Master's Thesis

The cross-modal congruency effect as an objective measure of embodiment

July 7, 2020

Master's Thesis (27,5 ECTS) – Master Applied Cognitive Psychology
Course Code: 201800483
Faculty of Social and Behavioral Sciences, Utrecht University

Author:
P.W.J. (Pim) Verhagen, MSc
Student number: 5542189
Email: p.w.j.verhagen@students.uu.nl

First assessor:
Dr. Krista Overvliet
k.e.overvliet@uu.nl

Second assessor:
Prof. dr. Stefan van der Stigchel

External supervisor:
Dr. Anne-Marie Brouwer
anne-marie.brouwer@tno.nl

Utrecht University

TNO innovation for life

Abstract

The remote control of robots (telerobotics) generally requires a high level of expertise and may impose a considerable cognitive burden on operators. A sense of embodiment over a remote-controlled robot might enhance operators’ task performance and reduce cognitive workload. In this study, we aimed to validate the cross-modal congruency effect (CCE) as an objective measure of embodiment under four conditions with different, a priori expected levels of embodiment, and by comparing CCE scores with subjective questionnaire reports. The conditions were (1) a real hand baseline condition, (2) a real hand seen through a telepresence unit, (3) a robotic hand seen through a telepresence unit, (4) and a human-looking virtual hand seen through VR glasses. We found no unambiguous evidence that the magnitude of the CCE was affected by the degree of visual realism in each of the four conditions. We proposed several factors that may explain this outcome: the degree of spatial uncertainty about the vibrotactile target location, the relative timing of the target and distractor stimuli, the spatial separation between the vibrotactile target and the visual distractor, and the high cognitive workload during the CCT in the VR condition. We also found no evidence to support the hypothesis that the CCE and embodiment scores as assessed by the subjective questionnaire reports are correlated. Based on these findings, it can be concluded that the CCE may not be a robust measure of embodiment. Therefore, it is recommended that future studies focus on other behavioral and physiological measures to quantify embodiment.

Keywords: teleoperation, embodiment, multisensory integration, visuotactile integration, cross-modal congruency effect, cross-modal congruency task

Introduction

Robotic teleoperation systems allow human operators to remotely perform tasks in inaccessible areas (e.g., space and deep-sea exploration) or in unpredictable and hazardous environments such as minimally invasive surgery, building inspection and disaster response (Toet et al., 2020). Typically, these tasks require a high level of expertise and may impose a considerable cognitive burden on operators (Hedayati et al., 2018). By letting operators feel as if they were physically present at the remote environment, they can interact more naturally, which might enhance task performance and reduce cognitive workload (Almeida et al., 2017; Sanchez-Vives & Slater, 2005). In an ideal situation, operators should have the (illusory) feeling that the robot's body and hands are their own body and hands, so that they do not notice the operation is being mediated. This feeling is often referred to as the sense of embodiment[1], or embodiment for short, and has been defined as the sense that emerges when an object's properties are processed as if they were the properties of one's own biological body (Kilteni et al., 2012). The concept of embodiment can be divided into three subcomponents: *sense of ownership* (e.g., Krom et al., 2019), *sense of agency* (e.g., Newport et al., 2010) and *sense of self-location* (e.g., Arzy et al., 2006). We use the term embodiment as the overarching construct of these three subcomponents for the remainder of this paper.

Numerous studies have found that it is possible to induce a sense of embodiment over extracorporeal objects with a varying degree of visual realism. First studies on this topic involved the classical rubber hand illusion (RHI), in which participants have the feeling that a rubber hand becomes part of their body when it is stroked synchronously with their hidden real hand (Botvinick & Cohen, 1998). This illusion is induced through the multisensory integration between what is seen on the rubber hand and felt on the real hand (Tsakiris & Haggard, 2005). Since then, feelings of embodiment have been induced over robotic hands (Aymerich-Franch et al., 2017a, 2017b; Marini et al., 2014; Romano et al., 2015) and virtual bodies and body parts (Krom et al., 2019; Ma & Hommel, 2013; Maselli & Slater, 2014; Slater et al., 2008, 2010) through multisensory stimulation. However, these studies have employed different measures (e.g. subjective reports, proprioceptive drift) to quantify embodiment, making it hard to compare their results. Therefore, this study aims to compare embodiment strength of robotic and virtual hands using both subjective reports and an objective measure of embodiment.

In the literature a range of different measures has been used to quantify embodiment and its subcomponents, including mainly subjective questionnaires in which people rate their agreement with several statements reflecting the sense of embodiment (Longo et al., 2008; Aymerich-Franch et al., 2017a; Pritchard et al., 2016), but also more objective measures such as *proprioceptive drift* toward fake body parts in RHI and full-body illusions (Botvinick & Cohen, 1998; Longo et al., 2008; Riemer et al., 2015) and behavioral and physiological responses. Proprioceptive drift is measured by asking participants to indicate the perceived location of their hand before and after induction of the illusion. Generally, participants perceive their hand to be closer to the artificial hand after induction, suggesting a stronger feeling of ownership of the artificial hand (Riemer et al., 2015). Behavioral and physiological responses include, among others, measurements of *brain activity* through functional Magnetic Resonance Imaging (fMRI; Ehrsson et al., 2004, 2007; Tsakiris et al., 2010) or Electroencephalography (EEG; González-Franco et al., 2014; Škola & Liarokapsis, 2016), *heart rate deceleration* (HRD; Slater et

[1] For a conceptual differentiation between embodiment and sense of embodiment see De Vignemont (2011).

al., 2010), *skin conductance response* (SCR; Armel & Ramachandran, 2003) and *electromyographic (EMG) onset activity* (Slater et al., 2008). It has been shown that subjects display similar levels of brain activity, HRD and SCR in response to threats when their artificial body (part) is threatened as when their real body (part) is threatened, indicating that they feel like the artificial body (part) is their own. Moreover, Slater et al. (2008) showed that after induction of the virtual hand illusion, EMG onset activity in the right arm increased after the virtual hand suddenly started to rotate, suggesting that they experienced a sense of embodiment over the virtual hand.

Although all the measures presented above are considered to reflect embodiment, they can be criticized as well. For example, questionnaires can be subject to the tendency to please the experiment leader (demand effects), to a person's interpretation of the given phenomenon, and their experience during the experiment such as their sense of comfort or involvement (Schwind et al., 2019). Regarding proprioceptive drift, several RHI studies have found that it is not correlated with the senses of ownership and agency (Kalckert & Ehrsson, 2014; Normand et al., 2011; Riemer et al., 2015). This dissociation seems to indicate the involvement of different neural mechanisms associated with proprioceptive drift and embodiment (Toet et al., 2020), suggesting that proprioceptive drift might not be a robust measure of embodiment. A disadvantage of behavioral and physiological responses such as brain activity, HRD, SCR and EMG onset activity is that they require relatively advanced equipment and signal processing.

Besides the measures mentioned above, the embodiment literature suggests that the cross-modal congruency task (CCT) would be a relatively simple objective tool to quantify embodiment, which enables the collection of multiple repeated measures *during* an embodiment illusion, and is less susceptible to demand effects than other behavioral responses of embodiment (Aspell et al., 2009). The CCT, introduced by Spence, Pavani and Driver in 1998, has originally been designed to study the multisensory integration of visual and tactile cues (Spence et al., 2008). The task consists of indicating the location of vibrotactile targets while ignoring visual distractors as much as possible (Pavani et al., 2000). In the traditional configuration, four vibrators and four LEDs are arranged on the thumb and index finger of the participant's left and right hand. On each trial, a vibration and a light flash are presented to the participant's thumb or index finger. This can be *congruent*: the light flash is presented on the same hand as the vibration; or *incongruent*: the light flash is presented on the opposite hand of the vibration (Maselli & Slater, 2014; Walton & Spence, 2004). Participants have to respond to the vibrotactile stimuli as quickly as possible by indicating on which finger they perceived the vibrotactile stimulation, irrespective of the location of the distractor light. A large number of studies have consistently shown that responses to the vibrotactile targets are delayed and less accurate when the light flash is incongruent, rather than congruent, to the vibrotactile target. This effect is quantified in terms of the cross-modal congruency effect (CCE), defined as the difference in average response times between incongruent and congruent trials. CCEs have shown to be associated with reported changes in hand ownership (Pavani et al. 2000; Zopf et al., 2010), self-location (Maselli & Slater, 2014) and full-body ownership (Aspell et al., 2009; Maselli & Slater, 2014). The CCE has also been used to measure the level of virtual robotic tool incorporation (Sengül et al., 2012; Sengül et al., 2013a; Sengül et al., 2013b; Grespan et al., 2019). Accordingly, it has been suggested that the CCE provides an objective measure of multisensory integration in the body schema and the resulting feeling of embodiment (Aspell et al., 2009; Pavani et al., 2000; Zopf et al., 2010). Hence, we argue that embodiment would be reflected by a congruent light being helpful to quickly localize the tactile stimulation, whereas an incongruent light would impair this localization. If there is no embodiment, congruency of the stimuli would neither be very helpful nor would it impair the detection. This implies that the CCE would be larger if someone experiences a high level of embodiment and smaller if someone experiences a lower level of embodiment.

