25. <https://doi.org/10.1525/mp.2019.37.1.1>
- Bishop, L., & Goebel, W. (2015). When they listen and when they watch: Pianists' use of nonverbal audio and visual cues during duet performance. *Musicae Scientiae*, 19(1), 84–110. <https://doi.org/10.1177/1029864915570355>
- Blakemore, S. J., & Frith, C. (2005). The role of motor contagion in the prediction of action. *Neuropsychologia*, 43, 260–267. <https://doi.org/10.1016/j.neuropsychologia.2004.11.012>
- Brass, M., & Heyes, C. (2005). Imitation: Is cognitive neuroscience solving the correspondence problem? *Trends in Cognitive Sciences*, 9, 489–495. <https://doi.org/10.1016/j.tics.2005.08.007>
- Bregman, A. S. (1990). *Auditory scene analysis*. Cambridge, MA: MIT Press.
- Chang, A., Kragness, H. E., Livingstone, S. R., Bosnyak, D. J., & Trainor, L. J. (2019). Body sway reflects joint emotional expression in music ensemble performance. *Scientific Reports*, 9(1), Article 205. <https://doi.org/10.1038/s41598-018-36358-4>
- Clarke, S., McEllin, L., Francová, A., Székely, M., Butterfill, S. A., & Michael, J. (2019). Joint action goals reduce visuomotor interference effects from a partner's incongruent actions. *Scientific Reports*, 9(1), Article 15414. <https://doi.org/10.1038/s41598-019-52124-6>
- Coey, C. A., Varlet, M., & Richardson, M. J. (2012). Coordination dynamics in a socially situated nervous system. *Frontiers in Human Neuroscience*, 6, 164. <https://doi.org/10.3389/fnhum.2012.00164>
- Cracco, E., Bardi, L., Desmet, C., Genschow, O., Rigoni, D., De Coster, L., Radkova, I., Deschrijver, E., & Brass, M. (2018). Automatic imitation: A meta-analysis. *Psychological Bulletin*, 144(5), 453–500. <https://doi.org/10.1037/bul0000143>
- D'Ausilio, A., Novembre, G., Fadiga, L., & Keller, P. E. (2015). What can music tell us about social interaction? *Trends in Cognitive Sciences*, 19(3), 111–114. <https://doi.org/10.1016/j.tics.2015.01.005>
- Davidson, J., & Broughton, M. C. (2016). Bodily mediated coordination, collaboration, and communication in music performance. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (pp. 573–595). Oxford, U.K.: Oxford University Press.
- De Jaegher, H., & Di Paolo, E. (2007). Participatory sense-making: An enactive approach to social cognition. *Phenomenology and the Cognitive Sciences*, 6, 485–507. <https://doi.org/10.1007/s11097-007-9076-9>
- Dumas, G., Martinerie, J., Soussignan, R., & Nadel, J. (2012). Does the brain know who is at the origin of what in an imitative interaction? *Frontiers in Human Neuroscience*, 6(128). <https://doi.org/10.3389/fnhum.2012.00128>
- Fairhurst, M. T., Janata, P., & Keller, P. E. (2014). Leading the follower: An fMRI investigation of dynamic cooperativity and leader-follower strategies in synchronization with an adaptive virtual partner. *Neuroimage*, 84, 688–697. <https://doi.org/10.1016/j.neuroimage.2013.09.027>
- Glowinski, D., Mancini, M., Cowie, R., Camurri, A., Chiorri, C., & Doherty, C. (2013). The movements made by performers in a skilled quartet: A distinctive pattern, and the function that it serves. *Frontiers in Psychology*, 4, 841. <https://doi.org/10.3389/fpsyg.2013.00841>
- Harburg, E. Y., & Arlen, H. (1939). *Somewhere over the rainbow. From the film "The Wizard of Oz"*. Leo Feist, Inc.
