UNIVERSITY
OF TWENTE.

A Multidisciplinary Investigation to Unravel the Complexity of the Sense of Embodiment in Teleoperation

ISBN (print): 978-90-365-5880-8
ISBN (digital): 978-90-365-5881-5
URL: https://doi.org/10.3990/1.9789036

A MULTIDISCIPLINARY INVESTIGATION TO UNRAVEL THE COMPLEXITY OF THE SENSE OF EMBODIMENT IN TELEOPERATION

SARA FALCONE

SARA FALCONE

A MULTIDISCIPLINARY INVESTIGATION TO UNRAVEL THE COMPLEXITY OF THE SENSE OF EMBODIMENT IN TELEOPERATION

INVITATION

to the public defense of my PhD thesis

A Multidisciplinary Investigation to Unravel the Complexity of the Sense of Embodiment in Teleoperation

Friday Octber 27th 2023 at 12:45

Room (WA4)
Waaier Building
University of Twente
Enschede

Paranymphs:
Davide Bicego
Federico Califano

Sara Falcone
*

A MULTIDISCIPLINARY INVESTIGATION
TO UNRAVEL THE COMPLEXITY OF
THE SENSE OF EMBODIMENT IN TELEOPERATION

Sara Falcone

A MULTIDISCIPLINARY INVESTIGATION
TO UNRAVEL THE COMPLEXITY OF
THE SENSE OF EMBODIMENT IN TELEOPERATION

DISSERTATION

to obtain
the degree of doctor at the University of Twente,
on the authority of the rector magnificus,
prof. dr. ir. A. Veldkamp,
on account of the decision of the Doctorate Board
to be publicly defended
on Friday 27 October 2023 at 12.45 hours

by

Sara Falcone

born on the 11th of August, 1993
in Napoli, Italy

This dissertation has been approved by:

Promotors
prof. dr. ir. D.K.J. Heylen
prof. dr. ir. J.B.F. van Erp

Co-promotor
dr.ing. G. Englebienne

ISBN (print): 978-90-365-5880-8
ISBN (digital): 978-90-365-5881-5
URL: https://doi.org/10.3990/1.9789036

Graduation Committee:

Chair / secretary:	prof.dr. J.N. Kok
Promotors:	prof.dr.ir. D.K.J. Heylen Universiteit Twente, EEMCS, Human Media Interaction prof.dr.ir. J.B.F. van Erp Universiteit Twente, EEMCS, Human Media Interaction
Co-promotor:	dr.ing. G. Englebienne Universiteit Twente, EEMCS, Human Media Interaction
Committee Members:	prof.dr. C.J. Rieffe Universiteit Twente, EEMCS, Human Media Interaction
	prof.dr. F. van der Velde Universiteit Twente, BMS, Cognition, Data and Education
	prof. dr. A.-.M. Brouwer Radboud Universiteit, Faculty of Social Sciences
	prof. dr. J.A. Taylor Princeton University

Contents

To Anyone reading these words, I wish you will never stop learning. I wish you will never stop challenging yourself. I wish you will never stop evolving into a better version of who you are.

Acknowledgements

Let's start these acknowledgments with a short little story. This is the second time that I have written them. The first time, I wrote six pages of heartfelt words. However, my laptop suddenly crashed, and those heartfelt words were not part of my last backup. Therefore, here I am again, in my second attempt at expressing gratitude.

I want to be honest with my readers. I thought about not including them. I believe that acknowledgements are a serious matter; they usually make me pretty emotional, and this is (probably) the last thesis in which I will have the possibility to do that. I wanted to do it right. I wanted to feel the moment while writing them. Now, instead, I mostly feel the pressure, the fear of not doing it as well as the first time, and a bit of a hurry. At the same time, a dear friend told me that a thesis without acknowledgements is something very miserable, especially if, as in my case, you have lots of people that you love and who love you back.

Just to guide the reader, a few remarks: I decided to divide these acknowledgements into categories: supervisors, paranymphs, blood family, Marina di Ravenna family, Enschede family, Around the World family, and Princeton family. These categories will not be presented in order of importance. I will take the freedom to switch between English and Italian to better convey my message according to the target.

Dirk, thank you for welcoming me to HMI, and for starting one of the most incredible journeys I have ever experienced (and you know that I am a great traveler!).

Jan, thank you for having been at the same time a great supporter, but also my highly learned opponent. I really needed it.

Gwenn, thank you for having been there in every part of the emotional spectrum of my PhD career. When I started, you told me that the PhD is a lonely journey. You were right, but you let me feel less lonely.

Jordan, you provided me with one of the greatest opportunities of my life, and I will never be grateful enough for that. In the VISA application, at the end of the endless paperwork, it is asked to suggest the name of a contact in the US. I put you, and I labeled you as a *friend*.

Thank you to my family, who gave me the freedom to navigate my own path, even in the face of my inevitable missteps. They taught me to go all-in in every aspect of life. They imparted the wisdom to wholeheartedly embrace life, to love deeply, to exhibit resilience in the face of challenges, to confront fears, and to engage in the ongoing journey of self-discovery. Above all, they taught me the invaluable lesson of cultivating happiness within myself, enabling me to offer the best of myself to those I hold dear. Ai miei nonni, a mia madre, a mio padre, a mio fratello, a mia zia Susi, grazie, vi sento vicini sempre.

To my paranymphs:
Federico, thank you for being here now, and thank you because you will be there also in the future. Our shared strength lies in our ability to deconstruct ourselves and dismantle the old in order to craft a better version of who we are. You can always count on me, and I find comfort in knowing that this is reciprocal.
Davide, thank you for being the purest soul I have ever met in my life. Your smile has reassured me so many times, and your eyes are just the incarnation of kindness. You are one of the bravest people that I know. You chose to be happy. You deserve all the happiness in the world. Nessuno ci toglierà in questi anni un birrino al tramonto, magari su una spiaggia, come i *tori a Pamplona*.

Alla mia famiglia di Marina di Ravenna, mi avete accolta fin dal giorno zero come se fossi sempre stata lì, come se fossi cresciuta con voi. A Marina si respira un'aria unica. È uno di quei pochi luoghi fisici in cui mi sento a casa. Siete il mio porto sicuro. La tappa che non può assolutamente mancare quando torno in Italia. Marina ti rovina, perchè ti ruba il cuore e non te lo restituisce più. Un ringraziamento speciale a quei folli che hanno preso un aereo per festeggiare questo traguardo con me (in ordine alfabetico, con i nomi con cui vi ho conosciuti, non litigate!): Cucci, Marcy, e Piullo, vi voglio bene.

To my Enschede family, there are so many people to mention here. There is so much to be grateful for. Thank you to: Federico Ongaro (I felt that I had to write your surname too), fratello, you will always be part of the Enschede family for me. You have been my best trip companion, and there are still lots of trips ahead of us! I miss you. Kennet, you are the little brother I have never had. Thank you for having taught me so many expressions that I will never forget (and, at this point, I will never even use again!). Luigi, the incarnation of a moviola, the greatest fun fact expert, a real Friend (yes, with capital F) who really respects the meaning of this word; Iris, the greatest latecomer, the most stylish, a dreamer, one of the most helpful and kind people that I know. Guido, you are Enschede; you represent the omen that a night could develop into something completely unexpected (really unexpected); Michela, you have been the mom of all of us; you managed to manage this unmanageable company, fortunately Ale (per sempre piccolo così, per zia Sara), the incredible human being that you (and Guido) created, accepted of sharing you with us. Quentin, mf, you are the out-of-the-box person; for sure we cannot have small talks, however sometimes we do not need to talk, I will always be the black hole in which you can shout your thoughts and let them go. Ani, the most Italian person that I know at this point, you were always there where I needed it, you always tried to help and you never told me no; you saved me many times bro'. Benedetta, zia, sorella, in the context of friendship, we are an example of: 'è intelligente, ma non si applica'. We did not use our time together to develop something that we knew could be incredible. However, we can do it now, and I think we are actually doing it in our own way, that I love. Thank you for sharing your disconnected thoughts, as much confused as mine (I feel less lonely). Raffus, you are the man of the many answers. I will never thank you enough for having been the source of so many solutions. Adedapo, amore, amore, amore, your energy and your attitudes make you so special, you are the person that everyone knows at a party. You still owe me one crazy night in Amsterdam, I do not forget! Francesco, you are the incarnation of the non-linearity, of the efficiency, and you have one of the greatest hearts that I have ever met. Lorenzo, gattuccio, you brought me in Enschede, you represent a game changer in my life. I am so glad that our paths randomly matched at a certain point, and that I met one of the best pun creators that I know (and also a good friend as a side). Jelte, alligatore, to each stage of my PhD path, I can associate one of our walks. They have been cathartic. We have been each other travel journal. I believe that letting you use voice messages in instant chats can be considered one of my greatest accomplishments.

Jacopo, I did not put you in the Enschede family because the truth is that you gave us all of you while you were there, but you never left a piece of you in town. With you, it is not a matter of the place, it is a matter of the people. The connections that you build go beyond any contextual cue. You know that we are so similar that we can almost read our minds (sorry about that). Meeting you was one of the luckiest things that ever happened to me. Ti voglio bene fratello sole.
Alberto, I think you are the friend with whom I laughed the most. We are here, after seven years (yes, seven years), still talking as if we have never left. As if we were still in your apartment in Rovereto, living as an elderly (vegetarian) couple. You are my certainty. I know that I will always find you if I need you, and you can always count on me.
Maria, la pazza (also in this acknowledgements, I decided to use your epithet, because I think that now it is a sort of tradition), we became so connected that we just removed any possible filter and boundary in our interaction. I know that you are the kind of friend that I could contact even after years to say the first thing that comes to mind, and you will just reply to that without expecting any pleasantries. This is also why our conversations are completely incomprehensible to the rest of the world sometimes, such as that time at the vending machine...
Manu, thank you for being a great model of strength, resilience, and intelligence. You are an example of something that I truly admire: never stop learning and never stop evolving into a better version of who we are. Moreover, thank you for the inspiration for the Chapter 7 of this thesis. Most of the credits go to you.

To the Princeton family, it was such an incredible year. Many people gave me so much. I had lots of new experiences. I met so many new and beautiful people. From the academic environment to everyday life, I observed the world with the eyes of a child who was seeing things for the first time. It was unique. There are a couple of people who deserve a special mention:
Sev, I will never forget the night in which we met, and even if you did not know me, you stayed with me and supported me as if we were old friends. You are the captain of the crew.
Isaac, you are one of the funniest, smartest, unfiltered, deepest, and craziest people

that I know. Every time that we do something together, it always develops into some crazy event. I am sure that we can also just take a coffee sometimes, but maybe simply this is not our style. You are a great friend.

Carlos, you are theoretically part of the Princeton family, but of course I reserved a special place for you in these acknowledgements, as you deserve a special place in my life. Thank you for being the choice that I decide to make every day, and for which I truly feel very thankful, every day more. Thank you for being the person who really sees me, and just let me be. Thank you for being the person with whom I want to let myself truly be. Thank you for being as curious as I am. Our career choices bring us to move from place to place, making difficult to call a physical place *home*. We got used to that. However, I have the fortune to feel at home when you are with me. I see the aleph in your eyes.

To myself (my psychologist would be so proud of me), I want to thank myself for my resilience, for my strength, for being independent, but also for understanding when it was necessary to ask for help, and for never stopping trying. I want to also thank myself for accepting, even if it is never easy, all the mistakes that I make, all the missteps that I take, the errors, the issues, and the bad reactions. A special mention to the anxiety monster and the you-are-an-impostor monster, who always put me in trouble, but also taught me a lot about myself. They taught me that, even if sometimes I forget about it, monsters do not exist. However, my mind can be very powerful and conditioning. It is *just* a matter of learning how to deal with this power to get the most out of it. This is a continuous work in progress, and I am definitely still learning. At least I could spot the right path.
To conclude, I have to thank myself for having built such strong relationships with so many people in my life. Even if it hurts to know that most of them are far away physically, I can state, without any doubt, that I have a family everywhere, and the most wonderful kind of family. Surely, this is the best thing that I have ever achieved in my life.