This study aims to validate the CCE as an objective measure of embodiment by measuring CCEs under four conditions with different, a priori expected levels of embodiment, and by comparing CCE scores with subjective questionnaire reports collected during the same four conditions. These conditions are (1) a real hand baseline condition (*real condition*) to replicate finding from previous CCE studies, (2) a real hand seen through a telepresence unit (*mediated condition*) to examine the effect of seeing the world through a telepresence unit, (3) a robotic hand seen through a telepresence unit (*robot condition*), (4) and a human-looking virtual hand seen through VR glasses (*VR condition*). The latter two conditions are especially relevant for applied teleoperation scenarios. A unimanual version of the CCT was used to examine embodiment strength in the four conditions. Previous research has also used a unimanual version of the CCT and found similar results to studies employing a classic CCT configuration (Zopf et al., 2010). The use of a unimanual version of the CCT is also justified by the fact that the CCE is typically larger for trials with visual and tactile cues on the same side (i.e. hand) with respect to the opposite side (Spence et al., 2004).

We expect that participants will experience a sense of embodiment in all conditions, as reflected by the CCE, but that the magnitude of the CCE will be affected by the degree of visual realism of the presented hand in each of the four conditions. Indeed, previous research has suggested that the visual realism of the fake body (part) might increase the relative strength of embodiment (Krom et al., 2019). In other words, embodiment might be stronger if the hand is realistic (i.e., human-looking) compared to a hand-like object with lower likeness such as a robotic hand (Toet et al., 2020). However, it is unclear whether this finding also holds when comparing embodiment strength for a robotic hand in reality and a human-looking hand in VR. Because we suppose that perceiving reality through a telepresence unit is akin to perceiving a world in VR, we tentatively expect that a human-looking virtual hand would induce a stronger sense of embodiment than a robotic hand in reality. This would result in the following magnitude of the CCE, from large to small: *Real, Mediated, Virtual, Robot.* Ultimately, a correlation between CCE magnitude and subjective questionnaire reports would validate the CCE as an objective measure of embodiment.

Method

Participants

Eight participants (5 male, 3 female; 7 right hand dominant; aged 23–44 years, mean ± SD = 28.9 ± 7.6 years; 2 wearing glasses) were recruited to take part in the experiment. Participants were informed about the general purpose of the research and received no compensation. All participants received oral and written instructions about the experimental procedures and gave their informed consent to participate in the study before the start of the experimental session. The study was approved by the TNO Institutional Review Board and was conducted in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki. Participants were invited using the following inclusion criteria: skin color similar to the rubber hand (Caucasian), age 18-50 years and a maximum hand length of 20 cm. Exclusion criteria were exceptional sensitivity to motion sickness and obvious properties by which the right arm or hand can be uniquely identified (e.g. prostheses, tattoos, scars).

Apparatus and materials

Vibrotactile stimuli were delivered on the thumb and index finger of the participant's right hand through a vibrotactile glove (Elitac, Utrecht, NL). On top of the glove, two LEDs (5 mm, red, diffuse) were attached to the thumb and index finger of the participant's right hand using Velcro. Responses were made by pressing two response buttons corresponding to 'thumb' and 'index finger' using the participant's left hand. A picture of the setup is shown in **Figure 1**. On each trial, a vibrotactile target stimulus (100 ms) and a visual distractor stimulus (100 ms) were delivered simultaneously at one out of the four possible locations (congruent thumb; congruent index finger; incongruent thumb; incongruent index finger). However, due to technical limitations, the duration of the light flash in the VR condition was a couple of milliseconds shorter (10-20 ms) than the programmed duration of 100 ms. While the onset was the same as of the vibrotactile target, the light flash stopped earlier than the vibration. Target and distractor stimuli were delivered in a pseudorandom order separated by a random interval between 2000-5000 ms, as depicted in **Figure 2**.

A telepresence unit was used to simulate mediated vision in a teleoperation setting. The telepresence unit consisted of a TNO-developed stereo camera that captured real-time video of the subject's hand (mediated condition) and the robotic hand (robot condition) and projected it onto a stereoscopic 3D head-mounted display (HMD, custom-built). The TNO-developed stereo camera followed the head movements of the participant (see **Figure 1**). It is worth noting that the custom-built HMD, contrary to the HTC VIVE, did not fully occlude the participant from the outside world.

The VR environment contained a 3D model of a human-looking arm and was calibrated to display the table (white surface) at the same height as the real table in front of the participant. The VR environment was modeled in a commercial game engine (Unity 3D, Unity Technologies, San Francisco, USA) using standard VR software (SteamVR, Valve Corporation, Bellevue, USA) and displayed in an HMD (HTC VIVE, HTC Corporation, New Taipei City, Taiwan).

The CCT was designed around an Arduino Mega 2560 Rev3 microcontroller board to achieve millisecond timing accuracy. A custom script on the Arduino Mega controlled the whole experiment. The Arduino Mega was connected to Robot Operating System (ROS). ROS was used as the 'central command hub', in which the experiment leader provided commands to run the CCT and which connected the Arduino Mega to Unity 3D. The Arduino Mega drove four physical distractor lights, two virtual distractor lights in Unity and two actuators of the vibrotactile glove through ROS. Two response buttons were interfaced with the Arduino Mega to measure subjects' responses to the vibrotactile targets.

Figure 1. Experimental setup (mediated condition). *Top right:* Robotic hand used in the robot condition. *Bottom right:* Human-looking virtual hand in the VR condition.

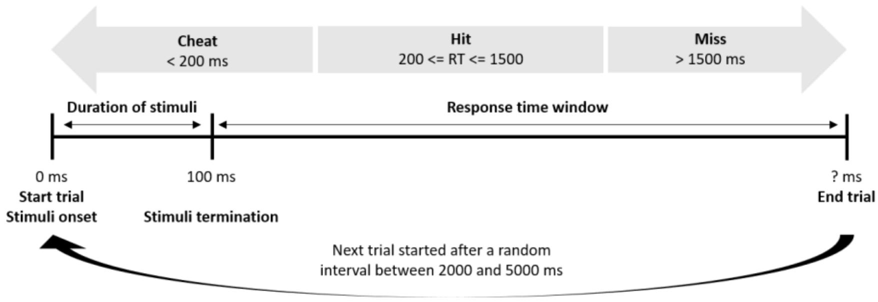

Figure 2. Trial timeline. Vibrotactile and visual stimuli were presented simultaneously for a duration of 100 ms. The trial ended when one of the two response buttons was pressed. The subsequent trial started after a random interval between 2000 and 5000 ms.

Experiment design

The design had three within-subjects factors: ***congruency*** of the location of the vibrotactile stimuli with respect to the visual distractors (congruent/incongruent), ***distractor location*** (thumb/index finger) and ***condition*** (real, mediated, robot and VR). The dependent variable was the cross-modal congruency effect in inverse efficiency (CCE-IE). The CCT was implemented in the following four conditions (see also **Figure 3**).

