Timing and correction of stepping movements with a virtual reality avatar

Omar Khan[ID][1⊙], Imran Ahmed[2‡], Joshua Cottingham[3‡], Musa Rahhal[4‡], Theodoros N. Arvanitis[1⊙], Mark T. Elliott[ID][1⊙]*

1 Warwick Manufacturing Group, Institute of Digital Healthcare, University of Warwick, Coventry, United Kingdom, 2 Warwick Medical School, University of Warwick, Coventry, United Kingdom, 3 Department of Computer Science, University of Warwick, Coventry, United Kingdom, 4 School of Engineering, University of Warwick, Coventry, United Kingdom

⊙ These authors contributed equally to this work.
‡ These authors also contributed equally to this work.
* m.t.elliott@warwick.ac.uk

Abstract

Research into the ability to coordinate one's movements with external cues has focussed on the use of simple rhythmic, auditory and visual stimuli, or interpersonal coordination with another person. Coordinating movements with a virtual avatar has not been explored, in the context of responses to temporal cues. To determine whether cueing of movements using a virtual avatar is effective, people's ability to accurately coordinate with the stimuli needs to be investigated. Here we focus on temporal cues, as we know from timing studies that visual cues can be difficult to follow in the timing context. Real stepping movements were mapped onto an avatar using motion capture data. Healthy participants were then motion captured whilst stepping in time with the avatar's movements, as viewed through a virtual reality headset. The timing of one of the avatar step cycles was accelerated or decelerated by 15% to create a temporal perturbation, for which participants would need to correct to, in order to remain in time. Step onset times of participants relative to the corresponding step-onsets of the avatar were used to measure the timing errors (asynchronies) between them. Participants completed either a visual-only condition, or auditory-visual with footstep sounds included, at two stepping tempo conditions (Fast: 400ms interval, Slow: 800ms interval). Participants' asynchronies exhibited slow drift in the Visual-Only condition, but became stable in the Auditory-Visual condition. Moreover, we observed a clear corrective response to the phase perturbation in both the fast and slow tempo auditory-visual conditions. We conclude that an avatar's movements can be used to influence a person's own motion, but should include relevant auditory cues congruent with the movement to ensure a suitable level of entrainment is achieved. This approach has applications in physiotherapy, where virtual avatars present an opportunity to provide the guidance to assist patients in adhering to prescribed exercises.

Introduction

Mirroring of movements has been described as an important aspect of social interactions[1], as well as providing a top-down process that allows humans to anticipate actions and their goals when interacting with others[2]. This process involves the mapping of the visual information about an agent's movement on to one's own interpreted motor representation[3]. The opportunity to exploit action mirroring for training the timing and coordination of movements is clear and has been investigated in the context of observing familiar and unfamiliar dance moves[4] and, in early experimental work, as visual cues for lower limb function exercise training[5]. However, existing timing research in the visual domain has currently focussed on human-human interaction or very simple artificial cues, such as flashing lights or moving dots; coordinating movements with a virtual avatar has not been explored in the context of responses to temporal cues.

In this study, we investigate how accurately participants are able to coordinate the timing of their stepping movements with that of an avatar in a virtual reality environment, through the use of a sensorimotor synchronisation paradigm. Simple auditory and visual cues have been used to investigate the timing and synchronisation of upper and lower limb movements to rhythmic stimuli [6–9]. However, the influence exerted by auditory cues versus visual cues is disputed. In some contexts, (e.g., side by side walking[10]) the visual component was found to have less of an influence on step synchronisation than the auditory component. In contrast, the use of visual stepping stones on a treadmill revealed that corrections to visual cues were faster than auditory cues, thus indicating gait may be more strongly influenced by visual cues in certain contexts[6]. The visual stimuli used in the majority of the published literature to date has been limited to simple representations, such as flashing lights or bouncing spheres (e.g., [8,11]). The abstract nature of the cues used may limit the scope of their effect, highlighted by the finding that maintaining a defined distance to an avatar is more accurate compared to the same task using simple spheres[12]. Recent research has highlighted the importance of a spatial element to facilitate accurate temporal coordination of movements with the cue, relative to a static visual cue [8,13,14]. Moreover, it has been shown that point-light figures are effective in inducing beat entrainment over auditory beats[15]. Hence, the presence of spatio-temporal information, in humanoid form, could invoke the mirror system [1], through the representation of familiar movements. Here, we hypothesise that a virtual reality avatar could provide effective cues for lower limb timing coordination, in the form of stepping actions.

We used a fully immersive environment created using a VR headset, such that participants could view the avatar in three-dimensions, including the ability to move around relative to the avatar's position. Based on evidence that multisensory cues can further improve the accuracy of sensorimotor synchronisation[7,16], the avatar was presented in either a visual-only format (no contextual audio cues relating to foot strikes) or visually with contextual auditory footsteps. Earlier experiments have provided clear evidence of the correction process in both finger tapping and stepping [14,17], and have demonstrated evidence that multisensory integration takes place with visual cues[7,16]. Hence, our focus in this study was the response to humanoid visual stimulus, in particular the effects this would have on synchronisation and corrections in response to temporal perturbations.

To test whether participants were truly synchronising with the avatar (rather than just generating 'similar' responses), we used a phase perturbation paradigm, by lengthening or shortening one of the step cycles during a trial by an amount that was unnoticeable by participants. This has previously been used to quantify the correction of movements in order to remain in synchronisation with the cue with similar correction responses observed for both finger tapping[18–20] and stepping[7].

Overall, we hypothesised that, due to the continuous spatial information portrayed by the avatar: 1) Participants would be able to match the tempo of their steps with the Avatar's, at both fast and slow stepping tempos; 2) Participants would be able to synchronise their steps with the Avatar's and subsequently correct their steps following a perturbation, but would exhibit greater accuracy (lower asynchrony variance and faster correction), when the avatar's movement was augmented with auditory footsteps.

Results

Participants wore a VR headset, in which they saw an avatar standing approximately 3m in front of them. This distance was chosen as it was determined to be a suitable compromise between the avatar being close enough to allow a clear view of avatar movements, but far enough away so that the whole avatar could be kept in view. The large field of view for the Head Mounted Display meant that the whole avatar, including the feet, would be visible to the participant at this distance.

The avatar was facing in the same direction as the participant, such that the participant was viewing the back of the avatar. Participants were fitted with a full body marker set (Plug-in Gait, Vicon[21]), such that their own movements could be accurately tracked. They were instructed to step in time with the avatar as accurately as they could. Each trial included a perturbation to a single step cycle of the avatar, equivalent to 15% of the inter-step interval. Participants completed either Experiment 1 (Visual-only) or Experiment 2 (Audio-visual), which involved 4 blocks of 5 trials (20 trials in total). For each between-subjects condition, there were within-subjects conditions of Tempo: Slow (800ms inter-step interval), Fast (400ms inter-step interval) and Perturbation Direction: Positive (longer step cycle interval), Negative (shorter step cycle interval).

Matching of inter-step interval

To evaluate whether participants were able to match their stepping tempo to the avatar, the mean inter-step interval, ISI, for each participant per trial type was calculated. A mixed measures ANOVA was used to compare mean and standard deviation of participant ISIs across conditions. In addition, to calculate this value, step onsets were calculated as the point the heel marker exceeded a defined vertical height, see Methods; Fig 1a. The intervals were calculated as the time between the current and preceding step onsets. Intervals were then averaged between the third step interval and the interval preceding the perturbed step.

A mixed measures ANOVA was used to compare mean and standard deviation of participant ISIs across conditions. In addition, each participant's mean ISI for each condition was compared to the avatar's corresponding mean ISI using a paired t-test (with adjusted p-value for multiple comparisons, using a Bonferroni correction). The mean Avatar ISI is included in the analysis because this varied slightly across participants and trials, largely due to slight variations in frame rate as well as the Avatar motion being based on real movement data from a pre-recorded step cycle. A real recording was used to ensure the movement profile was natural and could be seamlessly looped. Mapping a whole stepping recording, with multiple step cycles, would have resulted in complexities in measuring timing performance so one cycle was used to provide some control over the cue (see Methods).