1 Introduction

> No beings, except men, marvel at their own existence; for all animals it is something that is intuited for itself, nobody pays attention to it.
>
> *Arthur Schopenhauer*

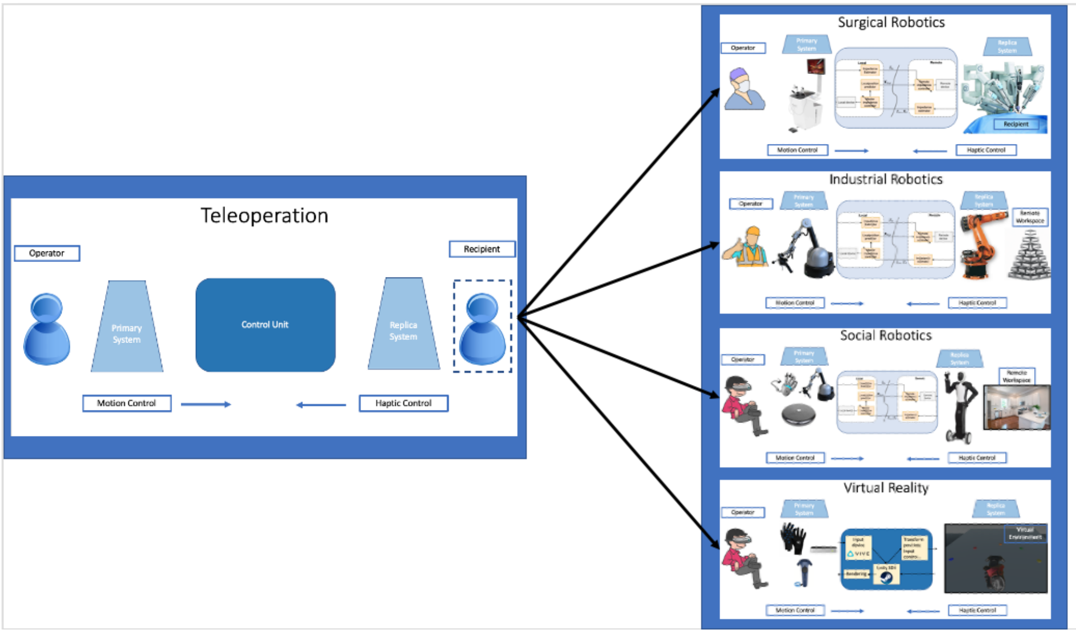
Once, I was explaining my research topic to someone completely unfamiliar with it. He provided me with the best example to introduce it. The conversation started with him asking: «Do you ski?» «No, I just tried once in my life, pretty recently.» «Ok, then you cannot understand. When you have skied for a long time, like me, you no longer perceive the skis as a tool, they are part of your body. You adapt your movements and your way of interacting with the environment according to the 'new shape' of your body. You modulate your body schema automatically!»
«Well, I can understand what you mean because I got the opposite experience: I could perceive the skis as a tool, and I was very aware of the fact that I needed to adapt my movements to the 'new shape' of my body. I had to learn a new way to turn, stop, and even stand! This was the worst part, considering my poor balance skills.» Now this example has to be applied to the teleoperation domain. Removing physical distances and experiencing remote events in real time is one of the big challenges in modern society. Telecommunication, teleoperation, and telepresence are becoming mainstream terms and generally expected achievements. Telecommunication is a branch of technology aiming at finding solutions to exchange information over distance [43, 66, 130]. Teleoperation is a specific sector of telecommunication, defined as remote control of a device or machine [2, 99, 159]. Finally, the concept of telepresence is deeply related to the previous two: It refers to the feeling of being in another location than one's physical body [49, 80, 269]. Nowadays, telepresence has been achieved in several ways; one example is by transferring sensory feedback using remote cameras and microphones to a user, usually through a Head Mounted Display (HMD) and a (stereo) headset. Including other sensory modalities other than

vision and audition can improve the user experience [257]. There are several ways to increase the operators' engagement and immersion, namely to completely absorb the operators' attention and their physical experience. An example is haptic feedback, which can be provided by devices and sensors attached to the user that respond to the information and signals collected from the remote environment. These signals can come from sensors that are part of a robotics platform, or from the input provided to a virtual avatar interacting with a Virtual Reality (VR) environment. Telepresence systems have been applied in several contexts: remote inspection, education, machine operation, explosive ordnance disposal, and entertainment. However, due to the current state-of-the-art of the technology, setups mostly focus on rendering vision and audition, and they do not focus on a design that could optimize the other senses. Tactile feedback, for example, is limited by the current haptic devices and sensors aimed at rendering this sense. These devices usually provide vibration as feedback, and cannot realistically reproduce textures. In Human-Computer Interaction, one of the current solutions is the mixed reality approach [239], which represents a mixture of Augmented Reality (AR) and VR. It is characterized by the combination of real world elements and the virtual reality environment to provide more realistic haptic feedback to the user. However, this cannot represent a definite solution if we are looking for a full immersive experience in the remote environment.

This work aims to enhance the teleoperation experience by building a Sense of Embodiment (SoE) in humans. SoE refers to the process of developing a sense of ownership (feeling of self-attribution), a trustworthy sense of agency (feeling of control), and a sense of self-location (the perception of being located in a volume of space) over a surrogate that could be represented by a mannequin, a robotic or virtual avatar. In this research, we focus on remote avatars controlled through the teleoperation setup [76, 82, 132, 160, 246] (see Figure 1.1 for a general representation of the architecture of a teleoperation system and three sample applications). As reported in [189], the term *avatar* lacks a standard definition. In this thesis, the term *remote avatar* refers to the surrogate operated by the user, which could be a virtual embodied agent or a robotic device. By leveraging theoretical frameworks and methodologies from multiple disciplines such as robotics, cognitive science, and neuroscience, we address the limitations encountered in designing a teleoperation setup aimed at achieving a high level of embodiment; limitations such as the haptic feedback rendering, the optimization of the perceptual cues, and reliable measures of embodiment.

In this thesis we address 11 main research questions to investigate and explore the role of the SoE in teleoperation. SoE has been studied originally and extensively using the paradigm of the Rubber Hand illusion (RHI) ([29]). The Rubber Hand illusion is a perceptual illusion in which individuals experience a fake model hand as being part of their own body: a sense of ownership is created. To do that, the experimenter provides simultaneous tactile stimulation to both the real and fake hand. Finally, the experimenter introduces a threat (e.g., stabbing the fake hand with a knife or hitting it with a hammer) to break the illusion and record the participants' reaction. If the illusion happens, the participants will be scared by the threat. The Rubber Hand illusion demonstrates that the combination of visual and tactile signals strongly influences the subjective experience of body ownership. Since the original Rubber

Figure 1.1: Schematic and general representation of an architecture of a teleoperation system, and three possible application scenarios.



Hand illusion studies, numerous studies have found that it is possible to induce a strong SoE over virtual and real extracorporeal objects such as fake limbs, robotic hands and arms, mannequins, virtual bodies and even empty volumes of space and invisible bodies ([42, 105, 140, 255]). These objects include ones that, in contrast to the original paradigm, can be controlled by the user, such as is the case in telerobotics. Although the importance of the SoE is acknowledged, there is no standard framework to test (checking the SoE levels in a task), measure (reliably addressing and collecting information coming from the SoE manipulation), and assess it (evaluating the quality and level of the SoE). Results are difficult to replicate and compare across studies. In the literature, we can find a large variability among the experimental setups, the method used to induce the embodiment experience, the quantification of its effects, and the experimental design and data analysis [212]. Moreover, we argue that the design of a teleoperation system or an embodiment experience could benefit from a standard procedure to assess them both in the development phase and also in the implementation tests. Therefore, we investigate:

1. How can we define a standard framework or guidelines to design a teleoperation setup aimed at optimizing the embodiment experience of the operator?

The optimization of the embodiment experience brings us to the introduction of another concept that we define as the *mediator*, which refers to the level of perception of the setup or control pod, and the distance that they create between the operator and the avatar, by reducing engagement and immersion. In an ideal SoE experience, the mediator perception should be null. This concept is particularly important when we investigate the phenomenon of SoE in relation to teleoperation, due to the multisensory integration and perceptual cues affected by the use of the setup to control the remote avatar.

2. To what extent does the perception of the mediator affect the SoE experience and task performance? Does this concept play an important role in the teleoperation applications?

The most common methods for quantifying the strength of the embodiment experience includes subjective self-reports assessed with questionnaires or interviews (explicit measures), and measures of the perceived position of the own hand, physiological measures, or task performance (implicit measures) [76, 212]. In RHI questionnaires, subjective experiences relating to the sense of ownership over the artificial hand, the sense of agency over movements of the artificial hand, and the sense of self-location are typically rated on a Likert scale [29, 198]. Verbal or behavioural judgments about the own hand's location usually reveal a systematic mislocation of the unseen own hand towards the artificial hand, a phenomenon commonly referred to as proprioceptive drift [127, 214]. Physiological measures reveal an arousal response to emotional stimuli [69, 75]. The current state-of-the-art presents lots of variability concerning the adopted measures. The research field, the context of application, and the setup can affect the choice and combination of those. Previous findings [212] show that the unconscious experience of embodiment, measured by implicit measures, appears to be more consistent among participants and to reflect an immersion that is not consciously perceived by the individuals, and therefore it is not reflected in the explicit measures. Collecting proprioceptive information from the participants seems to be the current most effective way to measure and assess the SoE. It is necessary to incorporate more informative and less noisy implicit and explicit measures, and use neuroscientific evidence to build a stronger theoretical understanding of SoE. This could explain how our brain works in experiencing the remote environment through the mediator filter, starting from the source of our behaviour instead of just looking at the effects. This triggered our curiosity to explore:

3. What different pieces of information about the embodiment are revealed, respectively, by the explicit and implicit measures?

4. Can the current embodiment measures disentangle the embodiment components?

The interest in designing and measuring embodiment is carried by one goal: implementing systems that enhance a telepresence experience beyond the feeling of being present at a remote location, by transporting both the functional and social self of the user [257]. The SoE and the multisensory integration of the perceptual cues become two relevant points to address in the design of a telepresence experience [74, 76]. In [246], the authors discuss two cognitive models that are currently used to describe the sensory processes underlying embodiment. One model builds on Bayesian perceptual learning ([11]), and postulates that multisensory brain areas, such as the premotor cortex and posterior parietal cortex ([70, 71]), integrate signals from different modalities that co-occur with a high probability in near-personal space ([71, 166]). Two perceptions from different modalities are 'bound' when they

co-occur with a high probability. For example, in a RHI setup, if participants observe the fake hand being touched by a brush in the same position as they feel their real hand being touched, they incorporate this into their bodily representation and expect new visual feedback to co-occur with coherent tactile feedback as well. According to this theory, the perceptual cues interact and influence each other. This means that the manipulation of one cue changes the perception of the other cues involved in the experience and, as a consequence, affects the SoE.
Another neuroscientific framework to understand SoE is the predictive encoding theory ([85, 86, 117]). In contrast with the Bayesian perceptual learning process, predictive encoding postulates that the brain produces models at each level of perceptual and cognitive processing to predict what information it should receive from the level below it (i.e., top-down). Then, the brain compares the bottom-up sensory information with the predictions from the model. The discrepancies between both (the prediction errors) are the only elements that are passed to higher levels, where they are used to update the model or resolved by activating a different model. Both these actions (model updates and activation) are aimed at minimizing or suppressing prediction errors at a lower level ([86, 87]). This theory relates to the SoE in the sense that a high error is associated with a low SoE ([9]). According to this theory, the model can update each perceptual cue individually and there is no interaction effect among the cues. The manipulation of the perceptual cues consists of affecting the multi-sensory integration of external stimuli obtained through the control of an avatar in order to increase or decrease the level of embodiment of the operator. This manipulation can be obtained by changing the field or point of view of the camera used in the surrogate cockpit, by activating or deactivating the tactile feedback, or by choosing a certain haptic device rather than another. The effect of the manipulation of the perceptual cues on the SoE remains still uncertain. Moreover, it is unclear how the manipulation of the SoE is reflected in task performance. This brought us to formulate the following research questions:

5. To what extent do perceptual cues affect the SoE components and task performance? Can we rank them in order of importance? And is the order consistent over the different SoE components and task performance?

6. Are there interaction effects between the perceptual cues? Is a simple additive model sufficient or do we need a more complex model to test the effect of the perceptual cues together?

7. How are SoE and task performance related?

Starting from the literature, it is assumed that different levels of SoE through an avatar, obtained by the manipulation of several perceptual cues, would have had an effect on the task performance. In cases in which operators feel strongly embodied, task performance would improve compared to situations in which the embodiment level is low [167, 217, 220, 262]. However, these results do not necessarily generalize to teleoperation, since they are mostly limited to VR studies, or they are conducted in the prosthetic field with participants who experienced an upper limb loss; another aspect is that they usually focus on cognitive performance and not on motor performance. Moreover, there is still a debate if high SoE really leads to better performance. The effect of SoE on task performance has not been widely replicated and there are also studies that found no effect or just different advantages in manipulating certain perceptual cues [62, 74, 101, 137]. Differences in motor learning can be an alternative explanation for the inconsistent results. Motor learning can be defined as any experience-dependent improvement in performance [142]. Explicit and implicit processes both contribute to how we learn new motor skills. Implicit adaptation serves to maintain motor performance in a fluctuating environment through a sensory prediction error-driven learning mechanism. Discrete sequence learning tasks reveal how we anticipate temporal regularities in the environment, but are not likely good models for skilled continuous sequential actions [171]. Many skills, like riding a bicycle, cannot be assembled from pre-existing skills and require building a new control policy. The same applies to the manipulation of a robotic avatar or device. The quality of movement execution can be improved through practice. The human operators are required to re-calibrate their body schema, namely the sensorimotor representations of the body that guide actions [60], considering the setup that mediates between their own body and the robotic surrogate in the remote environment. This implies that human operators need to build a new control to perform actions that they already know how to perform with their own bodies:

8. What is the effect of SoE on task performance in a perceptual-motor task?

9. What is the effect of SoE on the asymptote of the learning curve (i.e. the performance level after the learning curve reached a plateau) in a perceptual motor task?

Guided by the curiosity to get a full exploration of this complex concept of embodiment, we investigate the mind of the operators, and to also explore the recipient experience. In fact, a teleoperation scenario can involve two main users: the operator, who controls the avatar, and the recipient, who interacts with the avatar in the remote environment. We decided to approach this topic from a classic Human Computer Interaction (HCI) perspective. Many HCI researchers aim to create an 'artificial social entity' that is as human-like as possible, in both the (non-)verbal behaviour it exhibits and in the way its body looks. The term artificial social entities can cover a broad spectrum of technical artifacts, ranging from chatbots to virtual characters to physical social robots. Our interest focused on:

10. What enhances the telepresence perception of a social entity?

11. To what extent does the way in which the body and the mind of the operator are transferred to the remote avatar play a role in the recipient's experience and interaction?