- ***Real*** *condition:* The visual distractors and vibrotactile targets were positioned on the participant's right hand, which the participant viewed directly. This condition served as a baseline condition and was included to replicate findings from previous CCE studies.
- ***Mediated*** *condition:* Identical to the real condition differing in that the participant saw his/her hand through a head-mounted display (HMD, custom-built) coupled to a TNO-developed stereo camera that followed the head movements of the participant. This condition was included to examine the effect of seeing the world through a telepresence unit (mediated vision).
- ***Robot*** *condition:* The visual distractors were mounted on a robotic hand that was attached to a KUKA LBR iiwa robot arm (KUKA, Augsburg, Germany). Just as in the mediated condition, participants saw this hand through an HMD (custom built) that was coupled to a stereo camera that followed the head movements of the participant. The robotic hand was displayed in the same position as the participant's real hand.
- ***VR*** *condition:* The visual distractors were shown on a human-looking hand modeled in a virtual reality environment. Participants saw this hand through an HMD (HTC VIVE, HTC Corporation, New Taipei City, Taiwan). The virtual hand was displayed in the same position as the participant's real hand.

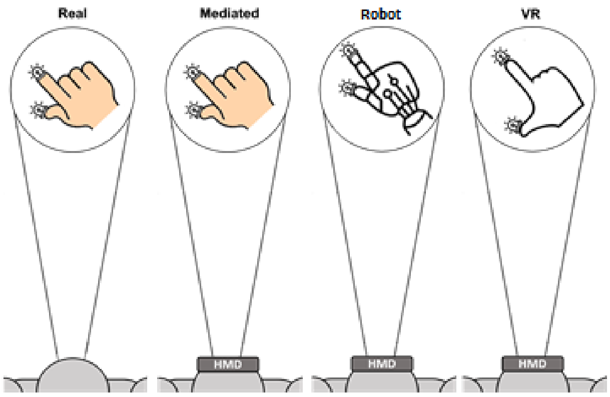

Figure 3. Four conditions in which the CCT was implemented. The conditions varied based on how the visual distractor lights were presented.

The four conditions were presented in a counterbalanced order across participants to compensate for possible learning effects (e.g., Blustein et al., 2019). The order was based on a balanced incomplete Latin square design (Ai et al., 2013). However, because the mediated and robot condition involved the same HMD (custom-built), these two conditions always followed one another to increase the flow of the experiment. In each of the four conditions, participants performed 100 CCT trials and completed an embodiment questionnaire immediately afterward. All participants performed all four conditions, each having an average duration of 10 minutes.

Procedure

During the whole duration of the experiment, participants sat comfortably on a chair, resting their forearms on foam sheets placed on a table. Participants were instructed to put on the vibrotactile glove on their right hand. Two LED strips were fastened tightly around the actuators of the vibrotactile glove. Participants were asked to lay their right and left arm on an indicated position on the table. In the mediated and robot conditions an HMD was put on, which was replaced by an HTC VIVE headset in the VR condition. Participants were then instructed to indicate as fast as possible the location of sequentially presented vibrotactile stimuli delivered on the thumb and index finger of their right hand by pressing one out of two response buttons (corresponding with 'thumb' or 'index finger') on which their left thumb and index finger rested. Participants were told that visual distractors would be presented simultaneously with the vibrotactile stimuli. They were instructed not to close their eyes and fixate on the visual distractors for the whole duration of the CCT run. Before the start of each CCT run, participants completed practice trials until they reached an accuracy level of 80%. After completion of 100 trials, the CCT was finished. Participants were instructed to take off the HMD or HTC VIVE headset and were asked to immediately fill in an embodiment questionnaire. Between each condition, participants had the opportunity to take a short break if necessary. The procedure was repeated for the other conditions. The total experiment lasted approximately 60 minutes per participant.

Questionnaire

Following each CCT condition, a 10-items questionnaire (reported in **Appendix A**) was administered in written form to assess the subjective level of embodiment. It contained questions relating to ownership, agency, and self-location, and control questions to rule out compliance, suggestibility, and possible placebo effects. Participants were asked to indicate the level of agreement or disagreement with the statements on a 7-point Likert scale which ranged from “strongly disagree” (-3) to “strongly agree” (+3). A response of 0 indicated that they ‘‘neither agreed nor disagreed”. An overall measure of ownership was computed by averaging across items 1–3 and 10, with items 2 and 3 being scored in reverse. An overall measure of agency was computed by averaging across items 4–6, with item 6 being scored negatively. An overall measure of self-location was computed by averaging across items 7–9, with item 8 being scored in reverse. Subsequently, these three measures were combined into a compound embodiment score reflecting the overall embodiment strength. The questionnaire was adapted from Longo & Haggard (2009) and Tsakiris et al. (2010) and was translated to the participants’ mother tongue (Dutch).

Data analysis and statistics

The CCT data were processed to extract the mean and median response time (RT) and error rate for each participant as a function of *congruency*, *distractor location* and *condition*. Outliers were identified via schematic boxplots. Given the debate on how to deal with RT outliers, the CCT data were initially analyzed using both mean and median RTs, and by using different cutoff methods, including absolute cutoffs and cutoffs based on three standard deviations above and below the mean. However, this led to essentially the same results. As demonstrated by Ratcliff (1993), using mean RTs and an absolute cutoff resulted in the most power and significance. Therefore, in order to enable comparison with other CCT studies, the analysis was continued with mean RTs. Trials with incorrect responses (144) and with RT smaller than 200 ms and larger than 1500 ms (18) were discarded (following the method of Maselli & Slater, 2014). This led to a rejection of a mean of 5.1% of all trials. The inverse efficiency (IE) score (Townsend & Ashby, 1983) was then calculated by dividing the mean RT by the percentage of correct responses for each condition, thereby accounting for possible speed-accuracy trade-offs in the RT data. The IE has been extensively used in previous studies that used the CCT (e.g., Marini et al., 2014; Maselli & Slater, 2014; Sengül et al., 2012; Spence et al., 2004).

Data from all resulting trials were first analyzed using a three-way repeated measures ANOVAs on the mean RTs. The three factors were: *congruency* (congruent/incongruent), *distractor location* (thumb/index finger) and c*ondition* (real, mediated, robot, and VR). Then, two one-way repeated measures ANOVAs were conducted to compare both congruent and incongruent RTs across conditions. Next, error rates were analyzed using the Friedman and Wilcoxon signed-rank test. Subsequently, a one-way repeated measures ANOVA on CCE-IE scores was performed with the factor *condition* to test the hypothesized difference in embodiment strength between conditions. Additional analyses were conducted to control for condition order effects as well as learning and fatigue effects.

Questionnaire data were processed to extract a compound embodiment score as well as mean responses per questionnaire statement and subcomponent of embodiment. The compound embodiment score was normalized using min-max normalization to allow a comparison between each of the four conditions. A one-way repeated measures ANOVA with the factor *condition* was conducted to examine differences in the compound embodiment score across conditions. Subsequent post-hoc tests were conducted to examine differences between conditions.

Finally, correlation analyses were conducted for each condition between the magnitude of participants’ CCE-IE scores and the compound embodiment scores as obtained from the questionnaire. Post-hoc tests for all ANOVAs were applied where appropriate. All statistical tests were performed at a significance level of alpha = 0.05 and were performed using IBM SPSS Statistics 26.

Results

Requirements for normality of residuals were checked with the Shapiro-Wilk test and reported with its *p*-value (p_{sw}). Mauchly's test indicated no violation of the sphericity assumption for all ANOVAs performed ($p > 0.05$). We avoided the debated practice of correcting for multiple comparisons (Perneger, 1998), and instead reported effect sizes together with the true *p*-value for each performed test.

Results of the CCT

The mean RTs, error rates, and CCE and CCE-IE scores are shown in **Table 2** and **Figure 6**. The residuals of mean RTs were normally distributed, except for one out of sixteen cases ($p_{sw} = 0.039$). To compensate for this small departure from normality a more conservative *p*-value of 0.01 was used. Analysis of mean RTs by using a three-way repeated measures ANOVA revealed a significant main effect of *congruency* $F(1, 7) = 62.6$, $p < 0.0001$, partial $\eta^2 = 0.90$, caused by faster responses when stimuli were congruent ($M = 489$ ms) versus incongruent ($M = 639$ ms). Furthermore, it revealed a significant main effect of *condition*, $F(3, 21) = 5.8$, $p = 0.005$, partial $\eta^2 = 0.45$, and a significant interaction between *congruency* and condition, $F(3, 21) = 8.7$, $p = 0.001$, partial $\eta^2 = 0.55$. Subsequent post-hoc comparisons of the conditions demonstrated a significant difference between the real and mediated conditions (529.3 ± 30.1 vs. 597.6 ± 48.6, $p = 0.017$), the real and VR conditions (529.3 ± 30.1 vs. 595.8 ± 31.7, $p < 0.0001$) and the robot and VR conditions (536.8 ± 39.4 vs. 595.8 ± 31.7, $p = 0.029$). A close to significant difference was observed between the mediated and robot conditions (597.6 ± 48.6 vs. 536.8 ± 39.4, $p = 0.067$). Importantly, none of the effects involving the factor *distractor location* were significant, indicating that visuotactile interactions were comparable when participants responded with either their thumb or index finger.