- Harry, B., & Keller, P. E. (2019). Tutorial and simulations with ADAM: An adaptation and anticipation model of sensorimotor synchronization. *Biological Cybernetics*, 113(4), 397–421. <https://doi.org/10.1007/s00422-019-00798-6>
- Heggl, O. A., Cabral, J., Konvalinka, I., Vuust, P., & Ringelbach, M. L. (2019). A Kuramoto model of self-other integration across interpersonal synchronization strategies. *PLoS Computational Biology*, 15(10), Article e1007422. <https://doi.org/10.1371/journal.pcbi.1007422>
- Hilt, P. M., Badino, L., D'Ausilio, A., Volpe, G., Tokay, S., Fadiga, L., & Camurri, A. (2019). Multi-layer adaptation of group coordination in musical ensembles. *Scientific Reports*, 9(1), Article 5854. <https://doi.org/10.1038/s41598-019-42395-4>
- Jeannerod, M. (2001). Neural simulation of action: A unifying mechanism for motor cognition. *Neuroimage*, 14, S103–S109. <https://doi.org/10.1006/nimg.2001.0832>
- Jeannerod, M., & Pacherie, E. (2004). Agency, simulation and self-identification. *Mind & Language*, 19, 113–146. <https://doi.org/10.1111/j.1468-0017.2004.00251.x>
- Kahl, S., & Kopp, S. (2018). A predictive processing model of perception and action for self-other distinction. *Frontiers in Psychology*, 9, 2421. <https://doi.org/10.3389/fpsyg.2018.02421>
- Kawase, S. (2014). Gazing behavior and coordination during piano duo performance. *Attention, Perception & Psychophysics*, 76, 527–540. <https://doi.org/10.3758/s13414-013-0568-0>
- Keller, P. E. (2001). Attentional resource allocation in musical ensemble performance. *Psychology of Music*, 29, 20–38. <https://doi.org/10.1177/0305735601291003>
- 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. (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, UK: Oxford University Press.
- 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. <https://doi.org/10.1016/j.concog.2005.12.004>
- Keller, P. E., & Koch, I. (2006). The planning and execution of short auditory sequences. *Psychonomic Bulletin & Review*, 13, 711–716. <https://doi.org/10.3758/BF03193985>
- Keller, P. E., & Koch, I. (2008). Action planning in sequential skills: Relations to music performance. *Quarterly Journal of Experimental Psychology*, 61, 275–291. <https://doi.org/10.1080/17470210601160864>
- Keller, P. E., Novembre, G., & Hove, M. J. (2014). Rhythm in joint action: Psychological and neurophysiological mechanisms for real-time interpersonal coordination. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 369(1658), Article 20130394. <https://doi.org/10.1098/rstb.2013.0394>
- Keller, P. E., Novembre, G., & Loehr, J. (2016). Musical ensemble performance: Representing self, other and joint action outcomes. In S. S. Obhi, & E. S. Cross (Eds.), *Shared representations: Sensorimotor foundations of social life* (pp. 280–310). Cambridge: Cambridge University Press.
- Keller, P. E., & Repp, B. H. (2008). Multilevel coordination stability: Integrated goal representations in simultaneous intra-personal and inter-agent coordination. *Acta Psychologica*, 128, 378–386. <https://doi.org/10.1016/j.actpsy.2008.03.012>
- Kilner, J. M., Paulignan, Y., & Blakemore, S. J. (2003). An interference effect of observed biological movement on action. *Current Biology*, 13(6), 522–525. [https://doi.org/10.1016/S0960-9822\(03\)00165-9](https://doi.org/10.1016/S0960-9822(03)00165-9)
- Kilner, J. M., Vargas, C., Duval, S., Blakemore, S. J., & Sirigu, A. (2004). Motor activation prior to observation of a predicted movement. *Nature Neuroscience*, 7(12), 1299–1301. <https://doi.org/10.1038/nn1355>
- Knoblich, G., Butterfill, S., & Sebanz, N. (2011). Psychological research on joint action: Theory and data. *Psychology of Learning and Motivation*, 54, 59–101. <https://doi.org/10.1016/B978-0-12-385527-5.00003-6>
- Knoblich, G., & Sebanz, N. (2008). Evolving intentions for social interaction: From entrainment to joint action. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 363, 2021–2031. <https://doi.org/10.1098/rstb.2008.0006>

- Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: Cognitive basis for stimulus-response compatibility—A model and taxonomy. *Psychological Review*, 97, 253–270. <https://doi.org/10.1037/0033-295x.97.2.253>
- Loehr, J. D., Kourtis, D., Vesper, C., Sebanz, N., & Knoblich, G. (2013). Monitoring individual and joint action outcomes in duet music performance. *Journal of Cognitive Neuroscience*, 25, 1049–1061. https://doi.org/10.1162/jocn_a.00388
- MacRitchie, J., Herff, S. A., Procopio, A., & Keller, P. E. (2018). Negotiating between individual and joint goals in ensemble musical performance. *Quarterly Journal of Experimental Psychology*, 71(7), 1535–1551. <https://doi.org/10.1080/17470218.2017.1339098>
- MacRitchie, J., Varlet, M., & Keller, P. E. (2017). Embodied expression through entrainment and co-representation in musical ensemble performance. In M. Lesaffre, P. J. Maes, & M. Leman (Eds.), *The Routledge companion to embodied music interaction* (pp. 150–159). New York: Routledge.