Overall, the aim of our research is to use a multidisciplinary approach to improve our understanding of SoE and its relation to teleoperation. Based on the literature, we predicted an initial model that could represent the relation between SoE and teleoperation (see Figure 1.2), with the aim of updating it and include the new gain

knowledge that we got at each step of the investigation. Initially, we described a process in which the stimuli from the remote environment are shaped by some perceptual cues (which were not well specified), that are integrated by the operators and that indistinctly affect the embodiment components. The sum of the levels of these components, dependent on the integration of the perceptual cues, gives the level of embodiment, that is positive correlated to task performance. In the Discussion chapter, we present the final updated model (see Figure 8.1), which represents our conclusions from this collection of studies.
Another goal is to start to build theoretical frameworks based on empirical evidence. We should define what we empirically observe to describe and understand it, share the knowledge, compare the observations, generalize the phenomena, and build new knowledge on previous research. This allows us to test, measure, and categorize SoE on a common basis, and to apply this knowledge to the new society based on telecommunication. The interest in telepresence systems has strongly increased over the past years, and the global Covid-19 pandemic, resulting in restrictions on traveling and social interaction, has provided an additional boost. The current necessity is to ground the current knowledge about embodiment, to allow its manipulation and to transpose it to the engineer dimension, to implement those concepts and findings in teleoperation systems. Finally, the next crucial step would be to integrate those systems in the modern society and make them accessible to everyone. Our final research goal would be to pursue the objective of multi-disciplines integration, and an integration between research and real-world applications in the society.

The next chapters will provide a global picture of what we know and, especially, about what we still need to investigate. Following the above presented research questions, Chapter 2 introduces the main concepts related to the Sense of Embodiment, its components, its relation to teleoperation, how to test and measure it, the concept of the mediator, and the state of the art (RQ1 and 2). The main purpose is to present a toolbox to achieve, based of the literature and practical examples, the desired SoE while designing and implementing teleoperation systems. The contents of this chapter are published in [76]. Chapter 3 concerns a topic when coming to embodiment: its assessment. It introduces the difference between explicit and implicit measures, and it provides an overview on the most commonly used assessment measures and some new approaches that we propose (RQ 3 and 4). The content of this chapter is based on a collection of publications [75, 78]. Chapter 4 presents a study that we realized to define and rank the effect of the most relevant (according to the literature) perceptual cues that affect the embodiment components and task performance (RQ 5, 6, and 7). Perceptual cues are fundamental to manipulate SoE, and their understanding is crucial while realizing a teleoperation system. The content of this chapter is published in [74]. Chapter 5 investigates the effect of SoE on task performance in teleoperation scenarios (RQ 8 and 9), going beyond the current literature mainly focused on prosthetic. Most of the contents of this chapter are collected in a paper that is now being peer reviewed. In the last chapters of this thesis, we will approach embodiment from different perspectives, trying to observe how the complexity of this phenomenon was recognized in several other fields. Chapter 6 explores the experience of the recipient. We present a reverse study on embodiment, in which we built a

Multimodal EchoBorg (MEB) system to try to observe from the inside what it means to be a remote avatar (RQ 10 and 11). The contents reported in this chapter are published in [77]. Chapter 7 presents a dissertation to address the concept of embodiment from different disciplines, focusing on the *Phenomenology of Perception* of the philosopher Merleau-Ponty, with a look at the technology perspective and the possible implications on the society. This work is still unpublished. Finally, there is a chapter dedicated to the General Discussion and Conclusions, to elaborate on and summarize all the previous presented material.

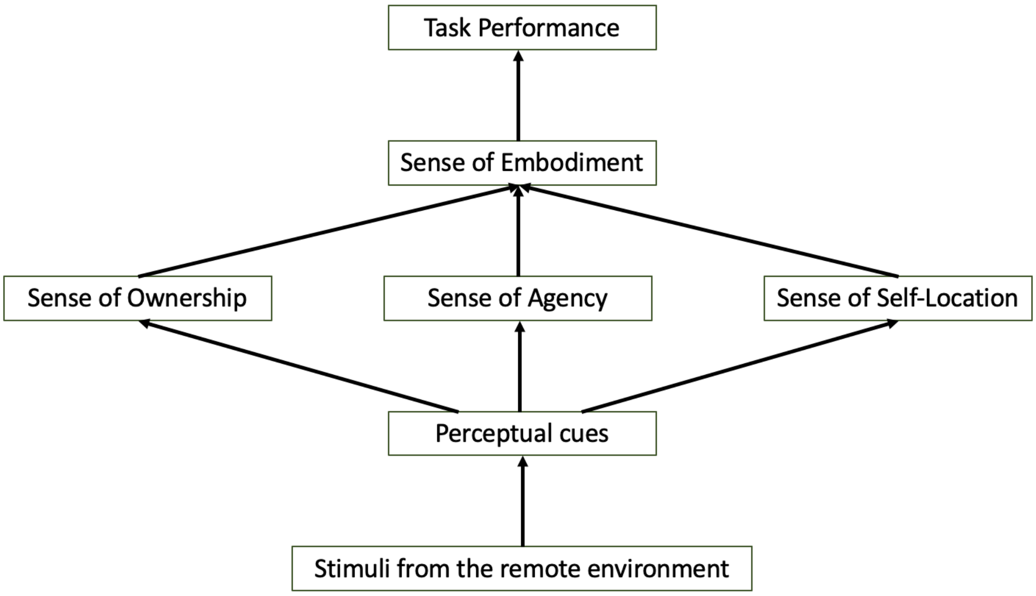


Figure 1.2: The figure represents an architecture of the predicted initial model of the Sense of Embodiment in teleoperation.

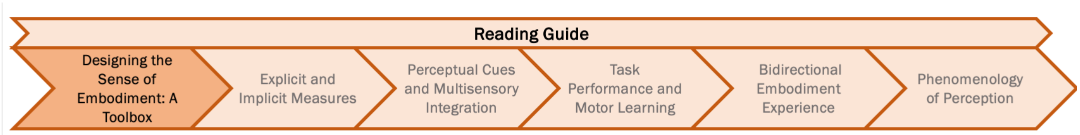
2

2

Towards Standard Guidelines to Design the Sense of Embodiment in Teleoperation Applications

Design is intelligence made visible.

Alina Wheeler



Abstract

We present a literature review and a toolbox to help the reader find the best method to design for and assess the Sense of Embodiment (SoE) in several application scenarios. The main examples are based on teleoperation applications, due to the challenges that these applications present. The three components of the embodiment that we consider to describe the SoE are the sense of ownership, the sense of agency, and the sense of self-location. We relate each embodiment component to the most often used assessment measures, test tasks, and application scenarios. The toolbox is built to efficiently design, test, and assess an embodiment experience, following 7 concrete steps. We provide four main contributions: 1) a literature review of the assessment measures and strategies used to measure the SoE; 2) a systematic categorization of the SoE measures; 3) a categorization of the main test tasks used in SoE assessment. These three contributions were used to arrive at the fourth one, namely 4) a toolbox consisting of 7 steps as guidance to design SoE. We included several examples and tables to guide the user step by step through the design of an embodiment experience.

2.1 Designing the Sense of Embodiment: A Review and Toolbox

The Sense of Embodiment (SoE) can be defined as the experience in which the external body, body part or object (such as a rubber hand, a mannequin, a robotic device or a virtual avatar) is perceived as their own [132]. Throughout the rest of this chapter, we will refer to the external body or part of it as *surrogate*, while the term *operator* will be used to refer to the person who goes through an embodiment experience over a surrogate. SoE lacks a standard definition, compared to the most used and acknowledged definitions [61, 132, 160, 175], we highlight that SoE has a subjective and objective component. The subjective component is related to the individual update of the level of embodiment that everyone performs intrinsically and unconsciously, mostly on the basis of perceptual cues. The objective one concerns the parameters of the system or the experience (such as the setup, the context, and the tasks) that can be manipulated in order to increase or decrease the SoE. This is strictly related to the concept of the mediator that we previously introduced. The mediator represents the level of perception of the setup, of motor adaptation to it, and of transparancy between, using the telerobotics lexicon, the primary (the controller) and the replica (the remote avatar or device) as perceived by the operator. Ideally, if the teleoperation system allows one to achieve a high level of SoE, the perception of the setup as a mediator between the operator and the avatar is low. The aim is to completely cancel the mediator perception, to create a one-to-one perception of the avatar. This means that the body schema of the operator is completely updated and the avatar becomes the new body.

The guidance provided in this study is applicable to different embodiment experiences and virtual reality (VR) setups. However, there will be a focus on teleoperation, since it offers the most difficult challenges for designers and developers.

Although the importance of the SoE in the previously cited fields is acknowledged, there is no standard framework to test, measure, and assess it. Results are difficult to replicate and compare across studies. Moreover, we argue that the design of a teleoperation system or an embodiment experience could benefit from a standard procedure to assess them both in the developing phase and also in the after-implementation tests. Therefore, we provide a literature review and a toolbox to guide the reader in designing, testing, and assessing the SoE in a system.

In this respect, our work is comparable to previous reviews, but it has some significant differences [100, 132, 246]. In [132], the authors present a literature review of the structure, measures, experimental manipulations, and challenges related to the SoE with a virtual body in a virtual immersive environment, but not in other applications such as teleoperation in a real environment. The authors define the SoE using the same components considered in this work. These same components are considered in [246], in which the authors focus only on telerobotics; they also provide a set of guidelines for applying the SoE in telerobotics, identifying some important challenges and research topics. Eventually, [100] focuses only on virtual bodies, and the aim of the authors is to define a standard questionnaire to assess SoE. The authors present a literature review of the most commonly used questionnaires and the SoE structure described by more components than those considered in this work. In particular, they include the *tactile sensation*, the *external appearance*, and the *response to external stimuli*. We consider these additional components as part of the sense of ownership, following the research lines of [132] and [246]. Our work considers both the SoE in virtual bodies and telerobotics. In addition, we provide the reader with a toolbox for designing, testing, and assessing the SoE in a teleoperation system or in an embodiment experience. We categorize the most used assessment measures and relate them to each component of the embodiment, the tasks, and the application scenarios.

Four reference figures and tables are presented in this chapter. We labelled, grouped, and categorized the design steps, assessment measures, and perceptual cues on the basis of the reported literature review:

- *Figure 2.1* represents a general guideline to apply the toolbox step by step;
- *Figure 2.2* sums up our findings related to the sub-categories of the categorized assessment measures;
- *Table 2.1* associates the perceptual cues considered in this study with the embodiment components that are mostly affected by them;
- *Table 2.2* shows if a specific assessment measure is suitable, non-specific, or unsuitable for each embodiment component.

The chapter is structured as follows: in Section 2.2 we present the methodology to realize the literature review. Section 2.3 is dedicated to a detailed presentation of the steps of the SoE toolbox. This section is structured following Figure 2.1. Section 2.3.9 includes our observations on the most used assessment measures and methodology, and reports the (dis)advantages for each assessment measure. Section 2.4 is an overview of the toolbox use and includes two example cases built with the toolbox application. Finally, in Section 2.5 we list conclusions and recommendations.

2.2 The Methodology

We examined the academic literature to provide a comprehensive review of the assessment measures and tasks that can be used to evaluate and test the SoE and its three components. After the categorization of the assessment measures and tasks, we identified if they can be applied to assess and test each single component, multiple components together, or if the distinction is not clear. The aim is to present a first step towards an SoE toolbox, associating the embodiment components to the involved sensory factors, the assessment measures, the test tasks, and the considered application scenarios.

Publications on testing and assessing the SoE and the most commonly used assessment tasks were identified through a systematic literature review. We included searches from Google Scholar, Elsevier Scopus, IEEE, and Mendeley, from 1991 to 2021 (with a particular focus on papers from 1991 and 2019). The reason for combining the different sources was that Google Scholar is not enough to be used alone for systematic reviews [98]. The search terms that were used were ‘sense of embodiment’, ’embodiment’, ‘ownership’, ’agency’, ’self-location’, ’evaluation’, ’assessment’, 'measures', 'metrics', 'teleoperation', 'telemanipulation', 'teleoperation task', 'teleoperation taxonomy', 'teleoperation toolbox', 'user case study', 'VR setup', 'virtual reality', 'embodiment experience', 'avatar', 'embodiment illusion', in various combinations (e.g., adding "and" between the terms or combining them: "teleoperation metrics"). Papers were selected from the domains of engineering, technology, psychology, and neuroscience. The initial search provided about 10300 results. By limiting our search to articles in English mostly reporting empirical studies in peer-reviewed journals, and excluding articles targeting animals and brain injured populations, we obtained 1800 articles. After reviewing the title and abstract of those articles, we excluded those that did not meet the inclusion criteria reported above. The search was further refined by adding search terms referring to the perceptual cues, test tasks and assessment categories (such as 'tactile feedback', 'visual feedback', 'questionnaire', 'proprioceptive drift', 'proprioception', 'skin conductance response', 'SCR', 'peg-in-hole', 'time delay'). This resulted in 161 articles, which were reviewed in chronological order to understand the evolution of the evaluation metrics, the test tasks, the research aims, the technical developments, and the related works.