On the whole, participants were clearly able to differentiate between the two avatar ISI conditions (400ms and 800ms) and adjust their own step timing accordingly ($F_{1,18}=719.169$, $p<.001$; Fig 2). This indicates that the avatar was successfully able to influence participants to step at a pace other than a self-selected ISI. An interaction between the tempo and modality of the cues ($F_{1,18}=20.805$, $p<.001$) highlighted differences between the Auditory-Visual and
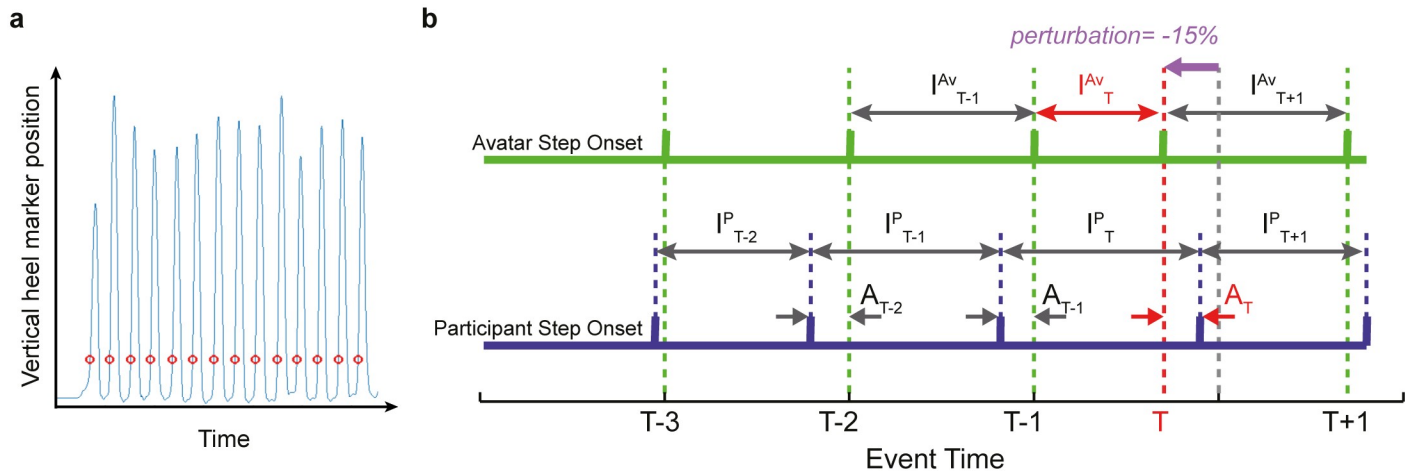
Fig 1. Definition of step onsets and resulting temporal performance measures. a). Schematic representation of step onsets (red circles) extracted from the vertical heel marker trajectory (blue line) during the stepping task. Onsets are defined by the time point when the vertical axis coordinate of the captured heel marker rose above a predefined threshold height above the floor surface. Onsets were extracted from the avatar and participants in the same way. **b).** Example schematic showing temporal performance measures based on the step onset times of the avatar and participant. Asynchrony (A) is measured as the time difference between the participant's and corresponding avatar step onsets. Inter-step intervals for the avatar (I^{Av}) and participant (I^P) are calculated as the time between a participant's/avatar's own step onsets. Step T, refers to the avatar step onset that was perturbed (shown here as a shortened interval by shifting the onset earlier). In this example, the participant onsets prior to the perturbation are before the avatars (denoted by a negative asynchrony).

Visual-Only conditions for the Fast condition (*p=.002*), with addition of the auditory cue resulting in the participants' step intervals becoming closer to those generated by the avatar (Fig 2). A similar result is seen for the Slow condition, but the change is not significant (*p=.087*; Fig 2b).

Despite the difference between modalities, participants consistently had slightly longer ISI (slower steps) than the avatar in the Fast condition for both Auditory-Visual (*t(11)=4.1100, p=.0017*) and Visual-Only conditions (*t(7)=13.7179, p<.0001*). However, at the slow tempo the ISI was not significantly different from the avatar across both Visual-Only and Audio-Visual conditions.

We found no significant differences in step interval variability, across either the tempo or modality conditions.

Synchronisation of steps

To analyse whether participants were able to synchronise their footsteps with the avatar, asynchronies were calculated between corresponding step onsets (see Fig 1 and Methods, Data Analysis section for how these were extracted). The asynchronies were then averaged across the trial up to the point of the perturbation, for each trial. Effects of each condition, on the mean and standard deviation of asynchrony, were analysed using a mixed measures ANOVA as before. The asynchrony was deemed negative when the participant led the avatar during a heel strike, and positive when the participant lagged behind as is commonly reported in the literature[7,9].

Mean asynchrony (Fig 3) was affected by both the avatar's stepping tempo ($F_{1,18}=8.533$, $p = 0.009$) and whether the cues included both auditory and visual or just visual information ($F_{1,18}=11.510$, *p=0.003*). Moreover, an interaction between Tempo and Modality ($F_{1,18}=17.445$, *p=0.001*; Fig 3a) highlighted that while asynchrony remained consistent for fast tempos regardless of modality (*t(15.682)=-0.066, p>0.05*), it switched from a negative (Visual-Only) to positive (Auditory-Visual) asynchrony for the Slow condition (*t(15.325)=6.219, p<.0001*).
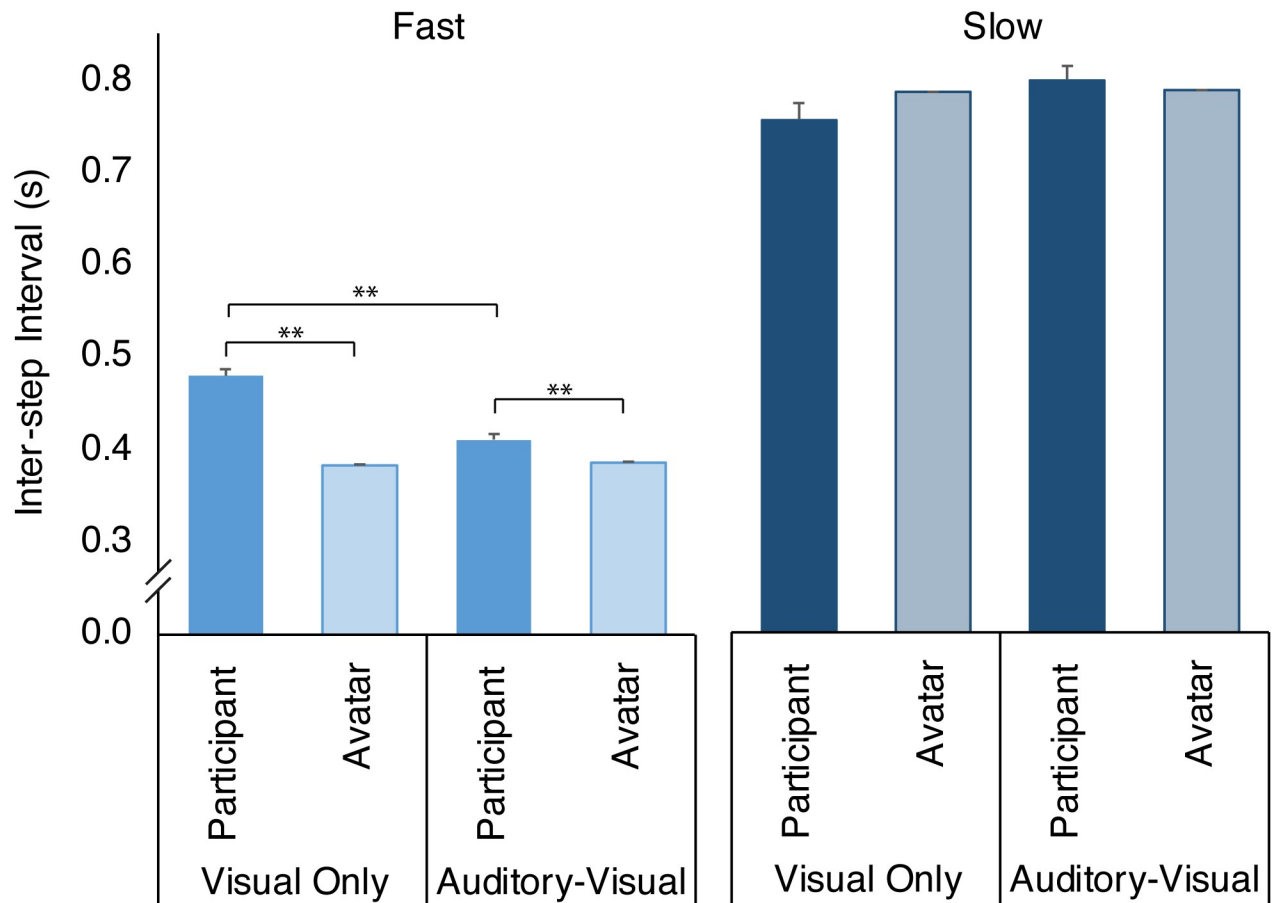
Fig 2. Mean Inter-step intervals (ISI) of participant steps (filled bar) compared to the target Avatar ISI (lighter shaded bars) pre-perturbation for all conditions. Error bars show standard error of the mean (SEM). * represents significance at p<.05, ** is p<.01.

Participants completing the Auditory-Visual condition showed similar positive mean asynchronies for both Fast and Slow tempos ($t(20.341)=-0.887$, $p>0.05$; Fig 3a), while those synchronising with the Visual-Only cues were on average ahead of the cue for the Slow tempo and lagged behind the cue for the Fast tempo ($t(13.148)=4.249$, $p<.0001$; Fig 3a).