- Müller, D., Widmann, A., & Schröger, E. (2005). Auditory streaming affects the processing of successive deviant and standard sounds. *Psychophysiology*, 42, 668–676. <https://doi.org/10.1111/j.1469-8986.2005.00355.x>
- Novembre, G., & Keller, P. E. (2014). A conceptual review on action-perception coupling in the musicians' brain: What is it good for? *Frontiers in Human Neuroscience*, 8, 603. <https://doi.org/10.3389/fnhum.2014.00603>
- Novembre, G., Mitsopoulos, Z., & Keller, P. E. (2019). Empathic perspective taking promotes interpersonal coordination through music. *Scientific Reports*, 9(1), Article 12255. <https://doi.org/10.1038/s41598-019-48556-9>
- Novembre, G., Sammler, D., & Keller, P. E. (2016). Neural alpha oscillations index the balance between self-other integration and segregation in real-time joint action. *Neuropsychologia*, 89, 414–425. <https://doi.org/10.1016/j.neuropsychologia.2016.07.027>
- Novembre, G., Ticini, L. F., Schütz-Bosbach, S., & Keller, P. E. (2012). Distinguishing self and other in joint action. Evidence from a musical paradigm. *Cerebral Cortex*, 22, 2894–2903. <https://doi.org/10.1093/cercor/bhr364>
- Novembre, G., Varlet, M., Muawiyath, S., Stevens, C. J., & Keller, P. E. (2015). The E-music box: An empirical method for exploring the universal capacity for musical production and for social interaction through music. *Royal Society Open Science*, 2 (11), Article 150286. <https://doi.org/10.1098/rsos.150286>
- Paas, A., Novembre, G., Lappe, C., & Keller, P. E. (2021). Not all errors are alike: Modulation of error-related neural responses in musical joint action. *Social Cognitive and Affective Neuroscience*, 16, 512–524. <https://doi.org/10.1093/scan/nsab019>
- Pacherie, E. (2012). The phenomenology of joint action: Self-agency vs. joint-agency. In A. Seemann (Ed.), *Vol. 93. Joint attention: New developments* (pp. 343–389). Cambridge, MA: MIT Press.
- Pecenka, N., & Keller, P. E. (2011). The role of temporal prediction abilities in interpersonal sensorimotor synchronization. *Experimental Brain Research*, 211, 505–515. <https://doi.org/10.1007/s00221-011-2616-0>
- Peper, C. E., Beek, P. J., & van Wieringen, P. C. W. (1995). Multifrequency coordination in bimanual tapping: Asymmetrical coupling and signs of supercriticality. *Journal of Experimental Psychology: Human Perception and Performance*, 21(5), 1117–1138. <https://doi.org/10.1037/0096-1523.21.5.1117>
- Plamondon, R., & Alimi, A. M. (1997). Speed/accuracy trade-offs in target-directed movements. *Behavioral and Brain Sciences*, 20(2), 279–303. <https://doi.org/10.1017/s0140525x97001441> (discussion 303–249).
- Prinz, W. (1997). Perception and action planning. *European Journal of Cognitive Psychology*, 9(2), 129–154. <https://doi.org/10.1080/713752551>
- Ragert, M., Fairhurst, M. T., & Keller, P. E. (2014). Segregation and integration of auditory streams when listening to multi-part music. *PLoS One*, 9, Article e84085. <https://doi.org/10.1371/journal.pone.0084085>
- Repp, B. H., & Keller, P. E. (2004). Adaptation to tempo changes in sensorimotor synchronization: Effects of intention, attention, and awareness. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, 57, 499–521. <https://doi.org/10.1080/02724980343000369>
- Repp, B. H., & Keller, P. E. (2008). Sensorimotor synchronization with adaptively timed sequences. *Human Movement Science*, 27, 423–456. <https://doi.org/10.1016/j.humov.2008.02.016>
- Richardson, M. J., Marsh, K. L., Isenhowe, R. W., Goodman, J. R., & Schmidt, R. C. (2007). Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *Human Movement Science*, 26, 867–891. <https://doi.org/10.1016/j.humov.2007.07.002>
- Richardson, M. J., Marsh, K. L., & Schmidt, R. C. (2005). Effects of visual and verbal interaction on unintentional interpersonal coordination. *Journal of Experimental Psychology. Human Perception and Performance*, 31, 62–79. <https://doi.org/10.1037/0096-1523.31.1.62>
- Sacheli, L. M., Arcangeli, E., & Paulesu, E. (2018). Evidence for a dyadic motor plan in joint action. *Scientific Reports*, 8(1), Article 5027. <https://doi.org/10.1038/s41598-018-23275-9>
- Schmidt, R. C., Bienvenu, M., Fitzpatrick, P. A., & Amazeen, P. G. (1998). A comparison of intra- and interpersonal interlimb coordination: Coordination breakdowns and coupling strength. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 884–900. <https://doi.org/10.1037/0096-1523.24.3.884>
- Schmidt, R. C., Carello, C., & Turvey, M. T. (1990). Phase transitions and critical fluctuations in the visual coordination of rhythmic movements between people. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 227–247. <https://doi.org/10.1037/0096-1523.16.2.227>
- Schmidt, R. C., & Richardson, M. J. (2008). Dynamics of interpersonal coordination. In A. Fuchs, & V. K. Jirsa (Eds.), *Coordination: Neural, behavioral and social dynamics* (pp. 281–308). Berlin Heidelberg: Springer.
