The Cross-modal Congruency Effect as an Objective Measure of Embodiment

Pim Verhagen
Utrecht University
Utrecht, Netherlands
pim_verhagen@live.nl

Irene Kuling
Eindhoven University of Technology
Eindhoven, Netherlands
i.a.kuling@tue.nl

Kaj Gijsbertse
TNO
Soesterberg, Netherlands
kaj.gijsbertse@tno.nl

Ivo V. Stuldreher
TNO
Soesterberg, Netherlands
ivo.studlreher@tno.nl

Krista Overvliet
Utrecht University
Utrecht, Netherlands
k.e.overvliet@uu.nl

Sara Falcone
University of Twente
Twente, Netherlands
s.falcone@utwente.nl

Jan van Erp
TNO & University of Twente
Soesterberg & Enschede, Netherlands
jan.vanerp@tno.nl

Anne-Marie Brouwer
TNO
Soesterberg, Netherlands
anne-marie.brouwer@tno.nl

ABSTRACT

Remote control of robots 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. We want to study the extent to which different factors affect embodiment. As a first step, we aimed to validate the cross-modal congruency effect (CCE) as a potential objective measure of embodiment under four conditions with different, a priori expected levels of embodiment, and by comparing CCE scores with subjective reports. The conditions were (1) a real hand condition (real condition), (2) a real hand seen through a telepresence unit (mediated condition), (3) a robotic hand seen through a telepresence unit (robot condition), and (4) a human-looking virtual hand seen through VR glasses (VR condition). 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 neither found evidence to support the hypothesis that the CCE and embodiment score as assessed by the subjective reports are correlated. These findings raise serious concerns about the use of the CCE as an objective measure of embodiment.

ICMI '20 Companion, October 25–29, 2020, Virtual event, Netherlands

ACM ISBN 978-1-4503-8002-7/20/10…$15.00
DOI: https://doi.org/10.1145/3395035.3425264

CCS CONCEPTS

• **Human-centered computing~Empirical studies in HCI**; **HCI design and evaluation methods**~*Laboratory experiments*

KEYWORDS

Teleoperation; embodiment; multisensory integration; visuotactile integration; cross-modal congruency effect; cross-modal congruency task

ACM Reference format:

Pim Verhagen, Irene Kuling, Kaj Gijsbertse, Ivo V. Stuldreher, Krista Overvliet, Sara Falcone, Jan van Erp and Anne-Marie Brouwer. 2020. The Cross-modal Congruency Effect as an Objective Measure of Embodiment. In *Proceedings of 2020 ACM International Conference On Multimodal Interaction (ICMI'20), October 25-29, Utrecht, Netherlands.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3395035.3425264

1 Introduction

Remote operation of robots generally requires a high level of expertise and may impose a considerable cognitive burden on operators [1]. To enhance task performance and reduce cognitive workload it has been proposed that 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 [2]. This feeling is often referred to as the sense of embodiment, 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 [3]. The concept of embodiment can be divided into three subcomponents: sense of ownership [e.g., 4], sense of agency [e.g., 5], and sense of self-location [e.g., 6]. 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. 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 [7]. This illusion is induced through the multisensory integration between what is seen on the rubber hand and felt on the real hand. Since then, feelings of embodiment have been induced over robotic hands [8, 9, 10] and virtual bodies and body parts [11, 12, 13, 14] through multisensory stimulation. A range of visual and other features have been found to affect embodiment [2]. We aim to assess the relative importance of these features in a single experiment to support the design of interfaces for remote-controlled robots. As a first step, we are looking for an objective measure to quantify embodiment. Quantifying embodiment is not straightforward. Different measures have been used, of which subjective verbal reports have been used most frequently. However, subjective reports may suffer from demand effects and differences in interpretation between individuals.