2.3 Toolbox and Guidance: Five Steps to the Sense of Embodiment

In this section, we explain the steps in the toolbox in more detail and provide guidance on how to complete the steps to design and assess the SoE.

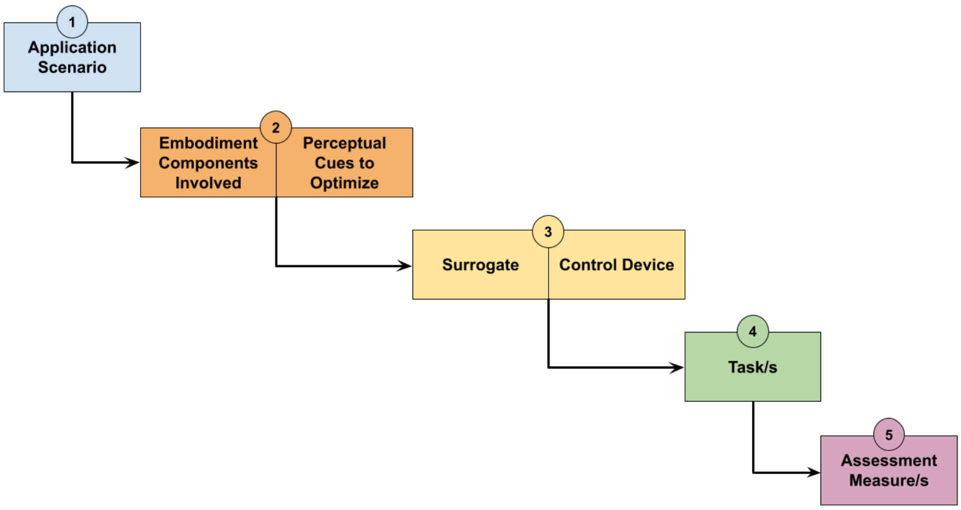


Figure 2.1: Schematic representation of the step-by-step guide to use the toolbox. The first step is to define the application scenario (step 1) and the embodiment components (step 2a) and perceptual cues (step 2b). The next step is to define the embodiment experience through the setup: the surrogate (step 3a) and the control devices (step 3b). Then, step 4 is to choose a task, or a combination of tasks, to test the system. Finally, in step 5, the assessment measures and their customization are selected.

2.3.1 Step 1. Application scenarios

We focus on four application scenarios and their relation with the components of SoE. These are high level scenarios that illustrate our general recommendations, though each individual case may have its own considerations.

2.3.1.1 Social

A social scenario refers to the situations in which the operator has to interact with other individuals in a dynamic and, often, partially unknown environment. In this scenario, the operated device must enable tasks such as: shaking hands, giving hugs, making eye contact, expressing gesture, adjusting posture, touching and manipulating everyday objects, moving around in an apartment, expressing emotions, and displaying other cues relevant in social interaction [31]. Usually, the robotic devices are humanoid in this application context. For social robotic telepresence, we agree with [45] in pointing the attention on four main aspects: 1) the mechanical design, 2) the user interface design, 3) the interaction between the surrogate and the operator's real body and 4) the subjective perception of the telepresence level of the system. Moreover, both [45] and this work consider the comparison between robotic and non-robotic systems. For a social scenario the relevant embodiment components are the sense of ownership, agency, and self-location.

2.3.1.2 Industrial

In industrial scenarios, the environment is usually static and the actions predictable. The operated device will have to manipulate tools, move objects, and move around in an environment that is not open air, such as a factory. The environment may not be human-friendly due to, for instance, high temperature, tiny dimensions, or high security risks [114]. Examples of tasks in this scenario are maintenance using tools or robotic arms, and moving large masses. While designing a teleoperation system that has to perform in an industrial scenario, the focus will be mostly on the performance and not on the SoE. The operator's experience refers to a more classical concept of user experience: efficiency, effectiveness, usability and ease of use [206]. In [247], for example, the authors present a design methodology to improve the operator's performance and experience in industrial scenarios, introducing the new concept *whole-arm manipulator* (WAM). The design goal is that the WAM should be able to control forces and torques robustly along all of the outer link surfaces. It also refers to the degrees of freedom for an operator to manipulate any intermediate structure independently of the end-effector by exploiting redundancy in the kinematics. For an industrial scenario the relevant embodiment components are the sense of agency and self-location.

2.3.1.3 Field

Field scenarios include tasks such as inspection & maintenance [202], and search & rescue [25] in unstructured environments. Unstructured environments are dynamic and unpredictable. These kinds of scenario are also called exploratory robotics. The robotic devices for this scenario are often tracked vehicles or animal shaped with multiple legs. In a literature review by [108], the authors compares different complications that could be encountered while teleoperating in a field scenario. The focus is on technical issues, such as data transmission, the choice of the setup and measurements. For a field scenario the relevant embodiment components are the sense of agency and self-location. The sense of ownership can be relevant if the operator needs to receive multimodal haptic feedback to optimize the teleoperation experience and to improve the performance.

2.3.1.4 Surgical

Robot-assisted surgery was developed to overcome the limitations of pre-existing minimally invasive surgical procedures and to enhance the capabilities of surgeons. The surgeon uses a direct telemanipulator, or a computer control, to control the device and the instruments. Another advantage of using robot-assisted surgery is that the surgeon does not have to be present, leading to the possibility for remote surgery. In this scenario, tasks of microassembly and microteleoperation are common. The main challenge is to create a connection and transparency between the macro world of the operators and the nano world in which they have to tele-operate the system. Particularly, the focus is on optimizing the motion control in constrained workspaces [90], and increasing dexterity and degrees of freedom [164] safely. For a surgical scenario, especially due to the importance of tasks involving hand-eye coordination, the relevant embodiment components are: the sense of agency and self-location.

2.3.2 Step 2a. The Sense of Embodiment and the embodiment components

Based on [132], we use the term SoE as an overarching construct including the sense of ownership [144], sense of agency [183, 184], and sense of self-location [12]. Unlike [100], as we already mentioned, we include the *tactile sensation*, *the external appearance*, and the *response to external stimuli*, as sub-components of the sense of ownership. Numerous studies have found that is possible to induce a strong SoE over virtual and real extracorporeal objects such as fake limbs, robotic hands and arms, mannequins, virtual bodies and even empty volumes of space and invisible bodies [42, 105, 140, 255]. While operating a machine remotely, with a high level of SoE, the operator's perception of the remote device as mediator decreases [38], increasing the teleoperation system transparency. Starting from this intuition, some studies try to demonstrate that a high level of SoE can improve teleoperation tasks performance [167, 217, 220]. In [246], the authors use the predictive encoding theory [85, 86, 117] as a neuroscientific framework to interpret and discuss their findings on how the SoE affects teleoperation performances. The predictive encoding theory postulates that the brain produces models at each level of perceptual and cognitive processing to predict what information it should receive from the level below it (i.e., top-down). Then, the brain compares the bottom-up sensory information with the predictions from the model. The discrepancies between both (the prediction errors) are the only elements that are passed to higher levels, in which they are used to update the model. The model updates are aimed at minimizing or suppressing prediction errors at a lower level. This theory can be applied to the SoE in the sense that the brain effectively updates the body model in a bottom-up way by minimizing prediction errors between its top-down predictions and actual sensory events. The predictive encoding framework allows to interpret and discuss data on embodiment experiments and to make predictions for embodiment effects in telerobotics. Moreover, in [246], the authors present four premises on the relation between the SoE and teleoperation, that are supported by the literature: 1) the brain can embody non-bodily objects (e.g., robotic hands, animal-shaped robots, non-humanoid virtual avatar), 2) embodiment can be elicited with mediated sensorimotor interaction, 3) embodiment is robust against inconsistencies between the robotic system and the operator's body, and 4) embodiment positively correlates to dexterous task performance. One of the main debates and challenges concerns how to clearly disentangle the SoE components. In [160], the authors present a psychometric approach to disentangle the embodiment components. Participants had to perform a proprioceptive judgment task while experiencing a rubber hand illusion (RHI). The authors confirmed previous findings about how the sense of ownership and agency reflect dissociable components of the embodiment [93, 240, 251, 252]. The same was claimed for the sense of self-location, even if this component presents a strong correlation with the sense of ownership. Another line of thought concerns the strict dependency and the mutual influence that the components, particularly the sense of ownership and agency, have on each other. In [13], the authors used the proprioceptive drift as a measure to assess the SoE during an RHI. They concluded that the explicit sense of agency can arouse an implicit measure of the sense of ownership. In [94], the author presents the ambiguity of the sense of agency, focusing on the difficulty of disentangling this component from other factors influencing the embodiment experience. The exact relation between the sense that one's body is one's own (body ownership) and the sense that one controls one's own bodily actions (agency) has been the focus of much speculation, but remains unclear. On the base of [251], we can consider two models to describe the relationship between the sense of ownership and sense of agency. First, an *additive model*, in which agency and body-ownership are strongly related, because the ability to control actions is a powerful cue to body-ownership; plus possible additional sub-components unique to ownership and agency. An alternative *independence model*, sustains that agency and body-ownership are qualitatively different experiences triggered by different inputs and recruiting distinct brain networks. We can divide the brain regions involved in two main groups. The first group of brain regions constitutes a network of sensorimotor transformations and motor control, whereas the second group of brain regions represents a set of hetero-modal association cortices implicated in various cognitive functions. Unfortunately, we still do not know the exact functions and contributions of these brain regions to the sense of agency. Several studies are aimed at identifying the neural correlates of two different judgments of attribution: experiencing oneself as the cause of an action (the sense of agency) or experiencing another person or object as being the cause of that action [79, 249].

2.3.3 Step 2b. Perceptual cues to optimize

The perceptual cues positively affect an embodiment component when they reassemble or recreate experiences to which the human brain is used, reducing the prediction

Table 2.1: On the basis of the reported literature review, we associate to each perceptual cue addressed in this study the SoE component(s) they support. Legend: O → Sense of Ownership; A → Sense of Agency; L → Sense of Self-Location.

Perceptual Cue	**SoE component(s) involved**
Point of view	O, L
Field of view	O, L
View direction control	O, L
Connectedness	O, A
Visuo-proprioceptive sync.	O, A
Visuo-tactile sync.	O, A
Visual likeness emb.	O
Haptic feedback	O, A
Visual likeness env.	L

error as explained in the previous section. The relevant perceptual cues and which embodiment components they mostly affect can be found in Table 2.1.

2.3.3.1 Point of view

The point of view is the perspective from which the operator observes the remote environment and experiences the surrogate. Usually, the main distinction is between first person perspective (1PP) and third person perspective (3PP). The extensive literature on this cue shows that a 1PP is sufficient to create the SoE, but is not strictly required, as the SoE also occurs in 3PP, in the absence of visual cues or with an incongruent visual perspective. However, as the authors conclude in the literature review from [246], 3PP alone (i.e. in the absence of other perceptual cues) is not sufficient to create an embodiment experience. In more detail, the sense of ownership over a virtual body can be obtained in both 1PP [168, 201, 235]) and in 3PP [102, 149, 150]. However, ownership is typically stronger from a 1PP compared to a 3PP [201, 203, 235]. For what concerns the sense of agency and self-location, both 1PP and 3PP induce the same level of SoE [101]. As we already stated, the sense of ownership and self-location are strongly correlated, and 1PP can consistently increase their level during an embodiment experience [63]. Among others, this is due to the better perception of the arms and the hands of the operators' avatars [101]. Ownership is less likely to occur when the apparent visual location or orientation of a body part conflicts with its real location [196]. Moreover, it was demonstrated that 1PP of a realistic virtual body can induce a strong embodiment illusion even after asynchronous visuo-tactile stimulation [168, 226, 235]. However, 3PP can provide a better awareness of the environment. Therefore, according to the context of application, it can be preferred over 1PP. The SoE can be obtained also over a distant body, seen from 3PP, when synchronous visuotactile information is provided [149, 150] or when the surrogate preserves spatial overlap with the real embodiment [168].

2.3.3.2 Field of view

The field of view is the observable area that the operators see without head or eye movements, directly or via an optical device [84], such as a VR headset. Normally, humans have a slightly over 210-degree forward-facing horizontal arc of their visual field, i.e. without eye movements (with eye movements included it is slightly larger). A high SoE is easier to obtain if the field of view allows for the coverage of an area similar to the human field. This cue can be manipulated in different ways: it can be dependent on the surrogate size or it can be increased or decreased compared to the human field of view (often decreased because of the limited field of view of head mounted displays). The manipulation of this perceptual cue affects the perception of the remote environment and the judgment of the peri-personal and distant space [254]. This can affect the level of sense of ownership and self-location, but it can also affect the sense of agency by creating movement impairments [266].

2.3.3.3 View direction control

This cue refers to the amount of control that operators have in directing their gaze in the remote environment. When the control is absent, imprecise, or not intuitive, this can affect the perception and experience of the embodiment. Different studies demonstrated the impact of the avatar control on the SoE. In [84], the authors tested the impact of three perceptual cues on the SoE. They demonstrated that a close match between the view direction control of the operators' intentions and the subsequent actions, mostly affected the sense of agency, compared to the other two embodiment components. Another example is given in [135], where participants experienced a high sense of agency when walking a virtual body, even though they were seated and only head movements were allowed.