- Schubotz, R. I. (2007). Prediction of external events with our motor system: Towards a new framework. *Trends in Cognitive Sciences*, 11, 211–218. <https://doi.org/10.1016/j.tics.2007.02.006>
- Schütz-Bosbach, S., Mancini, B., Aglioti, S. M., & Haggard, P. (2006). Self and other in the human motor system. *Current Biology*, 16, 1830–1834. <https://doi.org/10.1016/j.cub.2006.07.048>
- Schütz-Bosbach, S., & Prinz, W. (2007). Perceptual resonance: Action-induced modulation of perception. *Trends in Cognitive Sciences*, 11(8), 349–355. <https://doi.org/10.1016/j.tics.2007.06.005>
- Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, 10, 70–76. <https://doi.org/10.1016/j.tics.2005.12.009>
- Sebanz, N., & Knoblich, G. (2009). Prediction in joint action: What, when, and where. *Topics in Cognitive Science*, 1, 353–367. <https://doi.org/10.1111/j.1756-8765.2009.01024.x>
- van der Steen, M. C., & Keller, P. E. (2013). The ADaptation and Anticipation Model (ADAM) of sensorimotor synchronization. *Frontiers in Human Neuroscience*, 7, 253. <https://doi.org/10.3389/fnhum.2013.00253>
- Su, Y.-H., & Keller, P. E. (2020). Your move or mine? Music training and kinematic compatibility modulate synchronization with self- versus other-generated dance movement. *Psychological Research*, 84, 62–80. <https://doi.org/10.1007/s00426-018-0987-6>
- Tognoli, E., & Kelso, J. A. S. (2015). The coordination dynamics of social neuromarkers. *Frontiers in Human Neuroscience*, 9, 563. <https://doi.org/10.3389/fnhum.2015.00563>
- Vesper, C., Butterfill, S., Knoblich, G., & Sebanz, N. (2010). A minimal architecture for joint action. *Neural Networks*, 23, 998–1003. <https://doi.org/10.1016/j.neunet.2010.06.002>
- Vesper, C., & Richardson, M. J. (2014). Strategic communication and behavioral coupling in asymmetric joint action. *Experimental Brain Research*, 232(9), 2945–2956. <https://doi.org/10.1007/s00221-014-3982-1>
- Vorberg, D., & Wing, A. M. (1996). Modeling variability and dependence in timing. In S. W. Keele (Ed.), *Vol. 2. Handbook of perception and action* (pp. 181–262). London: Academic Press.
- van der Weiden, A., Liepelt, R., & van Haren, N. E. M. (2019). A matter of you versus me? Experiences of control in a joint go/no-go task. *Psychological Research*, 83(5), 842–851. <https://doi.org/10.1007/s00426-017-0903-5>
- van der Wel, R. P. R. D. (2015). Me and we: Metacognition and performance evaluation of joint actions. *Cognition*, 140, 49–59. <https://doi.org/10.1016/j.cognition.2015.03.011>
- van der Wel, R. P. R. D., Sebanz, N., & Knoblich, G. (2012). The sense of agency during skill learning in individuals and dyads. *Consciousness and Cognition*, 21(3), 1267–1279. <https://doi.org/10.1016/j.concog.2012.04.001>
- Wilson, M., & Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological Bulletin*, 131, 460–473. <https://doi.org/10.1037/0033-2909.131.3.460>
- 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. <https://doi.org/10.1007/s00426-010-0280-9>