Variability of asynchronies did not significantly differ between Fast and Slow tempos. However, there was a significant effect of Modality ($F_{1,18}=4.641$, $p=0.045$), with participants' timing being more stable with the addition of an auditory component to the cue (Fig 3b). This difference was significant for the fast condition ($t(17.616)=-2.145$, $p=0.046$; Fig 3b) though not for the slow condition($t(7.281)=-1.234$, $p>0.05$).

Correction in response to temporal perturbations

Relative asynchrony was calculated by subtracting the mean asynchrony prior to the perturbation from all resulting asynchronies in a trial. This measure eliminated the effects of individual participant tendencies to lead or lag behind the avatar when averaging across participants, with the relative asynchrony being approximately zero immediately prior to the perturbation event for each trial.

Participant asynchrony data was subsequently aligned so that the asynchrony corresponding to the perturbed step lay at step T. Plots for average asynchronies, from step T-4 to step T
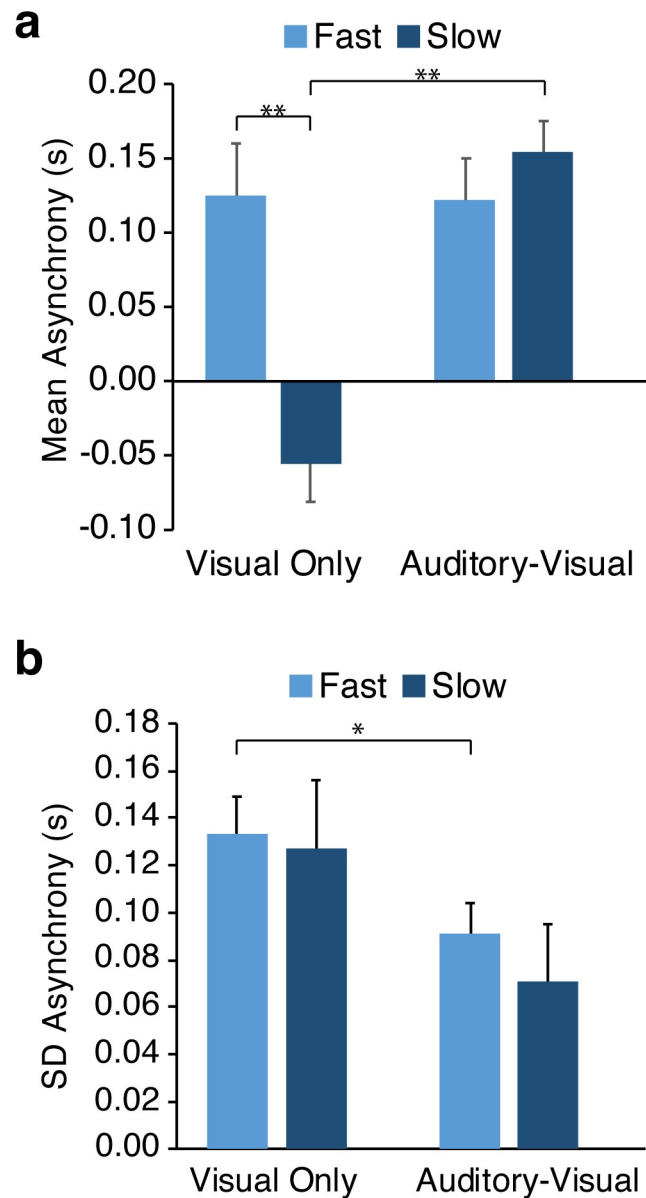
Fig 3. Asynchrony between participant step onsets and corresponding Avatar steps pre-perturbation. a). Mean Asynchrony between participant step onsets and corresponding Avatar steps, for all conditions. This asynchrony reflects baseline performance. Negative asynchronies indicate the participants are, on average, stepping ahead of the Avatar cue. Error bars show SEM. **b).** Mean standard deviation of asynchrony for all conditions. Error bars show SEM. * represents significance at p<.05, ** is p<.01.

+6, across participants for each trial type are shown in Figs 4 and 5. For the Visual-Only conditions asynchronies tended to exhibit slow drift in the Fast Visual-Only condition even before the perturbation, for both shortening and lengthening of steps, highlighting the lack of synchronisation between the participant and the avatar (Fig 4a and 4b). As the asynchrony continues to increase post-perturbation, it is assumed that there is no overall correction response to the perturbation for these conditions. For the Slow tempo, participants were able to maintain synchrony, shown by the asynchronies remaining close to zero. In particular, any drift seen in
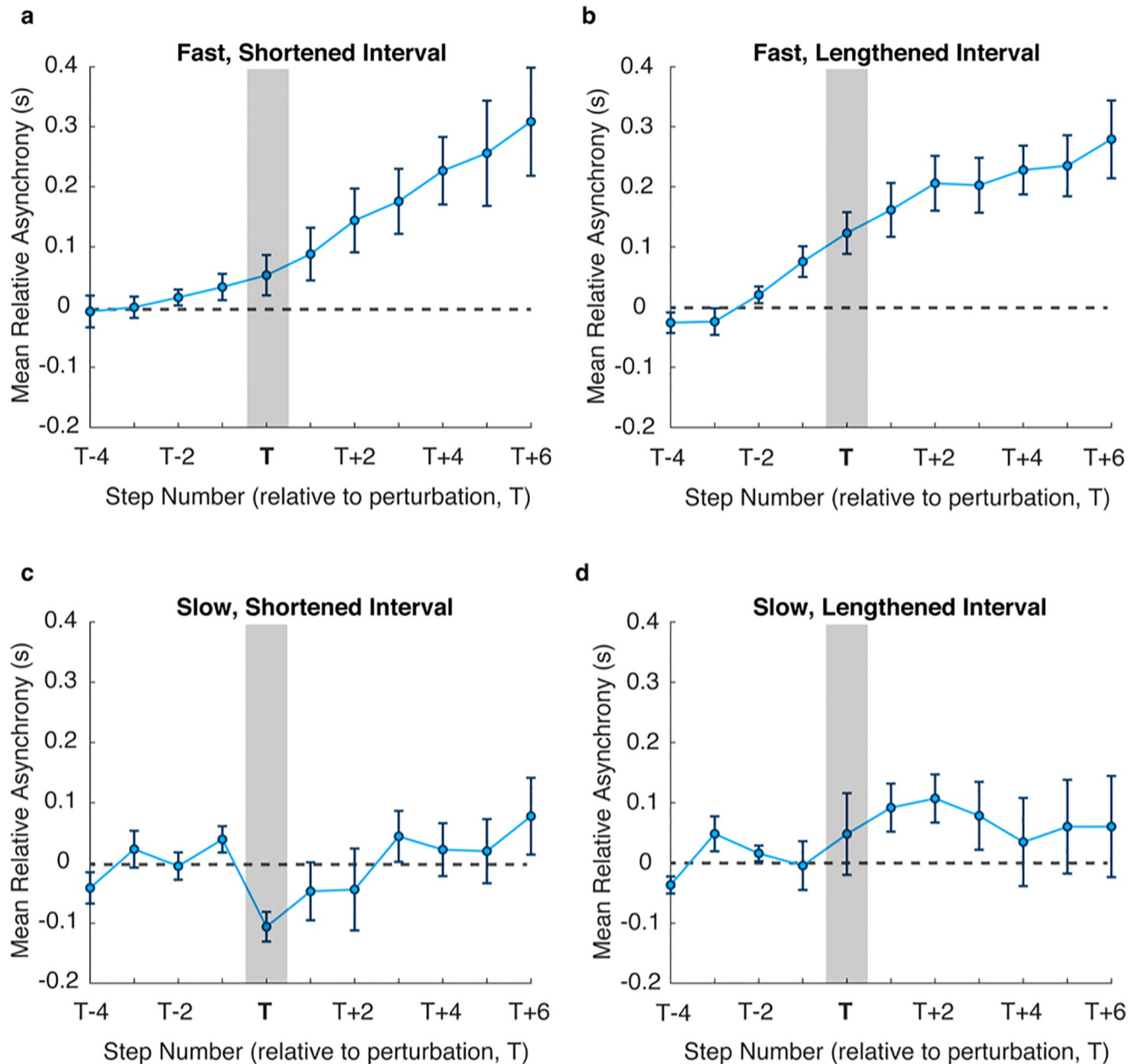
Fig 4. Mean relative asynchronies before and after the perturbation (vertical grey bar) at time T for Visual-Only cues. Dotted horizontal line shows the zero relative asynchrony measure, to which participants were expected to correct towards following the perturbation. Separate plots are shown for Fast, shortened (**a**) and lengthened (**b**) intervals, and Slow, shortened (**c**) and lengthened (**d**) intervals. Error bars show SEM.

the Fast conditions does not continue to exacerbate post-perturbation. The effect of the perturbation can be seen in Fig 4c at time T, with a sudden increase in negative asynchrony, and the participant correcting back to stability by around T+3. High inter-subject variability for the lengthened step (Fig 4d) results in a less pronounced increase in positive asynchrony at time T and less clear correction strategy.