2.3.3.4 Connectedness

The connectedness refers to the perception that the surrogate is attached to the operators' body. The operators perceive the real body as joined to an external object or device, such as a continuum of the their own body. This cue is especially helpful in increasing the joints perception and awareness in the space. Feeling the external device as attached to the real body helps the operators in choosing proper trajectories and better managing movements in the remote environment. In [158], the authors compared the time to accomplish a task in both connected and no-connected conditions. The results showed that the presence of connectedness improved the accomplishing time of the task. In ([200]), the authors investigated the importance of four factors on the SoE in a virtual rubber hand illusion: visuo-tactile synchronicity while stroking the virtual and the real arms, body continuity, alignment between the real and virtual arms, and the distance between them. The results showed that the subjective illusion of ownership over the virtual arm and the time to evoke this illusion are highly dependent on synchronous visuo-tactile stimulation and on connectivity of the virtual arm with the rest of the virtual body.

2.3.3.5 Visuo-proprioceptive synchronicity

The manipulation of proprioceptive information is one of the most used SoE evaluation tasks. Proprioception is based on the information from receptors in muscles and joints (capsules and surrounding tissue). These receptors provide information to the central nervous system about the position and movement of body parts (e.g., the angle of a joint or the length of a muscle). Proprioception is the sense that tells the body where it is located in space. It is very important to the brain as it plays a big role in self-regulation, coordination, posture, body awareness, the ability to attend and focus, and speech. To reach high SoE, it is important that the visual cues that the operator receives reflect the remote device and environment accurately. Therefore, the distances, points and objects in space have to be properly perceived in the remote environment, the perception of joints has to be paired, and the device response to the operator's movement has to be synchronous and, ideally, happen in real time (i.e. with negligible delay) [141, 144, 183].

2.3.3.6 Visuo-tactile synchronicity

The visuo-tactile synchronicity refers to how the operator detects visual and tactile cues (e.g. if the operator touches an object, the touch feedback should be in sync with the visual cues). The asynchronicity between visual and tactile feedback can easily and immediately break the embodiment illusion, providing to the operator an impaired perception of the external body. Usually, the asynchronicity is caused by delay in transmitting and receiving information from the operator to the device and vice versa. The asynchronicity makes the telemanipulation of objects and the interaction with the environment inefficient for the operators, as they need to adapt to a sensation of impaired position of the joints and tactile feedback. This perceptual cue is usually affected by time delay in teleoperation or by system lag in VR environments. There is an extensive literature of studies on the effect of synchronous and asynchronous strokes [16, 71, 81, 115, 160, 183] in a RHI setup or manipulating virtual or robotic limbs. The common finding is that asynchronous stimulation decreases the sense of ownership over the surrogate.

2.3.3.7 Visual likeness of the surrogate

The visual likeness of the surrogate refers to the human-likeness and appearance fidelity of the embodiment w.r.t. the operator who is manipulating it. Generally, the more the device one is controlling is similar to the real body, and the more the actions that the operator accomplishes are mirrored by the device, the higher will be the operator's SoE. For what concerns the sense of ownership, [229] showed that, in VR-based teleoperation, using human-like hands increased the risk perception and degraded workers' task performances in the execution of high-risk tasks. Moreover, some studies point to the importance of taking into consideration the operators' diversity while designing the avatar appearance. For example, [223] showed that female operators were more sensitive to the manipulation of male or female hands while operating an avatar in a VR setup. Female operators felt less embodied while they manipulated male avatar hands. Particularly, this condition negatively affected the sense of ownership. On the other hand, in [176], the authors tested the impact of the human-likeness on the sense of agency and task performance. They showed that this cue has a significant weight just in the initial phase of adaptation with the system. Meaning that, after a first phase of familiarization with the new joints, the operator will no longer perceive a low human-likeness of the surrogate as critical.

2.3.3.8 Haptic feedback

Haptic feedback is about the simulation of physical attributes, such as weight, pressure and stiffness, which allow the operators to interact directly with virtual or remote objects using touch and experiencing the physical attributes of them. Usually, they are reproduced by very small forces or cues (such as vibration), that are mostly only felt through mechanoreceptors in the skins surface. This cue determines a believable perception of and interaction with the remote environment. This is essential from both a practical point of view, in terms of task performance, and from the view of the operator's experience, in terms of the embodiment illusion. This cue is defined by the combination of tactile, kinesthetic and contact feedback, and the presence and magnitude of contact force. Realistic tactile feedback is complex to obtain with current technologies, but it can make a big difference in the task performance, especially in social, field, and surgical scenarios [37]. In the social scenario, it can make the interaction more believable by providing, for example, a proper skin perception [54]. In the field scenario it could help in exploring the unknown remote environment by providing important information about objects, such as the temperature, texture, and stiffness [95]. Finally, in the surgical scenario, it can improve task performance providing information about the texture, shape, and consistency of the internal body parts of interest [24, 221]. The proprioceptive feedback, instead, can provide information about the position and movement of the remote surrogate in the workspace. The presence of force and contact feedback makes the operator aware of the dimensions and shape of both the remote environment and the surrogate, while the magnitude of contact force is necessary to make the operator aware of the mass of the manipulated objects in the remote environment, and also of the power and strength of the surrogate.

2.3.3.9 Likeness of the Environment

There are different ways to present the remote environment to the operator, such as video streaming or by building a virtual environment. To allow the operator to properly interact with the remote environment and to make the experience immersive, the quality of the data transmission and of the environment reproduced has to be high. The highest and easiest to design immersive experience occurs when the operator wears a VR headset. However, there are also other solutions, such as: big screens (cinema effect), placing the cockpit of the surrogate in a silent and isolate room, augmented reality glasses. These cues help in moving through the environment and predicting the interaction effect. Simulated environments should also take the social norms applicable to a real life environment into consideration, if other actors are present in the simulation [267]. In other words, the likeness of the environment is not only affected by the quality of the realization and transmission of the workspace, but also by the way in which other living beings, external to the operator, can interact with it.

2.3.4 Step 3a. Surrogate

We use the term *surrogate* to refer to the device, fake body, or avatar embodied by the operator. On the basis of the application context, the surrogate can represent the entire body or just a part of it. Rubber hands [29, 71] and mannequins [41] are options just in empirical studies that focus on a better understanding of the embodiment components and their relation. VR avatars, instead, can be used for both theoretical studies [63, 235] and for testing the operator's SoE in a commercial teleoperation system (e.g., VR games). In telerobotics, we distinguish three categories of robotic surrogates: 1) *humanoid* robots are used in social scenarios, they resemble the human body and their design is based on the concept of bio-mimicking. The humanoid shape, due to the perception of a familiar appearance from the recipients, facilitates the expression of social cues and, therefore, the interaction. 2) *Industrial* robots are usually used for manufacturing. This category is mostly represented by robotic arms and the main focus, while designing these devices, is the optimization of the task performance. 3) *Explorative* robots can be divided in two sub-categories: a) *open-air* refers to animal-shape robots that are usually used in search & rescue scenarios since their anatomy is effective in hazardous and unpredictable environments; b) *nano-world* refers to robot applied in micro-surgery, anatomical exploration, or exploration in out-of-the-body world, such as pipe or tubes exploration [8].

In section 2.3.1 we described the tasks that these different categories of robots are required to accomplish.

2.3.5 Step 3b. Control device

With control device we mean the physical device, or combinations of devices, used by the operator to control the surrogate. The choice of the control device depends on the context of application and can affect the operator's experience [173]. For example, a joystick might be preferred in an industrial scenario to control a robotic arm, while a sensory glove may be a more adequate option to control a robotic hand in a social context. The control devices are part of the operator interface of the system and they are strongly linked to the perceptual cues (needed to achieve high SoE according to the described Step 2 in Section 2.3.3). In [215] and [52], several system controls and haptic interfaces are discussed with their challenges and opportunities, which could help deciding on a control device. A more updated and complete literature review can be found in [113].

2.3.6 Step 4. Test tasks

Most of the following tasks are suitable both in teleoperation and VR setups.

2.3.6.1 Positioning Objects

This category of tasks requires the operator to place an object in a specific point in the workspace [110, 161, 216]. Some examples are: 1) the peg-in-hole, that is a classic robotics task to test the interaction with a mechanical environment. It is a test of the operator's ability to achieve accurate positioning in spite of nonlinear mechanics and imprecise knowledge. It consists of a test participant grasping a peg with the end-effector and then inserting it into a specified hole. The tolerance of fit for the hole is usually varied, and the associated task time for inserting the peg into the hole is recorded [120, 170]. The 2) *pick and place* consists in grasping an object and then manipulate it (an extra condition is sometimes added, in which it is required to delicately manipulate the object), avoiding obstacles and moving along certain lines to avoid collision, and finally placing it in a target point [1, 104]. The 3) *tube task* (TT) is a test designed to gather measures of ownership and self-location. Participants are instructed to use a control device (such as a joystick) to adjust the size and position of a virtual tube, to match it with the perceived locations of their ankles. During the task no virtual body is displayed and the virtual tube is the only visible object in the environment.

2.3.6.2 Telemanipulation of flexible objects

This task consists of the telemanipulation of non-rigid objects, which require more complex control and force feedback strategies than the ones used to manipulate rigid objects. An example is attaching and detaching velcro fasteners. Its hook and loop fastening system has nonlinear mechanical properties which challenge manipulation capabilities. Tele-shaking hands is another example, the challenge is given by the unpredictable impedance of the dynamic system (in this case the hand of the recipient) to which the operator has to interact [3, 23, 110, 163, 238].

2.3.6.3 Micromanipulation & microassembly

This category of tasks is used to test systems for surgical or precision maintenance purposes. It consists of an operator interface which uses visual, haptic and control devices (macro world), a nano-manipulator, and sensors (nano world), to telemanipulate between macro and nano worlds [20, 44, 230, 231].

2.3.6.4 Tracking a sustained contact force

This set of tasks measures the ability to present information for tracking the magnitude of a force over time. It tests how the operator can manage the force feedback, by dosing the force and properly interpreting the information provided by the sensory feedback. An example is a telemanipulation task where a force must be exerted over a period of time, and has a maximum level above which damage and task failure will occur. These tasks are good to determine the presence and magnitude of contact force [7, 169, 195, 273]. Moreover, they test the operator's awareness of the dimensions, shape, and weight of the surrogate.

2.3.6.5 Changed Workspace

This task demonstrates the ability of the operator to deal with a dynamic workspace. It tests the ability to interact with a hazardous environment and unexpected stimuli. The operator is first trained in a workspace, and then this workspace is changed. For example, a new obstacle is placed right in the path of the learned trajectory and the operator has to circumvent the object [1, 241].

2.3.6.6 Motor imagery task

It consists of asking the operators to experience an embodiment illusion with the surrogate, only by imagining a movement (motor imagery) and watching the device performing it. Several studies demonstrate that the timing and accuracy of the performance feedback could improve operators' modulation of brain activities for the motor imagery task [4, 5, 21, 63]. Therefore, the motor imagery skills acquired through the training have long-lasting effects, which improve the operators' performances especially if they are using Brain Computer Interfaces (BCIs) as a control devices. Operators can explore and operate in a remote environment and train their distance perception and proprioceptive level of information to increase the sense of self-location. An example is the *mental drop ball* (MBD) task [149], in which the participant estimates the time a ball would take to fall down from their hand to the floor. The MDB is meant to address the question of where the self is localized or, more specifically, to detect whether the operator has similar time estimations in 1PP and 3PP. Consistently shorter times in 3PP could indicate a weak sense of self-location or that operators are better at judging distances than depth.

2.3.6.7 Multisensory congruency task

Multisensory congruency tasks are one of the most frequently used to test and evaluate SoE. They consist of a combination of multisensory stimuli, usually visual and tactile, that are presented to the participants in two conditions: synchronous or asynchronous. The operators have to judge their embodiment experience in both conditions. These stimuli are designed or used in order to evaluate or test different combinations of the embodiment components: the three components together [63, 103, 125, 160, 208, 225, 253, 261], the sense of ownership and agency [67, 124, 197, 219, 222, 235, 252], the sense of ownership and self-location [11, 34, 69, 83, 119, 150, 168, 186, 196, 201, 211, 218, 255, 256, 271], or just singularly the sense of ownership [29, 36, 72, 118, 146, 180], the sense of agency [134], or the sense of self-location [68]. The most common task of multisensory discongruency is the RHI setup, in which visual and tactile feedback are asynchronous [29, 103, 196]. These findings demonstrate how central body representation directly influences visual size perception by rescaling the spatial representation of the environment. There are also studies that focus on the importance of haptic information for multisensory integration, reporting that haptic information can be adapted to an illusory different size of the body [34, 225]. The majority of these studies are based on the relation and distinction between body image and body schema [60]. The body image consists of perceptions, attitudes, and beliefs concerning one's body. In contrast, body schema consists of sensory-motor capacities that control movement and posture.

2.3.6.8 Threat task

This is a passive task that tests the body ownership at the time of the threat. If the operator feels affected by the observation of the threat in the remote environment, this indicates the presence of the sense of ownership, which will be probably ended by the threat (since the operator becomes aware of the illusion) [72, 270, 272]. Designing a threat into the experiment (e.g., hitting the surrogate with a hammer or stabbing it with a knife) is the most used test to assess the sense of ownership with a proper physiological measure (such as skin conductance response or heart rate). A peak in the signals recorded by the physiological measure, at the moment of the threat, is considered a proof of embodiment. It is also a consequential measure of the sense of self-location, even if it is not necessarily considered as such in the papers. We claim that if one feels affected by the threat, it implies that the individual feels located and immersed in the remote environment, and therefore also affected by its dynamics.