Two one-way repeated measures ANOVAs showed a significant difference between congruent trials, $F(3, 21) = 3.823$, $p = 0.025$, partial $\eta^2 = 0.35$, as well as incongruent trials, $F(3, 21) = 7.304$, $p = 0.002$, partial $\eta^2 = 0.51$, across all conditions. Post-hoc comparisons of congruent trials demonstrated a significant difference between the real and mediated conditions (464.6 ± 30.1 vs. 523.4 ± 37.7, $p = 0.013$) and the real and VR conditions (464.6 ± 30.1 vs. 494.4 ± 28.0, $p = 0.004$). Post-hoc comparisons of the incongruent trials revealed a significant difference between the real and mediated (593.9 ± 31.5 vs. 671.8 ± 61.6, $p = 0.044$), real and VR (593.9 ± 31.5 ± vs. 697.2 ± 37.5, $p < 0.0001$) and robot and VR (591.8 ± 48.2 vs. 697.2 ± 37.5, $p = 0.007$) conditions. A close to significant difference was observed between the mediated and robot conditions (671.8 ± 61.6 ± vs. 591.8 ± 48.2, $p = 0.057$).

Residuals of error rates of incongruent trials and error rates in the mediated condition were not normally distributed ($p_{sw} = 0.02$ and $p_{sw} = 0.001$, respectively). A Friedman's test showed a trend in error rates between conditions, $\chi^2(3) = 7.581$, $p = 0.056$, with a higher error rate in the VR condition (6.6%) compared to the real (3%), mediated (4.5%) and robot (3.9%) conditions. Moreover, a Wilcoxon signed-rank test revealed a significant difference in error rates between congruent and incongruent trials across all conditions ($Z = -2.197$, $p = 0.028$), caused by a higher error rate in incongruent trials ($M = 7.8\%$) compared to congruent trials ($M = 1.2\%$).

The CCE-IE scores (shown in **Figure 7**) were analyzed using a one-way repeated measures ANOVA. This revealed a significant main effect of condition, $F(3, 21) = 8.7$, $p = 0.00022$, partial $\eta^2 = 0.60$. Subsequent post-hoc tests showed a significant difference between the real and VR conditions (137.5 ± 13.9 vs. 219.5 ± 19.3, $p < 0.00001$), mediated and VR conditions (159.5 ± 31.1 vs. 219.5 ± 19.3, *p* = 0.041) and robot and VR conditions (118.2 ± 19.4 vs. 219.5 ± 19.3, *p* = 0.00036). Residuals of mean CCE-IE scores were normally distributed ($p_{sw} > 0.36$).

Table 2. Mean response times, error rates, cross-modal congruency effects and cross-modal congruency effects in inverse efficiency, as a function of condition and the distractor's congruence with the target.

Condition	Target-distractor congruence	Response time (ms)[a]	Error rate (%)	CCE (ms)	CCE-IE (ms)
Real	Congruent	464 (7)	0.1	129	137
	Incongruent	596 (8)	2.9		
Mediated	Congruent	522 (8)	0.6	148	159
	Incongruent	674 (13)	3.9		
Robot	Congruent	481 (7)	0.6	110	118
	Incongruent	593 (10)	3.3		
VR	Congruent	490 (7)	1.0	203	220
	Incongruent	700 (10)	5.6		

[a] Standard errors are given in parentheses.

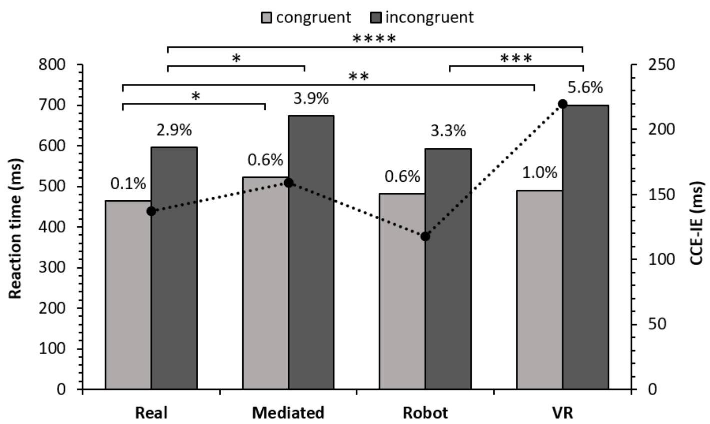

Figure 6. Mean RTs of congruent and incongruent trials per condition. The cross-modal congruency effect in inverse efficiency (CCE-IE) is shown in black dots. The error rate is given as a percentage above each bar in the graph. The graph shows that a robust congruency effect is present in all four conditions ($p < 0.0001$). Sig: ∗0.01 < *p* < .05, ∗∗0.001 < *p* < 0.01, ***0.001 < *p* < 0.01, ****0.00001 < *p* < 0.0001.

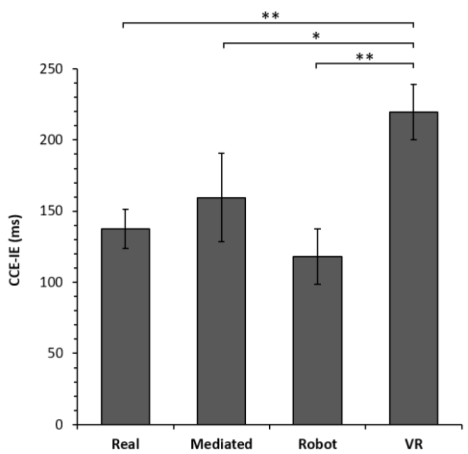

Figure 7. Means and standard errors of the cross-modal congruency effect in inverse efficiency (CCE-IE) as a function of condition. Sig: ∗0.01 < p < 0.05, ∗∗0.0001 < p < 0.001.

Additional analyses were conducted to control for possible order effects in the condition blocks as well as learning and fatigue effects in the RT data. A one-way repeated measures ANOVA with the factor *block number* (referring to the condition blocks, 1 through 4) revealed no significant effect. **Figure 8** shows the difference in mean CCE-IE scores as a function of the order in which the conditions were presented for each subject.

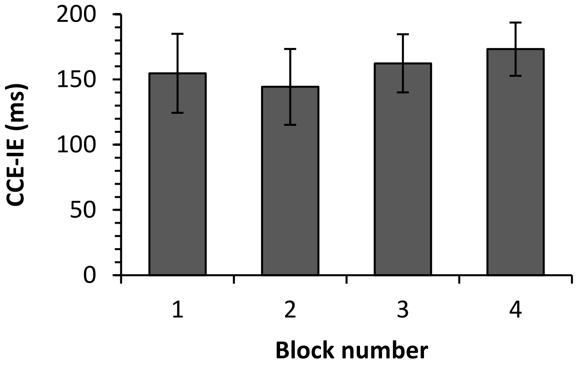

Figure 8. Means and standard errors of the cross-modal congruency effect in inverse efficiency (CCE-IE) as a function of condition block number.

RT data of all CCT trials were also checked for possible learning and fatigue effects utilizing scatterplots and Pearson *r* correlations. A significant negative correlation was found for the real condition ($r = -0.079$, $p = 0.028$), suggesting a very small learning effect. When examining correlations between participants, 6 out of 32 correlations were found to be significant (see **Table 3**), of which four suggest a weak learning effect and two suggest a weak fatigue effect. Thus, the effects of time were weak and non-systematic.

Table 3. Correlations between RT and trial number for each participant and condition.

	1	2	3	4	5	6	7	8
Real	-0.15	0.14	-0.08	0.03	-0.06	-0.08	**-,327****	**-,261****
Mediated	0.19	0.08	0.14	0.13	-0.08	0.13	-0.19	0.09
Robot	0.14	0.15	-0.03	-0.16	**,376****	-0.11	**-,346****	**-,230***
VR	-0.02	-0.05	0.10	**,315****	-0.16	-0.11	0.10	0.04

Note: **Sig. at 0.01 level (two-tailed), *Sig. at the 0.05 level (two-tailed).