Aspell et al. [15] propose the cross-modal congruency task (CCT) as a relatively simple objective tool to quantify embodiment, which enables the collection of multiple repeated measures during an embodiment illusion. The CCT has originally been designed to study the multisensory integration of visual and tactile cues [16]. The task consists of indicating the location of vibrotactile targets while ignoring visual distractors as much as possible. We used a unimodal version of the CCT similar to [17] in which two vibrators and two LEDs are arranged on the thumb and index finger of the participant's 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 finger as the vibration, or incongruent: the light flash is presented on the opposite finger of the vibration. Participants have to respond to the vibrotactile stimuli as quickly as possible by indicating on which finger they perceived the vibration, 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. It can be imagined that with a high level of embodiment, one is more hindered by incongruent flashes and supported by congruent flashes (high CCE), compared to a low level of embodiment (low CCE).

Indeed, CCEs have shown to be associated with reported changes in hand ownership [18], self-location [12], and full-body ownership [12, 15]. The CCE has also been used to measure the level of virtual robotic tool incorporation [19, 20]. 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 [15, 17].

To test whether this is indeed the case, 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 reports collected during the same four conditions. These conditions were (1) a real hand condition (real condition), (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), and (4) a human-looking virtual hand seen through VR glasses (VR condition). The latter two conditions are especially relevant for applied teleoperation scenarios.

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. Thus, we expect embodiment to be stronger if the hand is realistic (i.e., human-looking) compared to a robotic hand [2]. Because we presume 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, VR, Robot. A correlation between CCE magnitude and subjective reports would further validate the CCE as an objective measure of embodiment.

2 Methods

2.1 Participants

Eight participants (5 male, 3 female; 7 right hand dominant; aged 23–44 years, mean ± SD = 28.9 ± 7.6 years) participated in the experiment. All participants received oral and written instructions about the experimental procedures and gave their written informed consent to participate in the study before the start of the experiment. The study was approved by the TNO Institutional Review Board.

2.2 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) 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 left hand. On each trial, a vibration (100 ms) and a light flash (100 ms) were delivered simultaneously at one out of the four possible location combinations (congruent thumb; congruent index finger; incongruent thumb; incongruent index finger). Target and distractor stimuli were delivered in a pseudorandom order separated by a random interval between 2000-5000 ms.

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 microcontroller board to achieve millisecond timing accuracy. A custom script on the Arduino Mega controlled the whole experiment. The Arduino Mega drove two physical distractor lights, two virtual distractor lights in Unity 3D, 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.

Following each CCT condition, a 10-items questionnaire was administered in written form to assess the subjective level of embodiment (see Supplement). It contained statements relating to ownership, agency, and self-location. Participants were asked to indicate their level of agreement or disagreement with the statements on a 7-point Likert scale which ranged from "strongly disagree" (-3) to "strongly agree" (+3). Scores for each of the ten statements were combined into a compound embodiment score reflecting the overall embodiment strength. The questionnaire was adapted from [21, 22] and was translated to the participants' mother tongue (Dutch).

2.3 Design

The design had two within-subjects factors: ***congruency*** of the location of the vibrotactile stimuli with respect to the visual distractors (congruent vs. incongruent), and ***condition*** (real, mediated, robot, and VR). The dependent variable was the cross-modal congruency effect in inverse efficiency (CCE-IE). **Figure 1A** depicts the conditions in which the CCT was implemented. The four conditions were presented in a counterbalanced order across participants to compensate for possible learning effects [23]. In each of the four conditions, participants performed 100 CCT trials and completed the embodiment questionnaire immediately afterward.

2.4 Procedure

Participants sat comfortably on a chair, resting their forearms on foam sheets placed on a table in front of them. The vibrotactile glove was worn on their right hand and was visible during the real and mediated conditions. Two distractor lights were fastened tightly around the actuators of the vibrotactile glove. In the mediated and robot conditions a custom-built HMD was put on, which was replaced by an HTC VIVE in the VR condition. Participants were then instructed to indicate as fast as possible the location of sequentially presented vibrotactile targets 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 completing the 100 trials, participants took off the HMD (if applicable) and filled out the embodiment questionnaire. The procedure was repeated for the other conditions. The total experiment lasted approximately 60 minutes per participant.