In contrast, for the Auditory-Visual condition (Fig 5), asynchronies were much more stable post-perturbation, and inter-subject variability remained relatively low. There was a clear response following the perturbation in both fast and slow conditions. Participants corrected
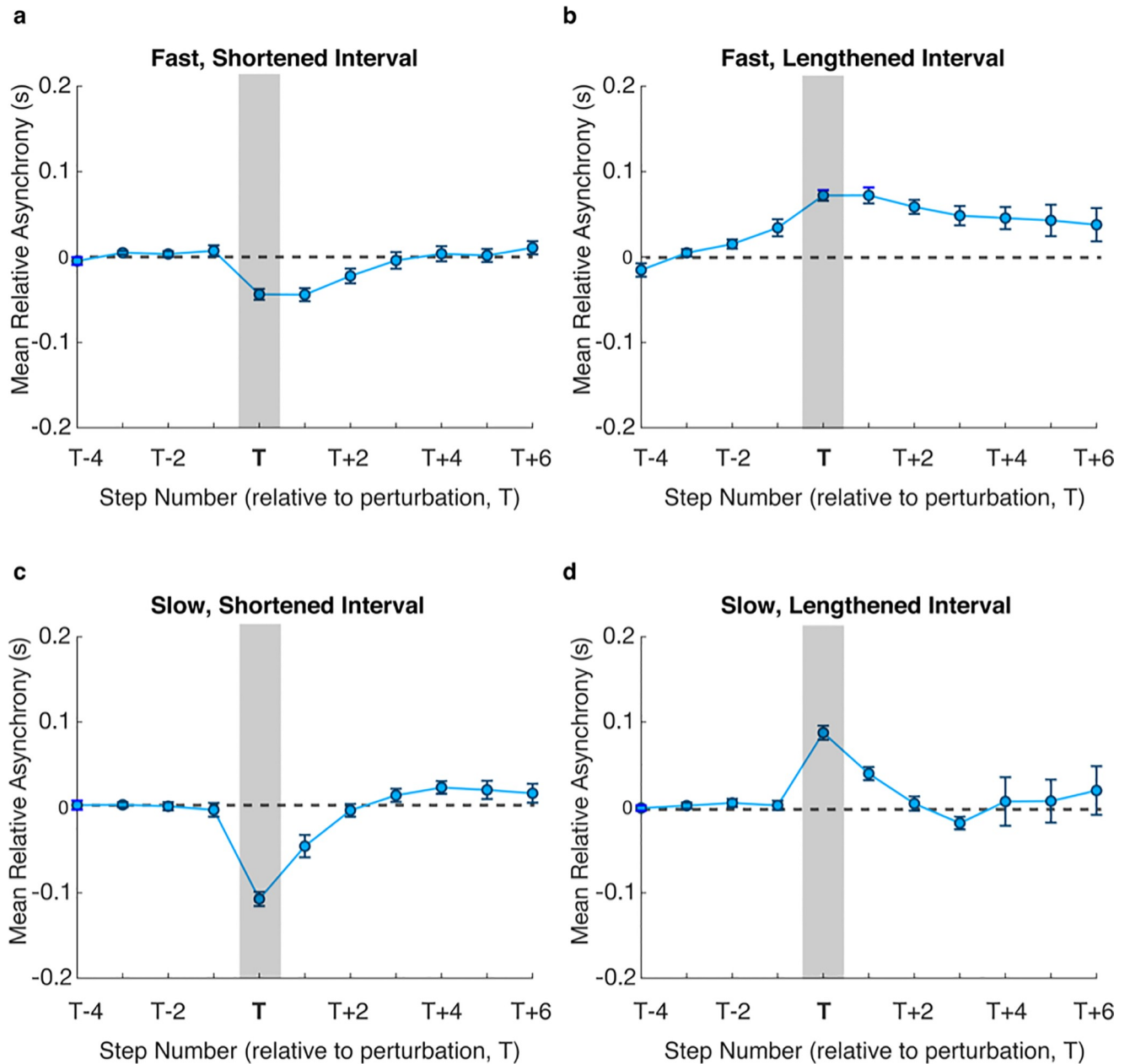
Fig 5. Mean relative asynchronies before and after the perturbation (vertical grey bar) at time T for Auditory-Visual cues. Dotted horizontal line shows the zero relative asynchrony measure, to which participants were expected to correct towards following the perturbation. Separate plots are shown for Fast, shortened (**a**) and lengthened (**b**) intervals, and Slow, shortened (**c**) and lengthened (**d**) intervals. Error bars show SEM.

within 3 steps, with correction normally initiated after the second step post-perturbation for the Fast condition. While there is some evidence of drift indicating a difference in the corrective response between fast positive and negative perturbation responses, comparing the correction gain showed there was no significant difference ($F_{1,18}=0.306$, $p=0.587$) between the directions of the perturbation.

To quantify the temporal corrections to the perturbation for each condition, a linear phase correction model was fitted to the participants' asynchronies following the perturbation, using

Fig 6. Mean correction gain as calculated by a linear phase correction model fitted to the asynchrony data of each participant. Higher values indicate participants are correcting their movements to a bigger proportion of the preceding step asy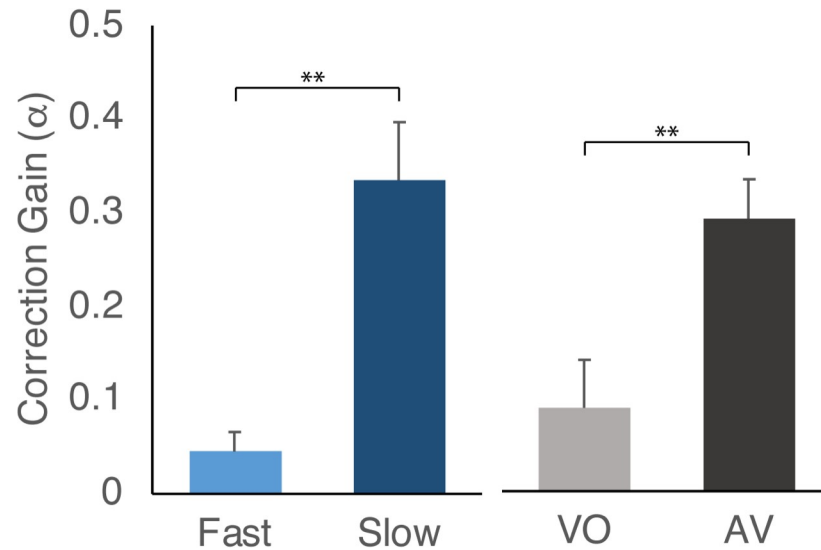nchrony between themselves and the Avatar. Correction gains are shown for Fast versus Slow and Visual-Only (VO) versus Auditory-Visual (AV) conditions. Error bars show SEM. * represents significance at p<.05, ** is p<.01.

https://doi.org/10.1371/journal.pone.0229641.g006

the method developed by Jacoby and colleagues[22]. Using the fitted model, the correction gain, α, was calculated (Fig 6) representing the average proportion of the last asynchrony corrected for on the next step ($0<\alpha<2$; see Methods). These were then analysed using a mixed measures ANOVA to investigate the effects of each condition on the participants' ability to correct their timing in response to the perturbed step cycle.

Substantiating the visual results from the plots of relative asynchrony (Figs 4 and 5), participants showed lower levels of correction to the Visual-Only condition, compared to the Auditory-Visual condition ($F_{1,18}=9.323$, $p=0.007$; Fig 6). Additionally, participants showed much lower correction when the Avatar was stepping Fast, compared to Slow step intervals, indicating that they struggled to keep up and synchronise with the shorter stepping cycles ($F_{1,18}=19.885$, $p<0.001$). There were no interactions between the conditions and the direction of the perturbation had no significant effect on the correction gain indicating that participants corrected similarly when speeding up their steps to recover synchrony as they did when slowing down.

Discussion

Here, we have investigated whether healthy participants are able to coordinate their movements accurately with a virtual avatar, using a sensorimotor synchronisation paradigm. Participants' step tempo and asynchrony, with respect to the avatar, were measured, while they attempted to step in time with the avatar at two different inter step intervals, 400ms (Fast) and 800ms (Slow). Correction gain was calculated in response to temporal perturbations to one of the avatar's step cycles, both speeded and slowed by 15% at both tempos resulting in four conditions.

We found that the multimodal cues (visual with auditory footsteps) reduced error in tempo and variability in asynchrony and increased participants' levels of temporal correction when stepping with an avatar. Participants were able to synchronise more accurately and showed

greater correction when the avatar stepped at a slow pace. However, participant step onsets appeared to consistently lag behind the avatar's by approximately 150ms across both tempos and modal conditions.