Results of the questionnaire

Mean ratings and standard deviations for the 10-items questionnaire are reported in **Table 1**. The compound embodiment scores for each condition are shown in **Figure 4**. This score reflects the sum of the ownership, agency, and self-location measures as assessed from the questionnaire. The mean questionnaire ratings for each of the subcomponents of embodiment are reported in **Figure 5**. No evidence was found for non-normal distributions in the overall measure of embodiment as obtained from the questionnaire for each of the four conditions (Shapiro–Wilk test, all $p > 0.09$). Results from the one-way repeated measures ANOVA determined that *embodiment score* differed significantly between *condition*s, $F(3, 21) = 31.4$, $p < 0.0001$, partial $\eta^2 = 0.82$. Subsequent post-hoc tests revealed that the *embodiment score* in the real condition differed significantly from the mediated condition (7.0 ± 1.0 vs. 4.0 ± 2.5, p = 0.038), the robot condition (7.0 ± 1.0 vs. -1.6 ± 2.4, p = 0.00016) and the VR condition (7.0 ± 1.0 vs. 1.3 ± 2.7, p = 0.003). Also, the mediated condition differed significantly from the robot condition (4.0 ± 2.5 vs. -1.6 ± 2.4, p = 0.002) and the robot condition differed significantly from the VR condition (-1.6 ± 2.4 vs. 1.3 ± 2.7, p = 0.018). A close to significant difference was observed between the mediated and VR conditions (p = 0.05).

Table 1. Mean rating scores with standard deviations of the 10-items embodiment questionnaire. Mean rating scores refer to participants' agreement with each statement on a 7-point scale, ranging from "strongly disagree" (-3) to "strongly agree" (+3).

	"During the experiment block there were times when..."	**Real**	**Mediated**	**Robot**	**VR**
1	"...it felt like the hand I was looking at was my own hand."	3.0 (0.00)	1.75 (1.04)	-1.63 (1.19)	0.75 (1.16)
2	"...it felt like the hand I was looking at wasn't mine."	-2.88 (0.35)	-2.00 (1.07)	2.00 (1.41)	0.00 (2.00)
3	"...it felt like the hand I was looking at was somebody else's hand."	-2.75 (0.46)	-1.88 (1.81)	0.13 (2.36)	-0.75 (1.49)
4	"...it felt like I was in control of the hand I was looking at."	2.38 (0.52)	1.88 (1.25)	-0.13 (1.64)	1.00 (1.41)
5	"...it felt like I could move the hand I was looking at if I wanted."	2.75 (0.46)	2.13 (0.99)	-0.63 (1.85)	0.38 (1.60)
6	"...it felt like the hand I was looking at was out of my control."	-2.28 (1.06)	-2.00 (0.93)	0.38 (1.69)	-0.38 (1.51)
7	"...it felt like my hand was somewhere between the table and the location where I saw the hand."	-0.88 (2.95)	1.13 (1.46)	-0.13 (2.03)	-0.75 (2.05)
8	"...it felt like I could not really tell where my hand was."	-2.75 (0.46)	-0.88 (1.55)	-1.25 (1.75)	-0.50 (2.00)
9	"...it felt like my hand was in the location where I saw the hand."	3.0 (0.00)	0.50 (2.07)	-0.63 (1.06)	1.25 (1.83)
10	"...it felt like I was looking directly at my hand rather than at an image of the hand."	2.88 (0.35)	-1.13 (1.36)	-1.75 (1.16)	0.13 (1.55)

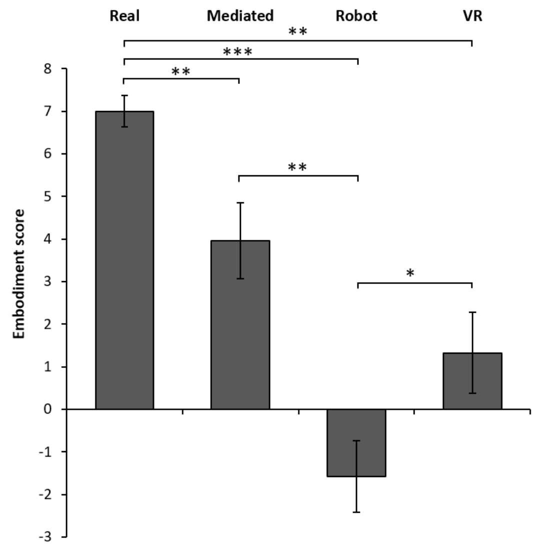

Figure 4. Compound embodiment scores and standard errors from questionnaire ratings a function of condition. Sig: ∗0.01 < p < .05, ∗∗0.0001 < p < 0.001, ***0.00001 < p < 0.0001.

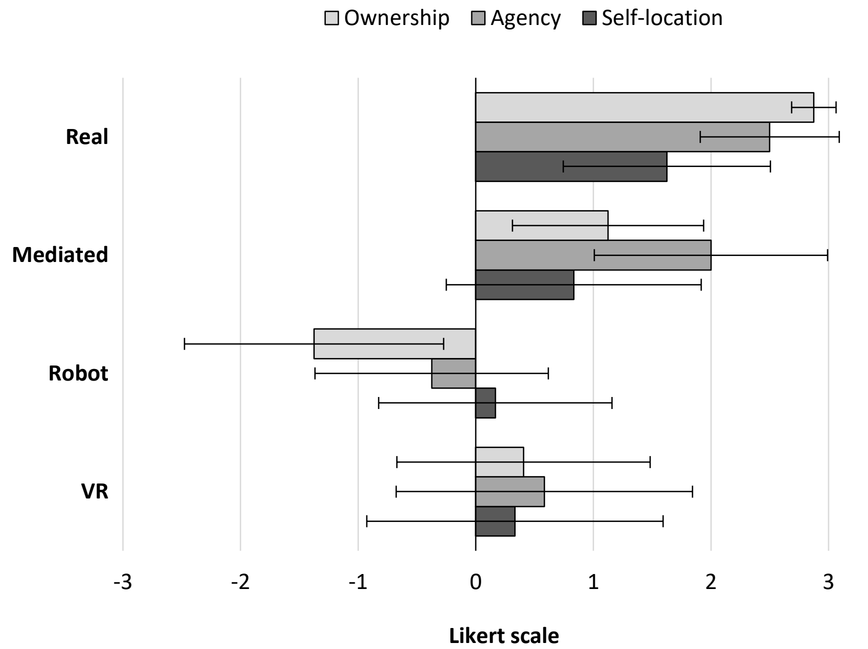

Figure 5. Mean questionnaire ratings for embodiment subcomponents with standard deviations.

Correlation analysis

For each condition, the correlation between the CCE-IE and the compound embodiment scores as assessed from the questionnaire was tested. This compound embodiment score was normalized using min-max normalization to allow a comparison between each of the four conditions. The Pearson's r showed no significant positive correlation between the two measures (p > 0.6), also shown in the scatterplots in **Figure 9**. The scatterplot in **Figure 10** is included to get an impression of the correlation between the CCE-IE and embodiment score across all conditions. The data points in the scatterplots represent subjects, which are color-coded to allow a comparison between conditions. Interestingly, the order in the magnitude of the CCE-IE tends to be similar across conditions. In other words, when subject X has a low CCE-IE score in the real condition, (s)he appears to have a low CCE-IE score in the other conditions. Furthermore, it is worth noting that the spread of both the CCE-IE scores and the compound embodiment scores in the real condition seems much smaller compared to the other conditions. No significant positive correlations between the CCE-IE and individual questionnaire statements as well as embodiment subcomponents were found either.