2.3.6.9 Time delay

For all the previous tasks, imagine an additional condition in which a delay is added. This is a difficulty that can be added to each task, and it can be used to test each component. Just to provide some examples, the delay could be added to test to what extent the presence of a time delay affects the embodiment experience, to test the operator's management level of an unstable control device, and to test how much delay the system control of a device can handle before becoming too unstable and dangerous [169, 227]. The time delay is one of the main issues which are encountered in teleoperation applications, and it can be used to test how its presence can affect the sense of ownership and agency of the operator in different contexts and conditions.

2.3.6.10 Miscellaneous tasks

There are even mixed and alternative stimulation methodologies. In [197], for example, the authors explore whether embodying the errors of an avatar may activate the error monitoring system in the brain of the observer (e.g., looking in 1PP at an avatar who drops an object which it should hold or who does not follow the instructions provided by the experimenter) by seeing it from 1PP. Other studies manipulate the perceived size of the external body or part of it, and in their findings they show that the perceived size of objects is determined by the size and the strength of the body in which the participant feels embodied [186, 255, 256].

2.3.7 Step 5. Explicit measures

We discern two categories of assessment measures: explicit and implicit measures. Explicit measures are based on explicit ratings and reports made by the user and observers. Implicit measures are based on, for instance, performance or physiology.

This section presents explicit measures (questionnaires and self-reports), and Section 2.3.8 describes implicit measures.

2.3.7.1 Questionnaire

Questionnaires are the most widely used explicit measurement, especially because they are adjusted and adapted to all types of experience. Furthermore, questionnaires allow one to focus on specific components of the embodiment.

Focus on the sense of ownership The most well-known and widely used questionnaire, possibly with variations, is from [29]. Participants submit their responses on a seven-step visual analog scale ranging from 'strongly agree' to 'strongly disagree' (eg, 'I felt as if the rubber hand were my hand'), 'It seemed as if the touch I felt came from somewhere between my own hand and the rubber hand'). Even if the items to be evaluated mostly address the sense of ownership, they also cover the sense of self-location, especially for what concerns the proprioceptive awareness of the operator. Similar studies can also be found in [256] and [118]. In the questionnaire presented in [218], the focus is on the sense of ownership, while the sense of self-location is measured through proprioceptive drift. In [201], the authors designed a questionnaire that focuses entirely on the sense of ownership. Participants were asked to complete a questionnaire on which they had to affirm or deny seven possible perceptual effects using a seven-point Likert scale. Three statements were designed to capture the illusory experience of being the artificial body (in this case a mannequin, e.g. "It felt like the mannequin's body was my body"), and the other four served as controls for suggestibility and task-compliance (mostly synchronous or asynchronous stimulation, e.g. "It seemed as the touch I felt was caused by the stick touching the mannequin's body"). In one experiment, participants observed a knife 'cutting' the mannequin's abdomen. There are studies in which the questionnaires previously reported are rephrased, mixed, or adjusted in order to be used in a particular embodiment experience, but with the same scales and question content [34, 150, 197, 245, 248, 255]. For example, [233] proposes a variation of the one presented in [29], in order to make it applicable also to virtual reality scenarios (e.g., "During the experiment, there were moments in which I felt as if the virtual arm was my own arm"). In [235], some of the questions were derived from previous work [29, 69, 150] and others were introduced after interviews with participants in extensive pilot trials. The participants responded to a 13-item questionnaire. Eight of these questions are related to the sense of ownership. The questionnaire scores (between 0 and 10) were recorded in ranges of Very Low (0), Low (1–3), Medium (4– 6), High (7–9) and Very High (10), based on the layout of the questionnaire. In [270], instead, the questionnaire is readapted from [233] in order to be applied to virtual reality scenarios. This questionnaire covers both the sense of ownership and agency. There are also studies in which the authors measure the sense of ownership using questionnaires designed for a different purpose. For example, in [92] participants were asked to complete two questionnaires aimed at measuring trait empathy (that is, the interpersonal reactivity index (IRI) [56] and the empathy for pain scale (EPS) [97]). Finally, other studies try to design a unique embodiment questionnaire in order to standardize this metric. An example is in [100], in which the authors review the questionnaires used in previous user studies and propose a standardized embodiment questionnaire based on 25 questions (the ones prevalent in the literature). The questions can be customized and used in studies involving virtual avatars, mannequins, and robotic devices. Moreover, the authors encourage the administration of this questionnaire in future embodiment experiments (especially in virtual reality scenarios) that include first-person virtual avatars. The main aim of the work was to further investigate the embodiment components and to increase the comparability and standardization of the measurement of embodiment across experiments by providing a standard embodiment questionnaire that is validated and reliable. They confirmed and updated this purpose also in their most recent work [198], in which they presented new topics of discussion on the embodiment components and also an updated questionnaire with a reduced number of questions, from 25 to 16.

Focus on the sense of agency In [55], the authors design a questionnaire that assesses the sense of agency and the sense of self-location; particularly, the participants indicate whether they experience a sense of agency during active agency task conditions and how they perform during the 3PP condition. In [36], we find an example of a questionnaire that focuses on the sense of ownership and agency. Another well-known and validated questionnaire is the one from [160]. The authors assess the three components of the SoE, but with a particular focus on the sense of agency, especially for what concerns the experiment design. In particular, they measure five components: 1) *embodiment*, reflecting feelings that the rubber hand, used as the artificial limb to create the embodiment illusion, belonged to the participant; it comprises three dissociable subcomponents: ownership, agency, and self-locatiagency, *Lossofownhand*, reflecting feelings of being unable to move one's hand; 3) *movement*, relating to the perception motion of one's own hand; 4) *affect*, relating to the experience of the experiment being interesting and enjoyable; 5) *deafference*, which is related to the experience of perceiving the hand less vivid than normal due to asynchronous visuo-tactile stimuli (such as seeing a brush touching the rubber hand but not feeling it on the real hand), which deceives the brain. Participants indicated their agreement or disagreement with 27 statements in each block using a 7-item Likert scale (from 'strongly agreed' to 'strongly disagreed'). The first two items were always related to the experience being interesting and enjoyable (e.g., "I found that experience enjoyable"); the order of subsequent items was randomized separately for each participant in each condition (e.g., "It seemed like I was in control of the rubber hand", "it seemed like my hand was in the location where the rubber hand was"). Regarding the *affect-component* , there is an ongoing debate on the way in which experiment experience can affect the embodiment components. In [14] and [18], the authors demonstrated, in two different experimental contexts, that the level of interest and enjoyability of an experience mainly affect the sense of ownership and then, possibly as a consequence in the long term, the sense of agency.

Focus on the sense of self-location It is not common to focus a questionnaire on the sense of self-location, because often this component is not assessed independently from the other embodimemt components. As previously reported, [55] evaluates the sense of agency and self-location in combination. Another example is in [63], in which the authors design a questionnaire to assess the senses of agency, body ownership, self-location to assess the effect of congruent visuomotor-tactile feedback both in active and passive (the participant is just an observer) conditions and in 1PP and 3PP conditions. It contains 10 questions: two for each component, two for the threat, and two control questions. Questions were formulated based on related experimental protocols [42, 150, 160]. Responses were given on a 7-point Likert scale, ranging from “Strongly Disagree” to “Strongly Agree".

2.3.7.2 Self-report

We define self-reports as participants’ reflections on their experience. It is an introspective report that can be semi-structured or without any kind of guidelines in the participant’s stream of thoughts. This differs from questionnaires, in which participants have to answer questions or evaluate an experience using a rating scale. The literature related to self-reports can be a good starting point to design more specific quantification metrics. The information obtained from self-reports can be interesting and relevant, but difficult to compare among studies and to report outside of an exploratory view. Often, self-report data is reported but not analyzed. Typically, self-reports take into consideration all three components of the embodiment. It is also common to combine questionnaires and self-reports, as in [11], combining free response descriptions of the experience and an intensity rating to determine the degree to which the participants embodied a fake hand. In [69], we can find a combination of questionnaire and self-report. In this case, Ehrsson reports a few sentences from the participants who described their experiences and feelings, but does not use a scale to assess the reported interview. Unlike [152], in which the authors analyzed the interview data using interpretative phenomenological analysis (IPA) [237]. The IPA was selected because of its emphasis on the experiences of the participant and how the participant makes sense of these experiences. The structure of the introspective interview was provided by a series of open questions: ’Do you have any unusual sensations and can you describe them?’, ’How do you feel about the rubber hand at the moment?’, ’How intense is this sensation?’, ’Compared to synchronous stroking, how does asynchronous stroking feel?’. Questions comparing the experience during synchronous stroking and the experience during asynchronous stroking were used to aid introspection and help the participants articulate their experience. The questions were usually presented to the participants in the order reported in the paper; however, the authors claim that the participants were encouraged to report out loud any thoughts as they occurred and this could influence the questions asked and the order in which they were presented.

2.3.8 Step 5. Implicit measures

2.3.8.1 Proprioceptive measures

Using proprioception as a measure of SoE is related to the operator’s awareness and perception of the size and shape of the surrogate in relation to the remote environment. There are three common ways to apply this measure: 1) asking participants to reach a point and then measuring the distance between the target point and the reached one [186]; 2) asking participants if they think that a part of their body is located [160]; or 3) asking participants if they felt that their location and the one of the controlled surrogate were the same [119, 150]. Proprioceptive measures provide useful information about all three components (sense of ownership, sense of agency, and sense of self-location).

2.3.8.2 Reaching-distance judgement

This measure is designed to assess the perception of the peripersonal space, in order to link it with the sense of self-location and also the sense of ownership. In [68], participants were asked to stop a confederate at the distance where they thought they could reach her/him.

2.3.8.3 Heart Rate

Heart rate (HR) is the speed of the heartbeat measured by the number of contractions (beats) of the heart per minute (bpm). HR can be used as a measure of embodiment to observe how much an operator is engaged with the surrogate, the remote environment, and the global embodiment experience (i.e., the extent to which, for instance, anxiety and stress are provided by the external environment). Please note that other factors, such as the physical activity of telemanipulation, are also reflected in changes in HR, and this should be compensated or controlled for. For example, in [92], participants were immersed in a VR scenario and observed a virtual: i) needle penetrating (pain), ii) caress (pleasure), or iii) ball touching (neutral) the hand of an avatar seen from 1PP or 3PP. In [235] they measured HR deceleration in response to a virtual scenario in which a woman slapped a girl, a parameter that has been associated with reports of aversive stress in the context of picture viewing. However, interoception is not always a good index of SoE; its variation could also be related to other factors [35]. In [92, 235], the authors do not state or prove a clear disentanglement among the three components while using HR as a measure of SoE.

2.3.8.4 Skin Conductance Response

The skin conductance response (SCR) is the phenomenon in which the skin momentarily becomes a better conductor of electricity when external or internal stimuli occur that are physiologically arousing. In [11], the authors report that a threat to a rubber hand in the RHI caused a skin conductance response (arousal in response to the expectation of pain) in the synchronous, but not in the asynchronous condition. Also in [103], the authors use SCR as a measure of autonomic nervous system (ANS) activity to quantify the experience of agency and ownership on a virtual hand. The ANS is the primary mechanism that regulates involuntarily physiological states, such as arousal due to anticipation of pain or fear. Participants who experienced the illusion show a marked increase in SCR. In [69], the author registered the SCR as a measure of the emotional response when the illusory body was 'hurt' by hitting it with a hammer after a period of synchronous or asynchronous stimulation. Several studies decided to apply the SCR as a physiological measure of embodiment (for all three components, but with particular attention to the sense of ownership) [63, 91, 92, 201, 211, 256, 270], and in order to maximize the measure they always insert a threat at the end of the embodiment illusion.

2.3.8.5 Skin Temperature

The skin temperature (ST) variation is a result of a physiological reaction of the body to a stressful situation. It is often used as an embodiment measure to observe the unconscious body reaction to the embodiment illusion, in both virtual and physical conditions. However, the literature presents contrasting opinions on the efficiency of this assessment, especially because there is no standard way to interpret the results and the replication of similar studies produces inconsistent results. For example, some studies report that any change in temperature is a proof of SoE. In [118], the ST of the participants was recorded to assess how changes in ST are related to the presence or absence of the embodiment illusion under different conditions. In [244], the authors investigate whether the SoE over a virtual hand is reflected in changes in the physiological mechanism of ST regulation and whether ST is modulated by the visual appearance of the virtual limb. This study focuses on both the sense of agency and the sense of ownership. However, it is difficult to state that ST addresses a specific component of SoE. Some studies are more specific on the kind of temperature change, and report that the sense of ownership is active if the temperature of the real limb decreases during the embodiment experience. In [180], the authors hypothesize that ST in a specific limb can be changed psychologically by causing the sense of ownership of that limb. By using an established protocol to induce the RHI, they demonstrated that ST of the real hand decreases when they take ownership of an artificial counterpart. Furthermore, they showed that the decrease in ST is limb-specific: It does not occur in the unstimulated hand. Also in [50], the authors explore the relationship between body ownership, thermoregulation, and thermal sensitivity in a mirror box illusion paradigm. Results showed a decrease in the ST of the hand, following the induction of the illusion of ownership towards the participant's reflected hand. Other studies point out the inconsistency of using the ST changes as a measure of the embodiment. In [57], the authors conducted several studies in which they recorded the temperature of the hands during an RHI under different circumstances, including continuous temperature measurements in a temperature-controlled room. They covered five attempts to replicate the traditional RHI experiment. The results did not show a reliable cooling of the real hand during the RHI. Therefore, they stated that hand cooling in the RHI is not causally related to changes in body ownership. [213] replicated the classical RHI, by inducing cooling of the stimulated hand using an automated stroking paradigm, where stimulation was performed with a robot arm. After they found no evidence for hand cooling in two experiments using this automated procedure, they tried a manual stroking paradigm, which is closer to the one applied in the original RHI. With this procedure, they observed a relative cooling of the stimulated hand in both the experimental and control conditions. The subjective experience of ownership, as rated by the participants in the questionnaire, was strictly related only to synchronous stroking in all three experiments, implying that hand cooling is not a strict correlate of subjective feeling of hand ownership in RHI.