A key finding from the experiments is the difference in participant performance between Visual-Only and Auditory-Visual conditions. The results from the stepping trials show that the addition of auditory cues significantly increased the accuracy of synchronisation and greater correction compared to visual-only cues. Moreover, in the visual-only condition, participants struggled to maintain synchrony and for the fast condition, even struggled to match the avatar's tempo. Multimodal cues have been found to increase accuracy in other movement timing contexts ranging from simple cues involving finger tapping[16] and larger full body movement such as gait[7]. Bimodal cues, featuring visual as well as auditory components, have been found to improve perception and synchronization to auditory rhythms[23] and reduce variability in asynchrony[24] over unimodal stimuli. In this study, we combined the complex spatio-temporal movements of the avatar with a simple relevant foot-strike sound effect equivalent to stepping on a gritted surface. This simple additional modality has provided the necessary sensory-information for participants to improve their performance, substantially. The fact the cue is contextual to the action could also have contributed to the improved performance, with evidence that footstep sounds alone can be used to infer the kinematic and kinetic properties of walking[25]. While it could be argued that introducing the auditory cues could have resulted in participants ignoring the avatar's movements and exclusively using the discrete auditory events to synchronise their steps, there is overwhelming evidence, in the literature, that multisensory cues are integrated in a statistically optimal fashion described by maximum likelihood estimation[26], provided cues are relevant to each other[27,28]. Hence, it is considered here that the improvement in timing performance is due to the visual avatar cues being integrated with the relevant footstep auditory cues.

Previously, sensorimotor synchronisation research with visual stimuli has focussed on simple cues such as flashing lights or bouncing spheres, which, while having significant effects on certain aspects of gait such as step timing and stride length[29], have limited scope for altering motion, given the amount of information they can convey. Existing research suggests spatio-temporal visual stimuli can result in improved synchronisation and coordination over temporal only stimuli as long as the spatial information is compatible[8,13,14]. In this study, participants failed to entrain fully to the spatio-temporal visual-only cues generated by the Avatar. This is highlighted further when comparing average asynchrony results against existing literature assessing stepping on the spot synchronisation with metronome cues [7,9]. Mean asynchrony and SD of asynchrony were both significantly larger in our study: 97ms vs. -64ms[7] and -61ms[9] for mean asynchrony, and 101ms vs. 26ms[7] and 17ms[9] for SD of asynchrony, though both studies used 500ms as the metronome cue interval. This difference could be due to the immersive nature of virtual reality headsets, causing participants to lose out on visual feedback on their true surroundings and importantly their own limbs, requiring greater conscious attention to maintain synchrony at the cost of stability. This removal of sensory information could have made it difficult to align their own movements with the Avatar. This could be examined using augmented reality, AR, avatars rather than virtual reality based ones. AR would allow the participants to view their true environment and their own body, while still getting a stereo view of the avatar. While synchronisation with the avatar was worse than that reported using metronome cues for similar studies, which indicates that an avatar alone is not suitable for assisting temporal coordination, avatars can potentially have an advantage in that they provide rich spatial information. This is especially important for exercises that are more complex and future work will investigate at how participants matched the avatar spatially during the exercise.

Participants showed high deviation in terms of target tempo when attempting to match to fast cues, which resulted in the lack of synchrony. This subsequently increased variability when, based on previously observed Weber's law properties, lower variance should be observed for faster tempos[30]. Other studies, in both gait and stepping on the spot, have shown that participants are able to step in time with auditory and simple visual cues at an ISI of 500ms or above[7,9], closer to normal self-selected walking speeds. This indicates that fast tempos, such as 400ms, while reasonable for simple finger movements[17], is too fast for complex conscious full-body movements without the addition of a corresponding auditory stimulus. While no normative values for self-selected step-in-place exercises could be found, others [31] have listed step frequencies for 'Fast Gait' equivalent to the inter-step-intervals chosen for this study. Repp[17] also reports the rate limit for finger tapping as between 150-200ms. As the leg is a much larger limb, a longer ISI is expected and 400ms would be closer to the rate limit than to the self-selected pace. We therefore assume that the fast tempo was sufficiently different to participants' self-selected pace to induce active attempts to synchronise steps, but too fast to control and correct motor actions to achieve synchrony with the avatar.

Participants' step onsets tended to lag behind the avatar by approximately 150ms. This asynchrony is opposite to that found in the majority of the tapping literature[17] but has been observed in some studies which involve movement over larger amplitudes[32]. The decision to use the step onset, such that we could observe the synchronisation in the start of the movement rather than the end, could also be a contributing factor to this observation; most finger tapping studies reference to the time of the 'tap'–i.e. the end point of the movement when it hits a surface. Other studies into gait entrainment have shown that participants tend to synchronise with the peak velocity rather than a particular position within the gait cycle[33]. The velocity profile of the cue has also been shown to affect coordination, with slowing to endpoints and turning points in trajectory increasing the stability of coordination[34]. Alternatively, this asynchrony could also be due to participants' inability to determine discrete onset times when following the avatar cue, however, the lag persisted even in the presence of an auditory component. These issues highlight the potential for movement alignment to vary across the trajectory and future studies will need to look at the temporal-spatial matching of the participants' trajectories to that of the avatar, rather than discrete time points.

Responses to perturbations in the avatar's step cycle were clearly related to participants' synchronisation performance. Participants were unable to correct their timing when attempting to synchronise with fast stepping on the spot cues without the aid of footstep sounds and, in general, participants corrected better to slow rhythms. Correction gain increased significantly in the presence of auditory cues, with participants on average correcting within 2 to 3 steps after the perturbation. This is in agreement with existing literature, which states that the bimodal condition lowers variability in asynchrony and provides an advantage over visual only cues when correcting to phase perturbations[24]. This indicates that if participants are able to keep in sync they have 'confidence' in the cue and hence make larger and quicker corrections, corroborating results from Wright, Spurgeon & Elliott[7]. In their study, which used an LED visual cue as opposed to a fully immersive virtual avatar, Wright et al. found that mean correction to phase perturbed metronome cues, calculated as the percentage correction compared to the initial perturbation, ranged from 7.5% in the Visual-Only case to 17% Audio-Visual at an ISI of 500ms when averaged across the 5 steps post-perturbation. Additionally, Bank and Roerdink's study[6], assessing responses when treadmill walking at self-selected speeds, found an average correction of 18% in the 5 steps following a perturbation and Chen, Wing and Pratt[9] found a mean correction at step T+1 of 34.4% at an ISI of 500ms, which would be expected to reduce in subsequent steps. In our study, at 400ms, we found an equivalent correction response in the audio-visual condition. At 800ms, however, we found a

significantly better correction response, with correction gain, α, ranging from 0.265 for the Visual-Only condition to 0.405 for audio-visual, equivalent to 26.5% and 40.5%, when calculated using the method applied by Wright et al. [7], assuming the correction is linear. However, an increased correction gain is expected with a longer inter-onset-interval [35]. This suggests that avatar cues presented at a relatively slow pace and in a multisensory format induce the highest level of entrainment from participants, resulting in high levels of synchronisation and rapid correction to perturbations in the cue.

An important application, where the use of virtual avatars could be beneficial, is in the context of physiotherapy and rehabilitation. Following lower limb or neurological injury, exercise is often required as an intense part of a physiotherapy or rehabilitation process. In the rehabilitation of lower limb function, patients are set goal-oriented functional tasks, such as walking, turning and stepping over objects[36], depending on the nature of their impairment. In the hospital setting, these are normally performed in the presence of physio- or occupational-therapists, who provide feedback to the patient and update the training regimen as required[37]. Outside the clinic, patients are usually left with limited guidance on their exercise routine. Subsequently, patients can quickly lose motivation or complete exercises incorrectly without the presence of physiotherapist supervision[38]. The way an exercise is completed affects the muscles recruited, meaning good temporal and spatial matching to an ideal exercise execution would result in muscle recruitment closer to that of the ideal. As an avatar can present spatial and temporal information during an exercise, they have the potential to improve matching to an ideal execution. Our results suggest that a simple visual representation of an avatar is not enough to provide accurate movement guidance in the temporal context. Rather a multimodal approach is required, where auditory or possibly haptic cues are used to highlight events or end points in trajectories that will assist patients to coordinate accurately with an avatar.

Conclusion

In conclusion, in this study, we have demonstrated that humanoid cues within virtual reality have the potential to be used as exercise cues for entrainment, as participants are able to match and correct the timing of their movements to an avatar. However, it is important to consider this was only achieved through the use of multimodal cues (i.e. the addition of contextual footstep sounds) along with relatively slow movement tempos. Moreover, a fully immersive environment potentially reduced performance due to the removal of visual feedback on one's own body movements; hence, an augmented reality based avatar may improve levels of entrainment.

These results open up the use of virtual avatars for rehabilitation, allowing physiotherapists to present realistic exercises to patients with an 'ideal' leader to follow. These avatars could provide patients with natural cues, taking advantage of their ability to mirror the motion of other humans. Additionally, providing a humanoid cue, which does not vary between repetitions of an exercise, allows for participant motion to be analysed against any target motion, aiding in the identification and quantification of deficiencies in a participant's limb function. We recommend that in these scenarios multimodal components are presented in the form of auditory or haptic cues to highlight events or end-points in movements, which will aid the individual to accurately align the timing of their actions with the avatar. However, key to evaluating the suitability of virtual reality avatars, as a method of movement training or guidance, is to evaluate the level spatial entrainment (e.g., limb positions and orientations), observed in an individual following the avatar cues. This will be investigated in our future work.