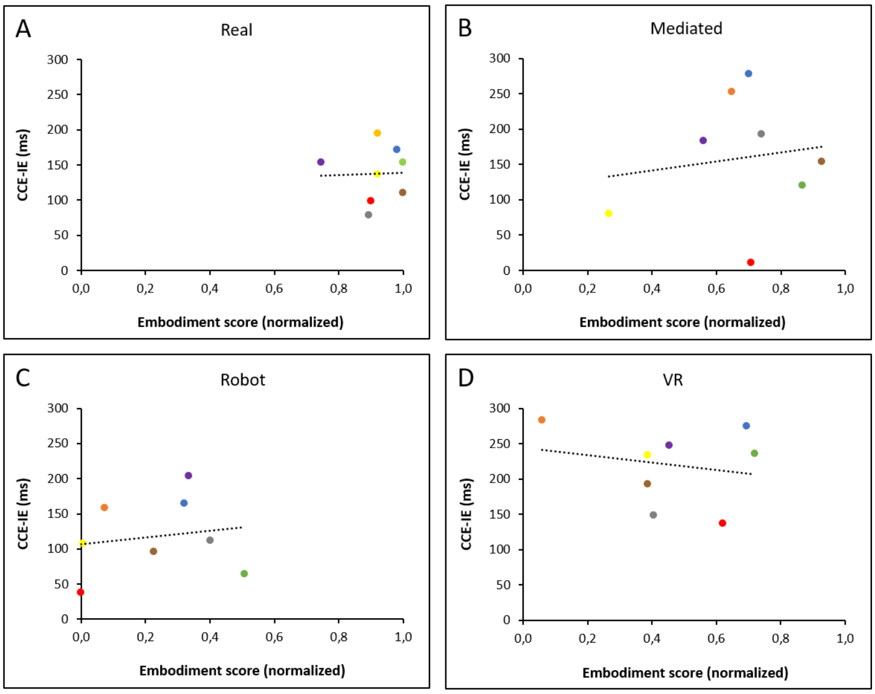

Figure 9. CCE-IE scores plotted against the embodiment scores as obtained from the questionnaire for each condition. Embodiment scores were normalized to enable a comparison between conditions. Subjects were color-coded.

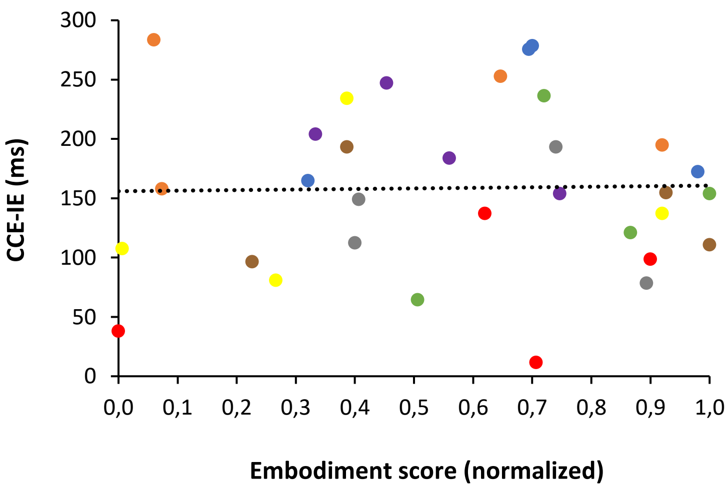

Figure 10. CCE-IE scores plotted against the embodiment scores as obtained from the questionnaire for all conditions. Subjects were color-coded.

Discussion

This study aimed to validate the CCE as an objective measure of embodiment by evaluating CCEs in four conditions and comparing CCE scores with subjective questionnaire reports. The study has two main outcomes. First, the magnitude of the CCE does not appear to be affected by the degree of visual realism of the hand. In contrast to our hypothesis, the magnitude of the CCE from strong to weak was: Virtual, Mediated, Real, Robot. This is striking because one would expect that the strongest multisensory integration or sense of embodiment would be demonstrated when visual distractors are presented on the participant's hand (real condition) instead of on a virtual hand (VR condition). Second, no significant positive correlation between the CCE and subjective questionnaire reports was found, suggesting that both measures do not measure the same phenomenon (in this case: embodiment strength). In the next paragraphs we provide possible explanations for both outcomes.

The results demonstrated a robust CCE (i.e., a significant difference between congruent and incongruent RTs), indicating a multisensory integration of visual and tactile events in all conditions. According to our hypothesis, this would imply that participants considered the hand presented in each of the four conditions to some extent as their own. However, we were not able to determine that the degree of multisensory integration as reflected by the CCE was related to the experienced strength of embodiment. Although the questionnaire revealed a clear order in the magnitude of embodiment strength across the conditions, which was in line with our hypothesis, the CCE did not demonstrate this expected order. A likely explanation for this finding could be that other factors than visual realism and whether or not the hand was viewed directly, affected the magnitude of the CCE, which were different across conditions. We will elaborate on the most likely factors by taking the unexpected CCE difference between the real and VR conditions as examples.

A first factor could be the *degree of spatial uncertainty* about the vibrotactile target location. In the VR condition, participants received no visual information about the location of their real hand at all and had to rely on their proprioception, which would result in a relatively large spatial uncertainty about the vibrotactile target location as compared to the real condition. On the one hand, this would explain the relatively fast RTs and low error rates of congruent (*M* = 464 ms, ER = 0.1%) and incongruent trials (*M* = 596 ms, ER = 2.9%) in the real condition, despite conflicting visual and tactile input during incongruent trials. On the other hand, the high spatial uncertainty about the target location, together with a conflict of visual and tactile input, would therefore explain the relatively high incongruent RT and error rate in the VR condition (*M* = 700 ms, ER = 5.6%). Interestingly, despite high spatial uncertainty about the target location, the congruent RT in the VR condition was relatively fast (*M* = 490 ms, ER = 1.0%), most likely because the light flash was of much help in determining the spatial location of the vibrotactile target. A study by Soto-Faraco et al. (2004) has shown that focusing attention on a particular location reduces, or even eliminates, distractor interference effects from task-irrelevant stimuli at other locations. However, whether this degree of spatial uncertainty indeed affects the magnitude of the CCE could be further investigated by assessing if the magnitude of the CCE would be reduced if participants focus their tactile endogenous spatial attention on a particular hand or finger slightly in advance of a trial (for example, see Spence et al., 2004).

A second factor could be the *relative timing* of the target and distractor stimuli in this condition. Due to technical limitations, the duration of the light flash in the VR condition was a couple of milliseconds shorter (10-20 ms) than the programmed duration of 100 ms. While the onset was the same as of the vibrotactile target, the light flash stopped earlier than the vibration, making it feel like the onset of the visual distractor slightly preceded that of the vibrotactile targets. Indeed, Spence et al. (2004) have demonstrated that subjects found it harder to ignore irrelevant visual distractors (i.e., they made significantly more errors) when their onset slightly preceded that of the vibrotactile targets than when the two stimuli were presented simultaneously or when the onset of the target preceded that of the distractor by 30 ms. This finding would also explain the relatively high error rate in the VR condition compared to the other conditions.

A third possible factor could be the *spatial separation* between the vibrotactile target and the visual distractor. Participants reported that the visual distractors in the VR condition seemed visually closer together and that the hand was further away compared to the real condition. In this light, Spence et al. (2004) have demonstrated that the magnitude of the CCE is increased as the spatial separation between the vibrotactile target and the visual distractor stimuli decreases, even when both the target and the distractor are presented from within the same hemifield. However, it is unclear whether this finding applies to the small difference between the target and distractor stimuli between the real and VR conditions in the present study.

Lastly, the relatively high *cognitive workload* during the CCT in the VR condition could explain the unexpected CCE difference between the real and VR condition. Some participants reported that it required more effort to concentrate during the CCT in the VR condition compared to the other conditions. This is supported by the work of Škola and Liarokapsis (2016), who found that inducing the RHI in VR produced more brain activity for gamma and beta waves than in a real-world environment. Beta and gamma waves are generally associated with attention, which might indicate that participants had to make more effort to remain concentrated in the VR environment. A reason for this could be that the pixelated screen of the VR headset caused eye strain which made it more difficult to maintain focused on the distractor lights, hence reducing the participant's capacity to ignore irrelevant light flashes. To summarize, although the VR environment resembled the experimental setup in the real condition, it introduced several issues that could have caused the relatively high CCE in this condition. It should be noted that one should be careful when comparing embodiment in real and VR environments.

We neither found evidence to support the expected positive correlation between the CCE and subjective questionnaire reports when this was examined within conditions. Regarding the subjective questionnaire, the subject's responses to the questionnaire items probably were biased to some extent, as some participants were aware of the purpose of the study. This implies that participants might have indicated a strong sense of embodiment in one of the conditions, while they did not experience this feeling that strongly and vice versa. Nevertheless, it remains unclear what the subjective questionnaire reports really measure. Therefore, it would be interesting to support the CCT with another objective measure of embodiment, such as physiological measurements.