2.3.8.6 Reaction time

Reaction time (RT) is a measure of how quickly an operator can respond to a particular stimulus. This measure can only be applied to certain types of task-oriented user study because it is strictly task related. A classical example of reaction time (RT) as a measure of embodiment, particularly of the sense of ownership, can be found in [196]. Participants had to recognize the position of vibro-tactile stimuli while ignoring the incongruent visual feedback (in this case, distractor lights). RT was compared between congruent and incongruent stimuli to examine response conflict. Instead, a study with a focus on task switching is the one from [268]. The authors investigated the relationship between RHI and higher cognitive functions by experimentally testing task switching by measuring RT. Task switching involves the ability to unconsciously shift attention between one task and another; therefore, the required attention span is high, and the RT becomes a valid assessment. A more unusual application of this measure, in the more peculiar form of onset time and temporal dynamics in general, can be found in [126, 146], where the authors try to detect the sense of disownership from the real hand in an RHI setup. In [103], the focus is on measuring the sense of agency, even if the other components are involved. In this work, the authors measure the reaction time by playing with the synchronicity of visuo-tactile stimuli. In [146, 196] the attention is also on the sense of self-location.

2.3.8.7 Neural activity

Recording and measuring neural activities can provide insight and evidence of SoE experience. In a less invasive way, it is possible to record electrical impulses in the brain using an electroencephalogram (EEG) [197]. It was also possible to observe that several areas of the brain are involved in SoE, thanks to functional magnetic resonance imaging (fMRI), positron emission tomography (PET) and direct stimulation through transcranial magnetic stimulation (TMS) [162, 190, 224]. These measures are expensive to design and invasive for operators; therefore, they present many constraints in designing the setup, the user study and the tasks. They also require the use of specific materials for the control device and the haptic devices and limit the action space of the operator. Several studies attempted to investigate the brain mechanisms involved in the sense of ownership over a surrogate. Usually, ownership is manipulated by using a perceptual illusion, such as RHI [71, 72, 83, 157, 204, 271]. In [156], the main focus is on the sense of self-location. The authors test the awareness of the body position in space and how the brain model is updated during an embodiment experience.

2.3.9 Assessment measures compatibility and discussion

The possibility to clearly and effectively disentangle the embodiment components is still an open debate, however, the tasks are usually designed to test them all together. Often, the task design is too generic: it does not allow for a clear distinction between

	Measurement type	Category	Sub-Categories	References
Sense of Embodiment	Explicit	Questionnaire	Physical scenario	Botvinick et Cohen (1998), , Hohwy at Paton (2010)
			VR scenario	Gonzalez et al. (2018), Slater et al. (2010)
			Illusion by image manipulation	Ehrsson (2007), Van Der Hoort et al. (2010)
		Self-Report	→	Armel et Ramachandran (2003), Ehrsson (2007), Lewis et Lloyd (2010)
	Implicit	Proprioceptive Drift	Target vs Reached point	Lenggenhager et al. (2007), Kammers et al (2009)
			Body parts awareness	Lewis et al. (2010)
			Real body parts vs embodied device position	Normand et al. (2011), Hoover et al. (2016)
		Heart Rate	Physical scenario	Riemer et al. (2014), Fusaro et al. (2016)
			VR scenario	Slater et al. (2010)
		Neural Activity	EEG	Pavone et al. (2016)
			fMRI	Ogawa et Inui (2007)
			rTMS	MacDonald et al. (2003)
		Reaction Time	Response conflict	Pavani et al. (2000)
			Task conflict	Lane et al. (2017)
			Task switching	Yeh (2017)
		Skin Temperature	Limb specific	Moseley et al. (2008)
			Physical scenario	Hohwy et Paton (2010)
			VR scenario	Tieri et al. (2017)
		Skin Conductance Response	Stress and Emotional Response	Armel et Ramachandran (2003), Ehrsson (2007), Yuan et al. (2010)
			VR scenario	Grechuta et al. (2017)
		Reaching Distance Judgement	→	D'Angelo et al. (2017)

Figure 2.2: The table reports a further sub-categorization for each assessment measure presented in Sections 2.3.7 and 2.3.8

.

the components, and therefore their assessment. The importance of disentangling the embodiment components arises when there is a need to singularly improve them in a system or an embodiment experience. If it is unclear how to address a specific component, it is not possible to understand how and which conditions and parameters have to be manipulated in order to change the level of a specific component. Sometimes, authors state that a particular measure was used to assess a specific embodiment component, but, considering the reported experiment design and the tasks, it is hard to disentangle the assessment among all the SoE components. For example, self-location is almost never tested individually, but it is usually indirectly tested when the sense of ownership is assessed. Moreover, the sense of self-location is rarely tested in big or open spaces, but mostly in the context of the peripersonal space of the surrogate.

Table 2.2 summarizes the compatibility between assessment measures and each embodiment component: 1) when a measure can assess a specific component, 2) when it is nonspecific, and 3) when it is not suitable for a specific component.

Assessment Measure	O	A	L
Questionnaire	+	+	+
Self-Report	+	+	+
Proprioceptive Drift	#	#	#
Reaching-Distance Judgment	#	-	#
Heart Rate	#	#	#
Skin Conductance Response	#	#	#
Skin Temperature	+	-	-
Reaction Time	+	+	+
Neural Activity	#	#	#

Table 2.2: We categorize the assessment measures and we group the SoE components that they can assess on the basis of the reported literature review. Legend: O → Sense of Ownership; A → Sense of Agency; L → Sense of Self-Location; - → not suited to assess this component; # → possible but non-specific (i.e. assesses multiple components); + → specific to measure this component.

The *questionnaire* and *self-report* can address each component specifically, since they can be customized. They can be applied to every kind of user study and context, since they are versatile and easily editable. Moreover, especially for questionnaires, there is a vast body of literature to support them. An advantage of self-reports is the possibility of obtaining insights of the operator experience and paying attention to unexpected aspects suggested by participants. Among the disadvantages of the two measures, there is a need to be supported by implicit measures, and they are usually time-consuming for the participants. Another drawback of self-reports is the lack of a unique and comparable way to evaluate them. For what concerns implicit measures, the *proprioceptive measures* are easy to assess and provide a good indication of the operator's perception of the surrogate. Among the disadvantages, it can be time consuming to first collect and measure a baseline before testing the embodiment experience. Moreover, it is hard to disentangle the assessment of each embodiment component, since the localization bias of the real body toward the surrogate can be an effect of all the three components. Finally, the operators' performance could be affected by some aspects of the experiment or setup design, such as the local and remote environment, sensory feedback, the surrogate or other factors involved in the design of the system or the embodiment experience. The *reaching distance judgment* is a good measure of the sense of self-location, but it is difficult to distinguish it from the effect of the sense of ownership. The distance is judged on the basis of the perception of the surrogate in the space. It is not a good measure for the sense of agency, since it does not allow one to directly assess it, but just to have an idea of how the operator would interact with the environment. Therefore, it cannot be considered a measure of the sense of agency but just a method to build predictions on the interaction between the operator and the environment. Among the advantages, this measure is also versatile and easy to measure, as it does not require expensive equipment and its application is not time-consuming. The *heart rate* and the *skin conductance response* are the most widely used implicit measures of SoE. They are easy to measure, but the results could be affected by each component of the embodiment. A high level of stress could be strongly related to the sense of ownership, but being immersed in the remote environment (sense of self-location) and performing an active task (sense of agency), which requires a high span of attention, could equally influence the results, without the possibility of understanding to what extent each component contributed to the embodiment experience of the operator. Currently, pupil dilation is studied compared to SCR and HR [78], since these three measures, apart from SoE evaluation, are used in similar applications and for similar purposes. They aim at reflecting the functions and reactions of the human body to the change of the outside environment and the inside of the body (e.g., that is why they correlate with cognitive effort, which can be a purely internal process). Using a combination of physiological approaches to measure human behavior, not specifically SoE, is not novel [106, 263], and neither compares nor investigates what information they can provide and how accurately [116]. However, this measure is still novel and needs more validation. *Skin temperature* is generally measured in a specific limb and is used as a measure of sense of ownership [180]. In the current literature, ST is not used as a measure of the other two components. There is an open debate in the scientific community about

the use of this assessment measure. Some studies did not replicate previous findings on temperature changes as a consequence of the illusion of embodiment [51, 213]. Moreover, it is still unclear if 1) the skin temperature is a measure of the stress levels and if 2) the temperature variation should be considered locally (on the body part) or globally (on the entire body).
The *reaction time* allows to design dedicated tasks to assess each component specifically. For example, a task that involves cognitive shift would allow one to assess the sense of agency by measuring the reaction time. The sense of self-location could be assessed by measuring the reaction time to a change in workspace task. Finally, the measure of *neural activities* is versatile and can be used as a measure of all the components of the embodiment and the experience of the operator. There is a vast body of literature to support it, especially noninvasive techniques. Moreover, it can provide interesting and unique insights into the SoE and embodiment experience. Among the cons, both non-invasive and invasive techniques create lots of restrictions for what concerns the setup design. The use of this category of measures can impose limitations, such as the impossibility to use certain kinds of material and the reduced action space for the operator. At the same time, this is also the category that has until now provided the most interesting insight and explanation of SoE, because it provides direct feedback from the brain. What emerged from this survey is that the most common and effective way to measure embodiment is a combination of one explicit measure and one or more implicit ones. In our opinion, this combination is the necessary basis for any good evaluation framework, since conscious perception and unconscious processes do not always coincide. What also arises is that there is a lack of a standard and common definition of SoE, making it hard to reuse the current SoE design and assessments. This leads to a vague SoE assessment. For what concerns the physiological measures, for example, each kind of variation from the standard signal is considered a proof of the SoE, without taking into account the context and conditions in which an embodiment experience is realized. As a final consideration, the initial assumption that achieving a high level of SoE can improve task performance is not always correct: SoE and teleoperation are positively correlated in some, but not all, application scenarios and tasks. Indeed, if we think about industrial scenarios and the kind of classical tasks that the operators have to achieve in this context of application (such as maintenance in unhandy or inaccessible environments or moving heavy objects), operators may prefer and perform better using a joystick as control device to teleoperate an industrial arm with a gripper attached to it than a sensory glove to teleoperate a humanoid arm and hand. This is because a humanoid hand would not be as effective as a gripper in carrying or moving heavy objects. Moreover, in this context, the main focus is on the task performance improvement which cannot be achieved in this case by a humanoid surrogate considering the current technology.

2.4 An Overview of the Toolbox

After having explained in detail the steps of the SoE toolbox, this section provides a global understanding of it.

We distinguish three components of SoE: the sense of ownership, the sense of agency, and the sense of self-location. As also depicted in Figure 2.1, the first step is to define the application scenario (step 1) and the embodiment components involved (step 2a). The components determine the relevant perceptual cues as listed in Table 2.1. These cues are key points for the setup and system designer to focus on while designing an embodiment experience (Step 2b). Certainly, the presence of overlapping embodiment components influenced by the manipulation of specific perceptual cues introduces both convenience and complexity to testing. While this overlap simplifies testing procedures, it also adds a layer of difficulty to precisely isolating and assessing the impact on individual components. This difficulty can be addressed by using accurate and customized qualitative measures that target a specific component (for example, with dedicated questions in a survey); while it usually persists with quantitative measures, unless a task is perfectly designed to address a specific component (for example, introducing a threat into the experiment to measure the sense of ownership). The next step is to define the embodiment setup (Step 3). The surrogate that can complete the application scenario (as determined in step 1) and the control device that can deliver the key perceptual cues (as determined in step 2b) have to be chosen. The next step (Step 4) is to choose a task, or a combination of tasks, to test the system according to the previous choices. Depending on the context of use, some tasks will be more recommended than others. Finally, in step 5, the assessment measures and their customization are selected. We make a distinction between explicit and implicit assessment measures. The explicit measures have to do with what people say they actually experience, such as questionnaires and self-reports. Implicit measures include task performance (irrespective of the experience during task execution) and physiological measures. In accordance with [128], we suggest using a combination of explicit and implicit measures to have a complete overview into the operator experience.

Now we present two examples of an application of the five steps outlined above.

2.4.1 Example 1: hugging in a computer game

Step 1. Let us say that we want to introduce the possibility of giving and receiving hugs in a virtual reality game (step 1: application scenario).

Step 2a. Since we will operate in a social scenario, we want to achieve the best level of all the embodiment components: sense of ownership, sense of agency, and sense of self-location. This information can be deduced from the description of the social scenario provided in Section 2.3.1.

Step 2b. It means that we want to optimize the perceptual cues that affect the embodiment components addressed in this situation (in this case, all of them). These cues can be taken from Table 2.1.