Materials and methods

Participants

Twenty, healthy participants (male, right-handed, aged 23–39) were recruited opportunistically from the University of Warwick, in August 2015 and April 2016. Participants were eligible to take part as long as they were in good health, had no lower limb impairments, had normal or corrected to normal vision and did not suffer from hearing impairments. No data on other specific impairments were collected. Of the participants recruited, none reported any lower limb function disabilities and all took part in and completed the experiments. Participants were unaware of the purpose of the experiment and gave written informed consent prior to taking part. One participant reported having taken musical training (drums) but the remainder did not have any musical or dance related experience. Eight participants completed the first experiment ("Visual-Only" condition); twelve completed a second follow-on experiment ("Auditory-Visual" condition). The audio-visual study was conducted based on the results of the first visual-only study. As such, participants completed either the visual-only or audio-visual conditions, not both.

The study was approved by the University of Warwick Biomedical Sciences Research Ethics Committee (BSREC), reference REGO-2015-1582, and was conducted in accordance with the ethical standards of the 1964 Declaration of Helsinki.

Design

Each experiment consisted of a single session per participant, which lasted approximately 40 minutes. In both experiments, participants wore an Oculus Rift DK2 (Oculus VR, California, USA) virtual reality headset to view an avatar in a virtual environment (Fig 7), developed in Unity3D (Unity Technologies SF, California, USA).

During the session, 40 reflective markers were attached to the participants, placed using the "Plug-In-Gait"[21] model for capturing full body motion. Participants stepped on the spot inside a 3m by 4m marked space and their movement was tracked and recorded by a 3D motion capture system (Vicon Vantage, 12 cameras) at a sampling rate of 100Hz.

Fig 7. View of the avatar seen by participants when completing the stepping task.

https://doi.org/10.1371/journal.pone.0229641.g007

A stepping exercise was chosen as it required cyclic coordination of the lower limbs in a very similar way to gait, but could be done on the spot reducing risks and practical issues of moving around a room whilst wearing a VR headset. Participants were instructed to step on the spot in time with the avatar (see Stimuli below), which was viewed on the headset. The orientation was designed such that the participant stood behind the avatar, with the whole body in view. This was to ensure the avatar's feet were not hidden as occlusion of end-points has been found to increase variability in coordination[39]. Participants completed either Experiment 1 (Visual-Only) or Experiment 2 (Auditory-Visual), which involved 4 blocks of 5 trials (20 trials in total) run in succession. Short breaks were offered between blocks, every 10–15 minutes, and additional breaks could also be requested by participants. Each trial consisted of the participant synchronising a total of 30 steps with the Avatar.

One interval from step 10–16 was chosen to be perturbed. The intervals pre-perturbation were used for the ISI and asynchrony analyses, with the first three steps excluded. The intervals, post-perturbation, were used for the correction gain analysis. More information about the trials and analysis is given later in this section.

The experiment was a 2 (Modality: Visual-Only, Auditory-Visual) x 2 (Tempo: Slow, Fast) x 2 (Perturbation Direction: Positive, Negative) design. Tempos represented the time between step onsets, denoted as the inter-step intervals. For the slow condition, this was approximately 800ms, while the fast condition was approximately 400ms. Step intervals generated by the avatar were not exactly at the target value as the avatar movement data was generated by real human motion following a metronome set to the target intervals. Therefore, some variation due to this and frame rate of the headset occurred.

Participants who completed Experiment 1, followed the avatar without any auditory cues, i.e. Visual-Only. This was performed in a quiet room with a relative soft surfaced floor, so auditory feedback from their own footsteps was minimal. Participants who completed Experiment 2 completed the experiment with the same visual cues but with footstep sounds (shoe onto gritty/dirt surface) accurately overlaid to align with the heel strikes of the avatar.

For all trials, one step cycle of the avatar was randomly chosen (between steps 10 and 16), to be temporally shortened or lengthened by 15%, creating a perturbation. A perturbation of 15% was chosen, as this was large enough to elicit a corrective response but small enough that, on the whole, participants were unaware of this perturbation occurring during the step cycle. This is similar to the perturbations used by Wright et al[24] of between 12.5% and 20% for slow and fast stepping trials, respectively.

Participants were not trained with, or acclimatised to, the system before the session, but did receive one practice trial to familiarise with the task. The order of the within-subject conditions were randomised to cancel out any potential learning effects which may have occurred over the course of the session.

Stimuli

The avatar cue was developed by recording marker trajectories on a single volunteer (Male, 37 years) using the same Vicon motion capture system and full body market set used to collect participant data. The volunteer was instructed to step on the spot in time with an auditory metronome at either a fast (metronome interval: 400ms) or slow (metronome interval: 800ms) tempo. These tempos were chosen to ensure participants had to actively attempt to stay in synchrony with the avatar rather than stepping at a speed close to what is reported as a typical self-selected pace (545ms[40] for gait, no normative values for stepping on the spot exercises could be found in the literature).

The virtual environment was developed using the Unity3D game engine. The scene presented in the application was chosen to be minimal, with no background movement or objects, to avoid distractions and maintain subject attention while the trials were being completed, as high density environments have previously been found to affect step kinematics in walking experiments[41].

Humanoid characters were obtained from Adobe Mixamo (Adobe Systems Incorporated, California, USA). A male character was selected to represent the male volunteer who provided the movement data to be mapped on to the avatar. The motion data was imported into Unity3D, and a single cycle of stepping on the spot, which matched closest to the original metronome cue, was mapped onto the avatar. This was done to minimise the temporal and spatial variability in the exercise cue, as variance in an exercise cue is known to adversely affect sensorimotor synchronisation, increasing the variance and phase correction and reducing correction gain[42]. While minimising the variance in the exercise cue would reduce the natural nature of the movement, this ensured that the variance found in participants would be largely due to their ability to follow the avatar rather than the avatar varying over the trial. The avatar developed was used across all trials and did not adapt based on participant motion.

When viewing through the headset, participants were positioned behind the avatar who was facing away from them. This was representative of tasks requiring a "follow the leader" approach, such as dance classes, minimising the incongruence of the stimuli presented. This view also allowed users to easily internalise the motion of the avatar and compare their own motion to it, as their orientation matched that of the avatar[43]. The avatar's anthropometrics did not adapt based on the subject but the position of the headset during the trial was fixed at a predefined height to ensure the top of the avatar was at eye level and each participant had the same view of the avatar. Orientation of the participant's view was obtained directly from the virtual reality headset used to present the scene to reduce the lag between user motion in the physical and virtual worlds.

The trajectories of the avatar limbs, from the end of the chosen step cycle to the start, were interpolated to allow the step cycle to play repeatedly without jittering. A single step cycle was chosen, instead of an average of multiple step cycles, in order to keep the cue motion as natural as possible.

To create the Perturbation, one step cycle, between the 10th and 16th step was chosen at random, and either accelerated (resulting in a shortened step interval, or negative perturbation) or decelerated (resulting in a lengthened step interval, or positive perturbation) by 15%. Accelerated or decelerated intervals were achieved by changing the playback speed during the step interval period. The step chosen randomly to be perturbed was achieved via a script within Unity3D.

Natural footstep sounds, representing a shoe on a gritty surface, were obtained from the Unity Asset Store and synchronised with the avatar heel strikes for the second Auditory-Visual condition. The phenomenon of spatial ventriloquism[44], where temporally synchronous sound is perceived to occur at the location of a visual stimulus, meant that the origin of the sound did not need to be placed at the base of the avatar in order for participants to perceive it as originating from the feet.

Procedure

A hardware trigger was developed to synchronise trial recordings from Vicon with the recording of the exercise cue within Unity3D. Within the application, and once a trial was started, triggering the movement of the avatar within Unity3D, a rising edge was sent to the computer's serial port. This was connected to the Vicon system, which was configured to start

recording when a rising edge was received, thereby synchronising the start times of the recordings from the two systems.

The scene, including the trial animations, logic for selecting trials and perturbing steps, was driven by the Unity3D application running on a Windows 7 desktop PC. This was presented to the participants through the Oculus Rift DK2 headset at 75Hz, with a resolution of 960x1080 pixels per eye.

Time-stamped trajectories for all limbs were captured, by both the Vicon and Unity3D systems, for the participant and avatar, respectively. As each system captured at a different rate, the Vicon system capturing at 100Hz versus the Unity3D/Oculus system recording at approximately 75Hz, the recordings were synchronised by timestamp, rather than frame, offline in MATLAB.

Participants self-initiated each trial by pressing a key positioned on the floor with their foot, after which the avatar started playing the exercise cue. Participants were instructed to step in time with the avatar, matching left and right leg movement. Participants were not informed of the presence of a perturbation in the step cycle.