The results have an important implication regarding the connection between multisensory integration and embodiment. In line with previous work, the results in the present study challenge the notion that multisensory visuotactile integration and embodiment strength are closely connected. Kanayama et al. (in press) investigated multisensory integration during RHI induction in a real and VR environment using the CCE and found that activity in brain areas related to multisensory integration diminished in VR compared to the real environment, while the CCE difference between the real and VR environments was not significant (107 at Real vs. 149 at VR). They suggested that VR, including an HMD, can alter our visuotactile integration process. Additionally, Marini et al. (2017) used a version of the CCE in which unimodal tactile trials were intermixed with crossmodal visuotactile trials and found no difference in RTs between unimodal tactile trials and congruent visuotactile trials. This implies that distractor lights merely slow down responses when delivered opposite to the vibrotactile target and do not speed up responses when delivered on the same location as the vibrotactile target. Hence, it can be concluded that multisensory visuotactile integration may contribute to only a small component of the CCE and therefore the CCE is likely to primarily reflect response conflict, together with other factors (Marini et al., 2017; Spence et al., 2004).

To confirm this notion, it would be of interest to conduct a follow-up study in which the CCE is investigated in one condition only, thereby keeping the environment constant while varying a range of factors that could enhance or diminish the realism of the environment. Examples of such factors include visuotactile synchronicity, stimulus onset delay (SOA), duration of the stimuli, and visual perspective. A corresponding question would then be whether the CCE increases as the visual realism increases. In this case, virtual reality would be the most suitable condition as it enables researchers to easily manipulate these factors (Kilteni et al., 2012). It also offers the opportunity to examine the effect of visual realism of the hand, by comparing a virtual human-looking hand with a virtual robotic hand (see for example Krom et al., 2019). However, even then we do not believe the CCE would be a robust measure of embodiment. Therefore, it is recommended that future studies focus on other behavioral and physiological measures to quantify embodiment.

In conclusion, this is the first study that has conducted a systematic comparison of embodiment strength under different conditions of visual realism by using an objective measure of embodiment. We found no unambiguous evidence that the magnitude of the CCE was affected by the degree of visual realism of the hand. We have proposed several factors that may explain this outcome: the degree of spatial uncertainty about the vibrotactile target location, the relative timing of the target and distractor stimuli, the spatial separation between the vibrotactile target and the visual distractor, and the high cognitive workload during the CCT in the VR condition. We also found no evidence to support the hypothesis that the CCE and embodiment scores as assessed by the subjective questionnaire reports are correlated. Based on these findings, it can be concluded that the CCE may not be a robust measure to quantify embodiment.

Acknowledgments

I thank Martijn van den Berg and Ioana Cocu for their help with developing the experimental setup and thank Kaj Gijsbertse for his assistance during the experiment. I am especially grateful to Irene Kuling, Anne-Marie Brouwer and Krista Overvliet for their feedback on early versions of this thesis.

References

Ai, M., Li, K., Liu, S., & Lin, D. K. (2013). Balanced incomplete Latin square designs. *Journal of Statistical Planning and Inference, 143*(9), 1575–1582. https://doi.org/10.1016/j.jspi.2013.05.001

Almeida, L., Menezes, P., & Dias, J. (2017). Improving robot teleoperation experience via immersive interfaces. *Proceedings of 2017 4th Experiment at International Conference: Online Experimentation (Exp.at'17)*, 87–92. https://doi.org/10.1109/EXPAT.2017.7984414

Armel, K. C., & Ramachandran, V. S. (2003). Projecting sensations to external objects: Evidence from skin conductance response. *Proceedings of the Royal Society B, 270*(1523), 1499–1506. https://doi.org/10.1098/rspb.2003.2364

Arzy, S., Thut, G., Mohr, C., Michel, C. M., & Blanke, O. (2006). Neural basis of embodiment: distinct contributions of temporoparietal junction and extrastriate body area. *Journal of Neuroscience, 26*(31), 8074-8081. https://doi.org/10.1523/JNEUROSCI.0745-06.2006

Aspell, J. E., Lenggenhager, B., & Blanke, O. (2009). Keeping in touch with one's self: Multisensory mechanisms of self-consciousness. *PLoS ONE*, *4*(8), 25–27. https://doi.org/10.1371/journal.pone.0006488

Aymerich-Franch, L., Petit, D., Ganesh, G., & Kheddar, A. (2017a). Non-human looking robot arms induce illusion of embodiment. *International Journal of Social Robotics*, *9*(4), 479–490. https://doi.org/10.1007/s12369-017-0397-8

Aymerich-Franch, L., Petit, D., Ganesh, G., & Kheddar, A. (2017b). Object touch by a humanoid robot avatar induces haptic sensation in the real hand. *Journal of Computer-Mediated Communication*, *22*(4), 215–230. https://doi.org/10.1111/jcc4.12188

Blustein, D., Gill, S., Wilson, A., & Sensinger, J. (2019). Crossmodal congruency effect scores decrease with repeat test exposure. PeerJ *7:e6976*, 1–16. https://doi.org/10.7717/peerj.6976

Botvinick, M., & Cohen, J. (1998). Rubber hands 'feel' touch that eyes see. *Nature, 391*, 756. https://doi.org/10.1038/35784

De Vignemont, F. (2011). Embodiment, ownership and disownership. *Consciousness and Cognition*, *20*(1), 82–93. https://doi.org/10.1016/j.concog.2010.09.004

Ehrsson, H. H., Spence, C., & Passingham, R. E. (2004). That's my hand! Activity in premotor cortex reflects feeling of ownership of a limb. *Science*, *305*(5685), 875–877. https://doi.org/10.1126/science.1097011

Ehrsson, H. H., Wiech, K., Weiskopf, N., Dolan, R. J., & Passingham, R. E. (2007). Threatening a rubber hand that you feel is yours elicits a cortical anxiety response. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(23), 9828–9833. https://doi.org/10.1073/pnas.0610011104

González-Franco, M., Peck, T. C., Rodríguez-Fornells, A., & Slater, M. (2014). A threat to a virtual hand elicits motor cortex activation. *Experimental Brain Research, 232*(3), 875–887. https://doi.org/10.1007/s00221-013-3800-1

Grespan, L., Fiorini, P., & Colucci, G. (2019). Operator interfaces for safe robotic surgery. *Springer Tracts in Advanced Robotics*, *126*, 125–142. https://doi.org/10.1007/978-3-030-03020-9_10

Hedayati, H., Walker, M., & Szafir, D. (2018). Improving collocated robot teleoperation with augmented reality. *ACM/IEEE International Conference on Human-Robot Interaction*, 78–86. https://doi.org/10.1145/3171221.3171251

Kalckert, A., & Ehrsson, H. H. (2014). The moving rubber hand illusion revisited: Comparing movements and visuotactile stimulation to induce illusory ownership. *Consciousness and Cognition*, *26*(1), 117–132. https://doi.org/10.1016/j.concog.2014.02.003

Kanayama, N., Hara, M., & Kimura, K. (in press). Disruption by virtual reality of the cortical oscillations related to visuotactile integration during an embodiment process. *bioRxiv*, 1–27. https://doi.org/10.1101/2020.03.27.010942

Kilteni, K., Groten, R., & Slater, M. (2012). The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments*, *21*(4), 373–387. https://doi.org/10.1162/PRES_a_00124

Krom, B. N., Catoire, M., Toet, A., Van Dijk, R. J. E., & Van Erp, J. B. F. (2019). Effects of Likeness and Synchronicity on the Ownership Illusion over a Moving Virtual Robotic Arm and Hand. *2019 IEEE World Haptics Conference, WHC 2019*, 49–54. https://doi.org/10.1109/WHC.2019.8816112

Longo, M. R., Schüür, F., Kammers, M. P. M., Tsakiris, M., & Haggard, P. (2008). What is embodiment? A psychometric approach. *Cognition, 107*(3), 978–998. https://doi.org/10.1016/j.cognition.2007.12.004

Longo, M. R., & Haggard, P. (2009). Sense of agency primes manual motor responses. *Perception, 38*(1), 69–78. https://doi.org/10.1068/p6045

Ma, K., & Hommel, B. (2013). The virtual-hand illusion: Effects of impact and threat on perceived ownership and affective resonance. *Frontiers in Psychology*, *4*, 1–6. https://doi.org/10.3389/fpsyg.2013.00604