2.5 Data analysis and statistics

The CCT data were processed to extract the mean response time (RT) for each participant as a function of *congruency* and *condition*. Trials with incorrect responses and with RTs smaller than 200 ms and larger than 1500 ms were discarded, similar to [12]. This led to a rejection of 5.1% of all trials. The inverse efficiency (IE) score [24] 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 used in previous CCT studies [e.g., 10, 12, 19, 25]. In a pre-analysis, we first tested for differences in RTs between responses made with the participant's thumb and index finger. We found no significant differences and, therefore, averaged the responses across thumb and index finger for both congruent and incongruent trials.

Data from all resulting trials were first analyzed using a two-way repeated measures ANOVA (4 x 2) on the mean RTs, with the factors *condition* and *congruency*. Then, 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 of RT data 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, which was normalized using min-max normalization to allow a comparison between each of the four conditions. A one-way repeated measures ANOVA was conducted to examine the effect of *condition*.

Finally, correlation analyses were conducted between the magnitude of participants' CCE-IE scores and the compound embodiment scores. Post-hoc tests for all ANOVAs were applied where appropriate. Requirements for normality of residuals were checked with the Shapiro-Wilk test and reported with its *p*-value (p_{sw}). All statistical tests were performed at a significance level of $\alpha = 0.05$ and were performed using IBM SPSS Statistics 26.

3 Results

The mean RTs and CCE-IE scores are shown in **Figure 1B**. Analysis of mean RTs using a two-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 versus incongruent. 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 that mean RTs were shorter in the real than in the mediated ($p = 0.017$) and VR condition ($p < 0.0001$). Mean RTs were also shorter in the robot than in the VR condition ($p = 0.029$). Residuals of mean RTs were normally distributed ($p_{sw} > 0.20$).

Analysis of CCE-IE scores using a one-way repeated measures ANOVA 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 ($p < 0.00001$), mediated and VR conditions ($p = 0.041$)

and robot and VR conditions (p = 0.00036). Residuals of mean CCE-IE scores were normally distributed ($p_{SW} > 0.36$).

Additional analyses of RTs revealed no significant order effects in the condition blocks. The effects of time were weak and non-systematic.