Data analysis

The stepping data was analysed by extracting the left and right step onset times of both the participant and the avatar for every trial. These were extracted in MATLAB by identifying the time point when the vertical axis coordinate of the captured heel marker rose above a predefined threshold height above the floor surface. Participant heel strike times were temporally aligned to the avatar onsets, allowing for calculation of timing errors, or asynchronies, between each heel strike and the corresponding cue. The duration between heel strikes was used to measure the inter-step interval (ISI) for comparison of tempo.

In the case where the participant ISI was consistently different from the avatar ISI, phase drift occurred, where participant motion moved from being in-sync to anti-phase to in-sync in later steps. This caused a wrapping effect in the asynchrony as, initially, the closest steps were chosen as being corresponding. When this was observed, corresponding steps were assigned to ensure continuous asynchrony between steps. In certain cases, this meant asynchronies appeared to grow larger than the cue's inter-step interval but this was required to correctly calculate mean asynchrony across steps (Fig 8).

After the unwrapping of asynchronies, a total of 7 fast tempo trials out of 76 for the Visual-only condition and 2 out of 120 or the Audio-visual condition were excluded from further analysis due to multiple steps being missed. Similarly, for slow trials, 2 out of 62 from the Visual-Only and 1 out of 120 from the Auditory-Visual condition were removed.

Mean asynchrony and ISI were calculated pre-perturbation, from step 3 to the perturbed step, for each participant within each condition. These were subsequently averaged across participants. The first 3 steps were eliminated from the analysis to remove the steps participants used to initially synchronise with the avatar after the abrupt start as it has been found that there is significant variability between subjects in time to synchronisation, and working memory plays an important role in synchronisation during the initiation phase[45]. This resulted in between 7–13 steps per trial being used for the inter-step interval and asynchrony analyses pre-perturbation.

Means from the recording of the avatar during the trial were used in the statistical analyses, instead of the original metronome tempo as the avatar was subject to the natural variance of the original volunteer. Dropped frames and decreases in frame-per-second (FPS) rates when the application was running also had an effect on the final tempo of the avatar observed by the participants, which varied between trials.
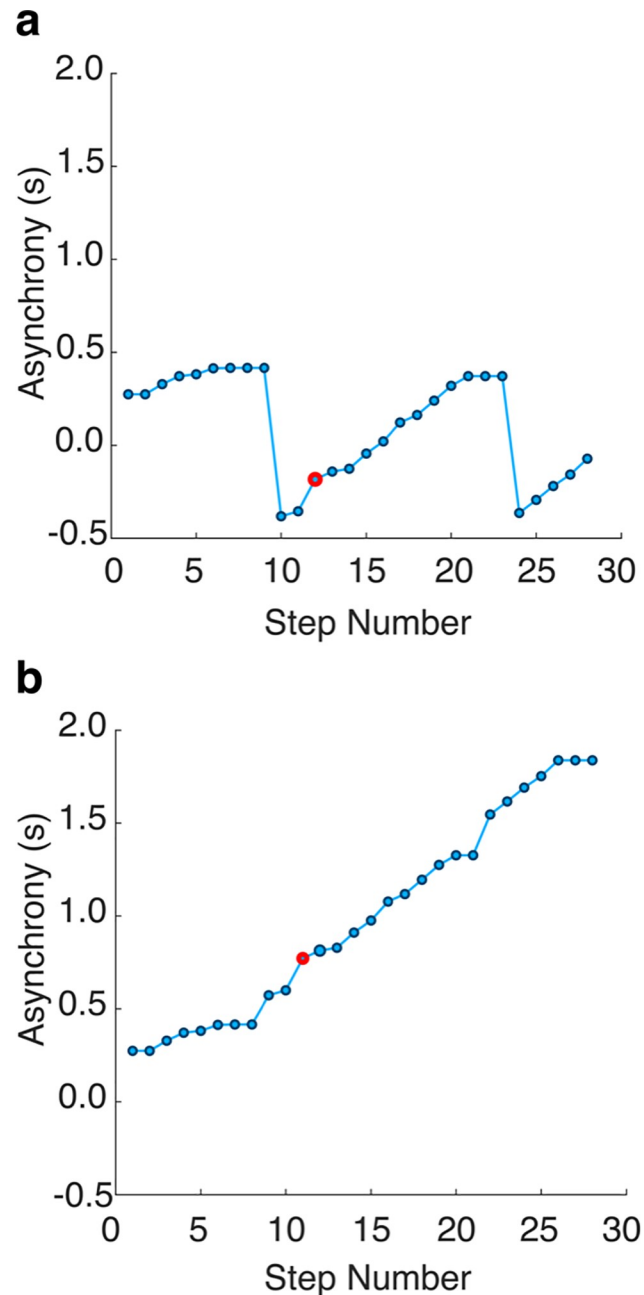
a

b

Fig 8. Example of asynchronies before and after phase wrapping removed. a). Asynchronies observed for a fast tempo trial when matching participant onsets to the nearest Avatar step onsets. Wrapping of the asynchrony can be seen between the intervals of +/-0.4 seconds which does not correspond to an attempt to regain synchrony. **b).** Asynchronies for the same trial but with phase wrapping removed and steps assigned to ensure asynchronies are continuous. This is a severe example of the drift observed in some fast tempo visual-only trials to demonstrate the unwrapping procedure.

https://doi.org/10.1371/journal.pone.0229641.g008

A mixed measures ANOVA was used to test for significant differences in mean ISI and asynchrony between conditions, auditory-visual versus visual-only cues and fast versus slow tempos. Paired t-tests (with adjusted p-value for multiple comparisons, using a Bonferroni

correction) were used to determine whether means and standard deviations of the asynchrony and inter-step interval were significantly different from the avatar cue.

Relative asynchronies were calculated by obtaining the mean asynchrony for all steps pre-perturbation per trial type per participant and subtracting these from the steps post-perturbation, to eliminate effects the of individual participant tendencies to lag or lead the avatar. The relative asynchronies for each step post perturbation were used to investigate the corrective response. This was achieved by fitting a linear phase correction model[22,46,47] to the asynchronies, based on the recurrence equation shown in Eq. 1:

$$A_{n+1} \;=\; (1-\alpha)A_n + T_n + M_{n+1} - M_n - C_n \tag{1}$$

where α is the correction gain, A_n is the current event asynchrony, T_n is the time interval generated by an assumed internal timekeeper, M_n is the current motor implementation delay and C_n is the corresponding step interval time of the avatar. A correction gain of 1 is therefore full correction of the previous asynchrony, while a correction gain of 0 indicates no correction.

The bounded General Least Square approach developed by Jacoby et al[22] was used to estimate the parameters of the model fitted to each trial. In particular, we extracted the correction gain, α, to determine the level of correction participants exhibited under each of the conditions. Correction gain values obtained per participant per trial type were analysed using a Mixed Measures ANOVA to find significant changes between conditions.

Acknowledgments

Author Contributions

Conceptualization: Omar Khan, Theodoros N. Arvanitis, Mark T. Elliott.

Data curation: Omar Khan, Imran Ahmed, Joshua Cottingham, Musa Rahhal.

Formal analysis: Omar Khan, Mark T. Elliott.

Investigation: Omar Khan, Imran Ahmed, Joshua Cottingham, Musa Rahhal.

Methodology: Omar Khan, Imran Ahmed, Musa Rahhal, Mark T. Elliott.

Software: Omar Khan, Joshua Cottingham.

Supervision: Theodoros N. Arvanitis, Mark T. Elliott.

Validation: Omar Khan.

Writing – original draft: Omar Khan.

Writing – review & editing: Omar Khan, Theodoros N. Arvanitis, Mark T. Elliott.

References

1. Rizzolatti G. and Craighero L., 'the Mirror-Neuron System', *Annu. Rev. Neurosci.*, vol. 27, no. 1, pp. 169–192, 2004.

2. Haggard P., Rossetti Y., and Kawato M., *Sensorimotor Foundations of Higher Cognition*. Oxford University Press, 2008.

3. Rizzolatti G., Fogassi L., and Gallese V., 'Neurophysiological mechanisms underlying the understanding and imitation of action', *Nat. Rev. Neurosci.*, vol. 2, no. 9, p. 661, Sep. 2001.

4. Su Y., 'Sensorimotor synchronization with different metrical levels of point-light dance movements', *Front. Hum. Neurosci.*, vol. 10, no. April, pp. 1–15, 2016.

5. Khan O., Ahmed I., Rahhal M., Arvanitis T. N., and Elliott M. T., 'Step in time: exploration of synchrony and timing correction in response to virtual reality avatars for gait re-training', in *Proc. 11th Intl Conf. on Disability*, *Virtual Reality and Assoc. Technologies*, 2016, pp. 323–326.

6. Bank P. J. M., Roerdink M., and Peper C. E., 'Comparing the efficacy of metronome beeps and stepping stones to adjust gait: Steps to follow!', *Exp. Brain Res.*, vol. 209, no. 2, pp. 159–169, Mar. 2011.