Marini, F., Tagliabue, C. F., Sposito, A. V., Hernandez-Arieta, A., Brugger, P., Estévez, N., & Maravita, A. (2014). Crossmodal representation of a functional robotic hand arises after extensive training in healthy participants. *Neuropsychologia*, *53*(1), 178–186. https://doi.org/10.1016/j.neuropsychologia.2013.11.017

Marini, F., Romano, D., & Maravita, A. (2017). The contribution of response conflict, multisensory integration, and body-mediated attention to the crossmodal congruency effect. *Experimental Brain Research, 235*(3), 873–887. https://doi.org/10.1007/s00221-016-4849-4

Maselli, A., & Slater, M. (2014). Sliding perspectives: Dissociating ownership from self-location during full body illusions in virtual reality. *Frontiers in Human Neuroscience*, *8*, 1–19. https://doi.org/10.3389/fnhum.2014.00693

Newport, R., Pearce, R., & Preston, C. (2010). Fake hands in action: Embodiment and control of supernumerary limbs. *Experimental Brain Research, 204*(3), 385–395. https://doi.org/10.1007/s00221-009-2104-y

Normand, J. M., Giannopoulos, E., Spanlang, B., & Slater, M. (2011). Multisensory stimulation can induce an illusion of larger belly size in immersive virtual reality. *PLoS ONE*, *6*(1). https://doi.org/10.1371/journal.pone.0016128

Pavani, F., Spence, C., & Driver, J. (2000). Visual capture of touch: Out-of-the-body experiences with rubber gloves. *Psychological Science*, *11*(5), 353–359. https://doi.org/10.1111/1467-9280.00270

Perneger, T. V. (1998). What's wrong with Bonferroni adjustments. *BMJ, 316*, 1236–1238. https://doi.org/10.1136/bmj.316.7139.1236

Pritchard, S. C., Zopf, R., Polito, V., Kaplan, D. M., & Williams, M. A. (2016). Non-hierarchical influence of visual form, touch, and position cues on embodiment, agency, and presence in virtual reality. *Frontiers in Psychology*, *7*, 1–14. https://doi.org/10.3389/fpsyg.2016.01649

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological bulletin, 114*(3), 510. https://doi.org/10.1037/0033-2909.114.3.510

Riemer, M., Bublatzky, F., Trojan, J., & Alpers, G. W. (2015). Defensive activation during the rubber hand illusion: Ownership versus proprioceptive drift. *Biological Psychology*, *109*, 86–92. https://doi.org/10.1016/j.biopsycho.2015.04.011

Romano, D., Caffa, E., Hernandez-Arieta, A., Brugger, P., & Maravita, A. (2015). The robot hand illusion: Inducing proprioceptive drift through visuo-motor congruency. *Neuropsychologia, 70*, 414–420. https://doi.org/10.1016/j.neuropsychologia.2014.10.033

Sanchez-Vives, M. V., & Slater, M. (2005). From presence to consciousness through virtual reality. *Nature Reviews Neuroscience, 6*(4), 332–339. https://doi.org/10.1038/nrn1651

Schwind, V., Knierim, P., Haas, N., & Henze, N. (2019). Using presence questionnaires in virtual reality. *Conference on Human Factors in Computing Systems - Proceedings*, 1–16. https://doi.org/10.1145/3290605.3300590

Sengül, A., van Elk, M., Rognini, G., Aspell, J. E., Bleuler, H., & Blanke, O. (2012). Extending the body to virtual tools using a robotic surgical interface: Evidence from the crossmodal congruency task. *PLoS ONE*, *7*(12). https://doi.org/10.1371/journal.pone.0049473

Sengül, A., Rognini, G., Van Elk, M., Aspell, J. E., Bleuler, H., & Blanke, O. (2013a). Force feedback facilitates multisensory integration during robotic tool use. *Experimental Brain Research*, *227*(4), 497–507. https://doi.org/10.1007/s00221-013-3526-0

Sengül, A., Rivest, F., Van Elk, M., Blanke, O., & Bleuler, H. (2013b). Visual and force feedback time-delays change telepresence: Quantitative evidence from crossmodal congruency task. World Haptics Conference, WHC 2013, 577–582. https://doi.org/10.1109/WHC.2013.6548472

Škola, F., & Liarokapis, F. (2016). Examining the effect of body ownership in immersive virtual and augmented reality environments. *The Visual Computer, 32*(6-8), 761–770. https://doi.org/10.1007/s00371-016-1246-8

Slater, M., Perez-Marcos, D., Ehrsson, H. H., & Sanchez-Vives, M. V. (2008). Towards a digital body: The virtual arm illusion. *Frontiers in Human Neuroscience*, *2*, 1–8. https://doi.org/10.3389/neuro.09.006.2008

Slater, M., Spanlang, B., Sanchez-Vives, M. V., & Blanke, O. (2010). First person experience of body transfer in virtual reality. *PLoS ONE*, *5*(5), 1–9. https://doi.org/10.1371/journal.pone.0010564

Soto-Faraco, S., Ronald, A., & Spence, C. (2004). Tactile selective attention and body posture: Assessing the multisensory contributions of vision and proprioception. *Perception & Psychophysics 66*, 1077–1094. https://doi.org/10.3758/bf03196837

Spence, C., Pavani, F., & Driver, J. (2004). Spatial constraints on visual—tactile cross-modal distractor congruency effects. *Cognitive, Affective, & Behavioural Neuroscience, 4*, 148–169. https://doi.org/10.3758/CABN.4.2.148

Spence, C., Pavani, F., Maravita, A., & Holmes, N. P. (2008). Multi-sensory interactions. In M. C. Lin, & M. A. Otaduy (Eds.), *Haptic Rendering: Foundations, Algorithms, and Applications* (pp. 21–52). Wellesley, MA: AK Peters.

Toet, A., Kuling, I. A., Krom, B. N., & van Erp, J. B. F. (2020). Toward enhanced teleoperation through embodiment. *Frontiers in Robotics and AI*, *7*, 1–22. https://doi.org/10.3389/frobt.2020.00014

Townsend, J.T., & Ashby, F.G. (1983). *Stochastic modeling of elementary psychological processes*. Cambridge, UK: Cambridge University Press.

Tsakiris, M., & Haggard, P. (2005). The rubber hand illusion revisited: Visuotactile integration and self-attribution. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(1), 80. https://doi.org/10.1037/0096-1523.31.1.80

Tsakiris, M., Longo, M. R., & Haggard, P. (2010). Having a body versus moving your body: neural signatures of agency and body-ownership. *Neuropsychologia, 48*(9), 2740–2749. https://doi.org/10.1016/j.neuropsychologia.2010.05.021

Walton, M., & Spence, C. (2004). Cross-modal congruency and visual capture in a visual elevation-discrimination task. *Experimental Brain Research*, *154*(1), 113–120. https://doi.org/10.1007/s00221-003-1706-z

Zopf, R., Savage, G., & Williams, M. A. (2010). Crossmodal congruency measures of lateral distance effects on the rubber hand illusion. *Neuropsychologia*, *48*(3), 713–725. https://doi.org/10.1016/j.neuropsychologia.2009.10.028

Appendix A: Questionnaire

Table A1. Subjective embodiment questionnaire.

	"During the experiment block there were times when..."	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
1	"...it felt like the hand I was looking at was my own hand."	□	□	□	□	□	□	□
2	"...it felt like the hand I was looking at wasn't mine."	□	□	□	□	□	□	□
3	"...it felt like the hand I was looking at was somebody else's hand."	□	□	□	□	□	□	□
4	"...it felt like I was in control of the hand I was looking at."	□	□	□	□	□	□	□
5	"...it felt like I could move the hand I was looking at if I wanted."	□	□	□	□	□	□	□
6	"...it felt like the hand I was looking at was out of my control."	□	□	□	□	□	□	□
7	"...it felt like my hand was somewhere between the table and the location where I saw the hand."	□	□	□	□	□	□	□
8	"...it felt like I could not really tell where my hand was."	□	□	□	□	□	□	□
9	"...it felt like my hand was in the location where I saw the hand."	□	□	□	□	□	□	□
10	"...it felt like I was looking directly at my hand rather than at an image of the hand."	□	□	□	□	□	□	□