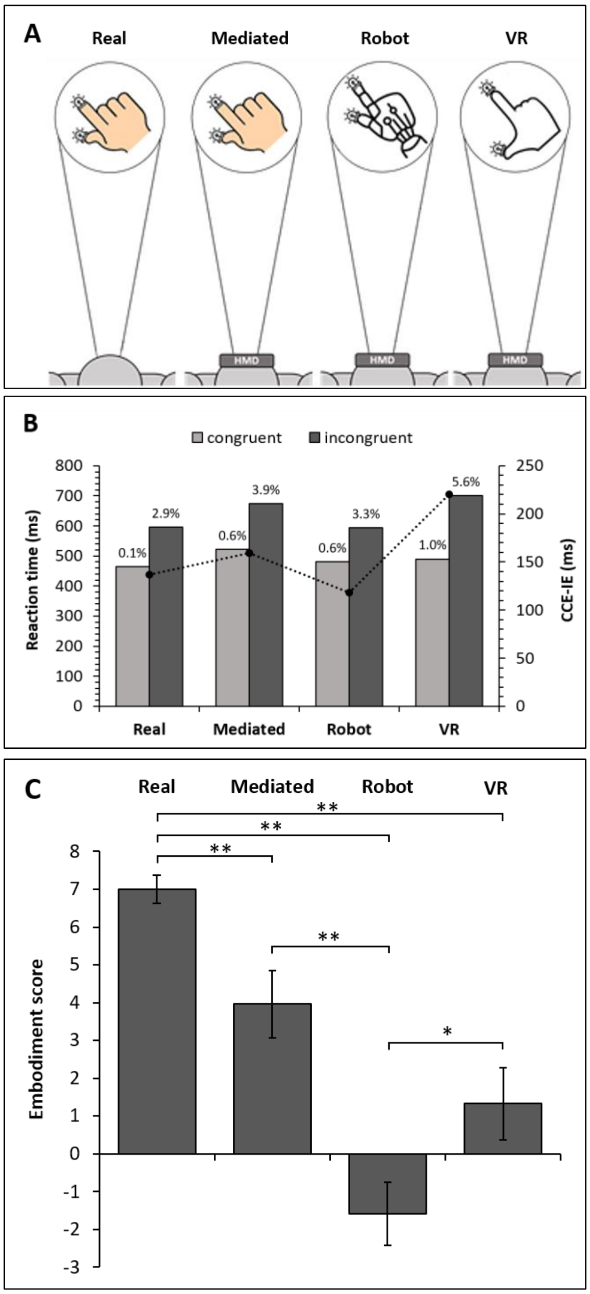

Figure 1: Conditions and results. (A) Four conditions with various visual realism in which the CCT was implemented. (B) Mean RTs of congruent and incongruent trials (left axis). The CCE-IE is shown in black dots (right axis). The error rate is given as a percentage above each bar in the graph. (C) Compound embodiment score as obtained from the questionnaire with standard errors. *p < .05, **p < 0.01.

The compound embodiment scores as obtained from the questionnaire are shown in **Figure 1C**. Results from the one-way repeated measures ANOVA showed that the embodiment score differed significantly between conditions, $F(3, 21) = 31.4$, $p < 0.0001$, partial $\eta^2 = 0.82$. The significant differences in the compound embodiment score between conditions resulting from post-hoc tests are indicated in **Figure 1C**.

For each condition, the correlation between the CCE-IE scores and the compound embodiment scores was tested. Pearson's r showed no significant correlation between the two measures for any of the conditions (all p-values > 0.6). No significant correlations between the CCE-IE and individual questionnaire statements or embodiment subcomponents were found either.

4 Discussion and conclusion

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 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. 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 instead of on a virtual hand. A possible explanation could be that other factors, that (may have) differed between the conditions, affected the magnitude of the CCE, e.g., the *degree of spatial uncertainty* about the vibrotactile target location [see 26], differences in the *relative timing* of the target and distractor stimuli, the *spatial separation* between the vibrotactile target and the visual distractor, or differences in *cognitive workload* in the different conditions. Second, no significant correlation between the CCE and subjective reports was found, suggesting that the two measures do not measure the same phenomenon. Our results challenge the notion that multisensory integration and embodiment are closely connected. Recently, Kanayama et al. [27] conducted an experiment similar to the one in the present study but using the RHI paradigm. They suggested that VR, including an HMD, can disrupt the visuotactile integration process. Additionally, Marini et al. [28] 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 that the CCE is likely to primarily reflect response conflict [25, 28], together with the abovementioned factors. Therefore, we raise serious concerns about the use of the CCE as an objective measure of embodiment.

ACKNOWLEDGMENTS

We are grateful to Ioana Cocu and Martijn van den Heuvel from TNO for their valuable contributions to the experimental setup.

REFERENCES

[1] 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

[2] 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

[3] 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

[4] 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

[5] 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

[6] 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

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

[8] 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

[9] 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

[10] 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

[11] 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

[12] 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

[13] 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

[14] 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

[15] 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

[16] 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.

[17] 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

[18] 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

[19] 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

[20] Sengül, A., Rivest, F., Van Elk, M., Blanke, O., & Bleuler, H. 2013. 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

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

[22] 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

[23] 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

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

[25] 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

[26] 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

[27] Kanayama, N., Hara, M., & Kimura, K. (2020). Disruption by virtual reality of the cortical oscillations related to visuotactile integration during an embodiment process. bioRxiv. https://doi.org/10.1101/2020.03.27.010942

[28] 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