7. Wright R. L., Spurgeon L. C., and Elliott M. T., 'Stepping to phase-perturbed metronome cues: multisensory advantage in movement synchrony but not correction', *Front Hum Neurosci*, vol. 8, p. 724, 2014.

8. Hove M. J., Spivey M. J., and Krumhansl C. L., 'Compatibility of motion facilitates visuomotor synchronization', *J Exp Psychol Hum Percept Perform*, vol. 36, no. 6, pp. 1525–1534, Dec. 2010.

9. Chen H. Y., Wing A. M., and Pratt D., 'The synchronisation of lower limb responses with a variable metronome: The effect of biomechanical constraints on timing', *Gait Posture*, vol. 23, no. 3, pp. 307–314, 2006.

10. Zivotofsky A. Z., Gruendlinger L., and Hausdorff J. M., 'Modality-specific communication enabling gait synchronization during over-ground side-by-side walking', *Hum. Mov. Sci.*, vol. 31, no. 5, pp. 1268–1285, Oct. 2012.

11. Booth A. J. and Elliott M. T., 'Early, but not late visual distractors affect movement synchronization to a temporal-spatial visual cue', *Front. Psychol.*, vol. 6, 2015.

12. Meerhoff L. A., De Poel H. J., and Button C., 'How visual information influences coordination dynamics when following the leader', *Neurosci. Lett.*, vol. 582, pp. 12–15, Oct. 2014.

13. Armstrong A. and Issartel J., 'Sensorimotor synchronization with audio-visual stimuli: limited multisensory integration', *Exp. Brain Res.*, vol. 232, no. 11, pp. 3453–3463, Nov. 2014.

14. Armstrong A., Issartel J., Varlet M., and Marin L., 'The supplementation of spatial information improves coordination', *Neurosci. Lett.*, vol. 548, pp. 212–216, Aug. 2013.

15. Su Y.-H., 'Audiovisual beat induction in complex auditory rhythms: Point-light figure movement as an effective visual beat', *Acta Psychol. (Amst.)*, vol. 151, pp. 40–50, Sep. 2014.

16. Elliott M. T., Wing A. M., and Welchman A. E., 'Multisensory cues improve sensorimotor synchronisation', *Eur. J. Neurosci.*, vol. 31, no. 10, pp. 1828–1835, May 2010.

17. Repp B. H., 'Sensorimotor synchronization: A review of the tapping literature', *Psychon. Bull. Rev.*, vol. 12, no. 6, pp. 969–992, 2005.

18. Repp B. H., 'Automaticity and voluntary control of phase correction following event onset shifts in sensorimotor synchronization.', *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 28, no. 2, pp. 410–430, Apr. 2002.

19. Repp B. H., 'Phase Correction, Phase Resetting, and Phase Shifts After Subliminal Timing Perturbations in Sensorimotor Synchronization', *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 27, no. 3, pp. 600–621, 2001.

20. Elliott M. T., Welchman A. E., and Wing A. M., 'Being discrete helps keep to the beat', *Exp. Brain Res.*, vol. 192, no. 4, pp. 731–737, Feb. 2009.

21. Davis R. B., Ounpuu S., Tyburski D., and Gage J. R., 'A gait analysis data collection and reduction technique', *Hum. Mov. Sci.*, vol. 10, no. 5, pp. 575–587, 1991.

22. Jacoby N. et al., 'Parameter Estimation of Linear Sensorimotor Synchronization Models : Phase Correction, Period Correction, and Ensemble Synchronization', *Timing Time Percept.*, vol. 3, no. 1–2, pp. 52–87, May 2015.

23. Su Y.-H., 'Visual enhancement of auditory beat perception across auditory interference levels', *Brain Cogn.*, vol. 90, pp. 19–31, Oct. 2014.

24. Wright R. L. and Elliott M. T., 'Stepping to phase-perturbed metronome cues: multisensory advantage in movement synchrony but not correction', *Front Hum Neurosci*, vol. 8, no. September, p. 724, 2014.

25. Young W., Rodger M., and Craig C. M., 'Perceiving and reenacting spatiotemporal characteristics of walking sounds', *J. Exp. Psychol. Hum. Percept. Perform.*, 2013.

26. Alais D., Newell F. N., and Mamassian P., 'Multisensory processing in review: From physiology to behaviour', *Seeing and Perceiving*. 2010.

27. Körding K. P., Beierholm U., Ma W. J., Quartz S., Tenenbaum J. B., and Shams L., 'Causal inference in multisensory perception', *PLoS ONE*, 2007.

28. Elliott M. T., Wing A. M., and Welchman A. E., 'Moving in time: Bayesian causal inference explains movement coordination to auditory beats', *Proc. R. Soc. B Biol. Sci.*, 2014.

29. Sejdić E., Fu Y., Pak A., Fairley J. a, and Chau T, 'The effects of rhythmic sensory cues on the temporal dynamics of human gait', *PLoS ONE*, vol. 7, no. 8, 2012.

30. Repp B. H., 'Variability of timing in expressive piano performance increases with interval duration', *Psychon. Bull. Rev.*, 1997.

31. Oberg T., Karsznia A., and Oberg K., 'Basic gait parameters: reference data for normal subjects, 10–79 years of age.', *J. Rehabil. Res. Dev.*, vol. 30, no. 2, pp. 210–23, 1993.

32. Rodger M. W. M. and Craig C. M., 'Timing movements to interval durations specified by discrete or continuous sounds', *Exp Brain Res*, vol. 214, no. 3, pp. 393–402, Aug. 2011.

33. Su Y.-H., 'Peak velocity as a cue in audiovisual synchrony perception of rhythmic stimuli', *Cognition*, vol. 131, no. 3, pp. 330–344, Jun. 2014.

34. Varlet M., Coey C. A., Schmidt R. C., Marin L., Bardy B. G., and Richardson M. J., 'Influence of stimulus velocity profile on rhythmic visuomotor coordination', *J Exp Psychol Hum Percept Perform*, vol. 40, no. 5, pp. 1849–1860, Oct. 2014.

35. Repp B. H., 'Tapping in Synchrony With a Perturbed Metronome: The Phase Correction Response to Small and Large Phase Shifts as a Function of Tempo', *J. Mot. Behav.*, vol. 43, no. 3, pp. 213–227, May 2011.

36. Bovend'Eerdt T. J. H., Botell R. E., and Wade D. T., 'Writing SMART rehabilitation goals and achieving goal attainment scaling: a practical guide.', *Clin. Rehabil.*, vol. 23, no. 4, pp. 352–361, 2009.

37. Johnson L., Burridge J. H., and Demain S. H., 'Internal and external focus of attention during gait re-education: an observational study of physical therapist practice in stroke rehabilitation.', *Phys. Ther.*, vol. 93, no. 7, pp. 957–66, 2013.

38. Campbell R., 'Why don't patients do their exercises? Understanding non-compliance with physiotherapy in patients with osteoarthritis of the knee', *J. Epidemiol. Community Health*, vol. 55, no. 2, pp. 132–138, Feb. 2001.

39. Hajnal A., Richardson M. J., Harrison S. J., and Schmidt R. C., 'Location but not amount of stimulus occlusion influences the stability of visuomotor coordination', *Exp. Brain Res.*, vol. 221, no. 3, pp. 351–355, 2012.

40. Pietraszewski B., Winiarski S., and Jaroszczuk S., 'Three-dimensional human gait pattern—reference data for normal men', *Acta Bioeng. Biomech.*, vol. 14, no. 3, pp. 9–16, 2012.

41. Jelić A., Appert-Rolland C., Lemercier S., and Pettré J., 'Properties of pedestrians walking in line. II. Stepping behavior', *Phys Rev E*, vol. 86, no. 4, p. 46111, Oct. 2012.

42. Repp B. H., Keller P. E., and Jacoby N., 'Quantifying phase correction in sensorimotor synchronization: Empirical comparison of three paradigms', *Acta Psychol. (Amst.)*, vol. 139, no. 2, pp. 281–290, 2012.

43. Rio K. W., Rhea C. K., and Warren W. H., 'Follow the leader: Visual control of speed in pedestrian following', *J. Vis.*, vol. 14, no. 2, pp. 4–4, 2014.

44. Chen L. and Vroomen J., 'Intersensory binding across space and time: A tutorial review—Springer', *Atten Percept Psychophys*, vol. 75, pp. 790–811, 2013.

45. V Kurgansky A., 'Visuomotor synchronization: Analysis of the initiation and stable synchronization phases', *Hum Physiol*, vol. 34, no. 3, pp. 289–298, Jun. 2008.

46. Vorberg D. and Schulze H.-H., 'Linear Phase-Correction in Synchronization: Predictions, Parameter Estimation, and Simulations', *J. Math. Psychol.*, vol. 46, no. 1, pp. 56–87, Feb. 2002.

47. Vorberg D. and Wing A., *Handbook of perception and action*. 1996.