Step 3a. To do that, it is important to choose the proper surrogate (Step 3.1: surrogate). Since the application will be a virtual game, we can design a virtual humanoid avatar.

Step 3b. We need to design the operator interface. We will immerse the user in a VR scenario, using a VR headset with 3D vision and audition, and a haptic suit that covers the upper body of the operator so we can optimize the perceptual cues listed under 2b.

Step 4. The best tasks to test our system are in the categories of sustain contact force (to test the hugging experience) and threat (to test the SoE). The guidelines to select the tasks are presented in Section 2.3.6.

Step 5. The most suitable measures to assess the system are: 1) explicit measures: a questionnaire on the SoE and telepresence, and a structured self-report to have more insights on the experience (e.g., what still misses to be believable, which are the differences with respect to a real hug); 2) implicit measures: skin conductance response or heart rate, to measure the level of stress and the emotional state to the threat task chosen in Step 4, and the proprioceptive drift to measure the awareness of distances and space.

2.4.2 Example 2: maintenance in a hostile environment

Step 1. We want to design a system to allow the operator to perform maintenance in an industrial scenario, particularly moving blocks of different weights in an environment in which the temperature is too high to be tolerated by a human being.

Step 2a. The scenario is industrial and it involves the manipulation of objects in space; therefore, the focus will be on the sense of agency and self-location (step 2a: embodiment components involved).

Step 2b. The perceptual cues that affect those embodiment components should be optimized (Step 2b: perceptual cues to optimize). We need to give greater priority to optimizing task performance than SoE. Haptic feedback is important for task execution because the operator has to distinguish among the different weights of the blocks. In this case, the temperature and texture of the manipulated objects or the environment are less relevant. However, it could be useful to alert the operator with visual or tactile feedback if the temperature is high at the level that can damage the surrogate or the manipulated objects. Force and tactile feedback provide information that helps the operator to accomplish the task.

Step 3a. The best surrogate will be a robotic arm capable of handling the weight of blocks and high temperature (step 3.1: surrogate). Moreover, the surrogate will need to have enough degrees of movement to reach all the necessary points in the workspace. In this case, we will choose a gripper attached to an industrial robotic arm. This is because we do not need an end-effector to accomplish precision tasks or with multiple degrees of freedom. The operator will have to always perform the same simple movements, which will be differently combined with respect to the situation. Therefore, a setup that supports the operator in accomplishing the allowed movements, by providing a complete visual overview of the workspace, will reduce the risk of making mistakes.

Step 3b. As a control device, a force feedback joystick would be one of the best comfortable choices to perform in this environment (step 3.2: haptic device), since the movement would be more intuitive and easier to learn.

Step 4. Positioning objects and changed workspace are three categories of tasks that could be useful to test SoE (step 4: tasks). The first tests the dexterity of the operator and the second tests how the operator faces the dynamic environment and the awareness of the surrogate in it.

Step 5. The best assessment would be: 1) explicit measure: questionnaire on the user experience (efficiency, effectiveness, ease of use of the system); 2) implicit measure: proprioceptive drift, reaching distance judgment, and task performance (number of errors, time needed to accomplish the task, number of moves needed to accomplish the task) (step 5: assessment measures).

2.5 Conclusions and recommendations

We present a toolbox to assess SoE, starting with a review of the literature on the most frequently used assessment measures, test tasks, perceptual cues and application scenarios, with particular attention to teleoperation, VR setups and embodiment experiences. Our conclusion consists of the following considerations and recommendations:

- We miss a standard definition of the SoE and a clear picture of what we would like to assess while testing it. In this chapter, we try to integrate the previous well-known definitions into a unique one, highlighting, with respect to the previous definitions, the distinction between subjective and objective aspects of SoE, and underlining the importance of the role of the embodiment components;

- Authors often create a strict dependency between the task and the measure. This makes searching and designing a standard measure for SoE difficult;

- Tasks are often designed to test the three SoE components at the same time, even if it is not explicitly stated by the authors. Our recommendation is to have clearly in mind which embodiment components one wants to address before designing or assessing an embodiment experience. This will make the assessment more reliable and valid;

- The most used explicit measures are questionnaires. In [100, 198], the authors present a good starting point for a standardized questionnaire. It is useful to define specific rules to customize the questionnaire to a particular application scenario and embodiment experience, defining how to choose a subset of questions (e.g., this could be useful in long and repetitive experiments, in which 25 questions would be too many). [198], with the new validation studies, addresses some of these points;

- The most used physiological measures are the skin related ones and the heart rate. However, they can become unreliable if, while designing an embodiment experience, unrelated factors to SoE (which can affect the measurements anyway) are not taken into account. Therefore, even if these are the most frequently used implicit measures, it is important to choose the proper assessment based on the setup and system requirements, the task and the application scenario;

- It is hard to disentangle the assessment of the SoE components using physiological measures;

- We suggest measuring the SoE using a combination of explicit and implicit measures. The first would measure the conscious embodiment experience of the individuals, while the second would measure the intrinsic and unconscious changes in the SoE levels.

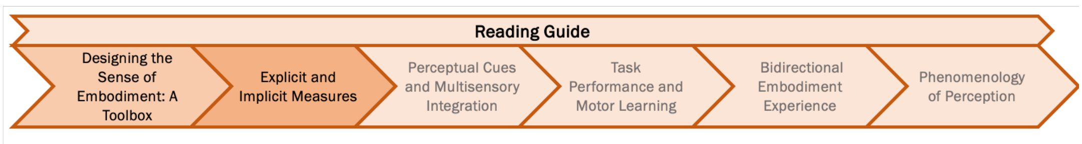
This study guides the reader in choosing the appropriate tasks and measures to evaluate SoE. We also present and underline the reasons that led us to perform this review, namely, the lack of a standard assessment framework. Moreover, this chapter aims to provide the first complete SoE toolbox that can guide the application of existing measures and tasks. A SoE design toolbox will help to define a clear idea of what researchers can and want to assess, test, and obtain from their SoE studies and setup. To facilitate the SoE design, we also define Tables 2.1 and 2.2. However, further work and investigations are required to confirm the information that we reported and structured in the tables. The aim of the proposed toolbox would even be to help the research community to compare the different studies and, ideally, to create a predictive model of the level of SoE in a task-oriented system before its implementation. This predictive model, object of our current investigations, would become a starting point for the improvement and optimization of new teleoperation systems, VR setups, and, more generally, embodiment experiences.

3

Measuring Sense of Embodiment: Implicit and Explicit Approaches

No beings, except men, marvel at their own existence; for all animals it is something that is intuited for itself, nobody pays attention to it.

Arthur Schopenhauer



Abstract

In the realm of assessing the Sense of Embodiment (SoE), measurement approaches can be categorized into explicit and implicit measures. Explicit methods involve self-reports and standardized questionnaires, attempting to capture specific SoE components, but they are subject to user biases and language influences. On the other hand, implicit measures, such as Heart Rate (HR) and Skin Conductance Response (SCR), gauge the body's physiological responses to stimuli. Combining both explicit and implicit measures is recommended for a comprehensive SoE assessment. In this Chapter, we embarked on a comprehensive exploration of pupil diameter (PD) as a potentially more consistent and informative implicit measure of embodiment, and even more suitable for teleoperation setups. As baseline, in user study 1, we test pupil diameter in a rubber hand illusion designed in an augmented reality setup, in comparison with SCR and HR. After the first study, we hypothesize a direct and indirect effect of the SoE on pupil dilation. In the presence of emotional stimuli, we hypothesize that when

a robust SoE is established, there will be a discernible dilation of the pupil. This dilation is attributed to heightened engagement and arousal induced by the emotional stimuli. Therefore, we design a second user study, using the same setup as user study 1, to test this direct effect of the SoE on PD through a rubber hand illusion involving emotional stimuli. To assess the indirect effect of SoE, in user study 3, we employ pupil diameter as a measure in a telerobotics setup. In scenarios devoid of emotional stimuli, we hypothesize that if the operator is immersed in a sense of embodiment, the cognitive workload will be low. This, we anticipate, would manifest as an invariant dilation of the pupil. Even if the results were not completely coherent among the three studies, this multi-faceted approach not only expands our understanding of the direct and indirect effects of SoE on pupil diameter, but also contributes to the broader discourse on implicit measures of embodiment. By triangulating pupil diameter with established physiological indicators and applying it to diverse scenarios, our study seeks to provide valuable insights into the intricate dynamics of embodiment, emotion, and cognition.

3.1 Pupil Diameter as Implicit Measure to Estimate Sense of Embodiment

In the section 2.5 of Chapter 2, we reported an overview on the SoE measures and their advantages and disadvantages. Generally, the measurement of SoE can be divided into explicit and implicit measures. Explicit measures include self-reports and standardized questionnaires (e.g., [198]). Implicit measures refer to the body response to certain stimuli and include Heart Rate (HR) and Skin Conductance Response (SCR) [72]). Although explicit approaches try to address specific components of SoE, implicit measures may not exclusively reflect a specific SoE component. However, explicit approaches are subjective measures and depend on different factors of the user experience, certain biases, and language (e.g., see [123]). Therefore, it is recommended to use a combination of explicit and implicit measures to assess SoE. The most used implicit measures are HR and SCR. However, the field may benefit from exploring other physiological measures to assess SoE, which are more consistent, informative, and easier to apply in a teleoperation setup. This is why in this chapter we attempt to validate the dilation of the pupil diameter (PD) as a measure of the SoE.

PD, heart rate (HR), and SCR are physiological measures that reflect the functions, behaviors, and reactions of the human body to changes in both the outside environment and inside the body itself. These physiological measures are known to reflect the emotional and cognitive state [106, 263], and to compare the type and accuracy of information they can provide [116]. PD is considered to reflect autonomic arousal raised by emotional stimuli in an individual [193, 194]. It is also a good measure of cognitive workload [109, 178], and may even be more sensitive than HR and SCR [116]. An increased diameter is usually associated with an increased degree of difficulty in a task.

SCR and HR are currently the main assessments of SoE. We are interested in exploring whether and how PD is correlated with SoE. When individuals feel strongly embodied, stimulating the surrogate will produce the same effect that could be detected while stimulating their own body (such as arousal, emotional changes, and cognitive workload). PD reflects uncertainty, surprise, and reflects reward prediction errors [246]. Stimulation of the sympathetic branch of the autonomic nervous system induces pupil dilation, whereas stimulation of the parasympathetic system causes constriction. We assumed that during an embodiment experience that stimulates SoE, the pupil will be restricted on average (unless a stressful event happens or is introduced in the embodiment experience). In case of an embodiment experience with low SoE, the individual will feel uncomfortable and stressed in embodying the surrogate. This will stimulate the sympathetic system and cause pupil dilation.

We report three user studies in which we collected the PD as a potential measure of SoE. PD was always collected in combination with other measures (either explicit, implicit, or both). We investigate a direct and indirect effect of the SoE on PD through two main hypotheses:
H1) We expect a positive correlation between PD and SoE in the presence of emotional stimuli such as a threat to the surrogate. We expect higher arousal when participants are more embodied and thus a larger PD.
H2) We expect a negative correlation between PD and SoE in the absence of emotional stimuli. This is caused by the indirect effect of workload. Higher SoE will reduce workload, and lower workload results in smaller PD. For conflicting sensory cues, we expect a lower SoE, a higher workload, and thus a larger PD.

3.2 User Study 1

This first study served as a baseline to test the potential of PD as implicit measure of the SoE, compared with the two most established physiological measures of embodiment: SCR and HR.

3.2.1 Method

We wanted to validate PD as an implicit approach to measuring SoE. To do that, we designed an embodiment experience in which we focused on the manipulation of the sense of ownership and sense of self-location in a between-group design with two conditions: one group experienced visuo-tactile synchronous stimuli (embodied condition), while the other experienced asynchronous stimuli (not-embodied).

3.2.1.1 Participants

33 right-handed participants (between 20 and 34 years old, 16 females and 17 males) were recruited from the student body of the University of Twente, with 5 euros as a raffle. Participants were divided into two groups: 17 participants experienced the embodied condition, while 16 participants the non-embodied one. The ethics committee of the University of Twente approved the study (RP 2021-111).

3.2.1.2 Setup and Procedure

The experiment lasted 20 minutes. The embodiment experience was pre-recorded and participants experienced first person perspective (1PP) of a surrogate (a confederate). They were asked to read and sign the consent form. Then, they were asked to sit in a rest position, putting their right hand on the table as indicated by markers, and the left hand on their left leg. We asked to look at their right hand on the table for the duration of the embodiment experience. Participants were asked to wear a latex blue glove on their right hand. A white tissue was used to cover the participant's right wrist. In this way, they would not focus on the different features of their skin or clothes compared to the surrogate hand (we also placed a white tissue on the right wrist of the surrogate). The participants were primed with a cup of water that was placed both next to the participants and also next to the surrogate hand. Then, they were asked to wear the Empatica E4 wristband on the left wrist, which was used to collect SCR and HR. Finally, they put on the HTC VIVE Pro Eye that was used to collect PD. After the eye tracker calibration, the video was displayed on the HMD and the experiment started. The surrogate was displayed in the same room and in the same position as the participants. Participants observed stimuli administered to the surrogate's hand while synchronous or asynchronous (depending on the condition) stimuli were administered to their own hand.

3.2.1.3 Stimuli

The experiment consisted of four different stimulation phases focused on the right hand (see Figure 3.1): 1) a pen crossing the lunate, 2) a pen alternatively touching the trapezium and the lunate, 3) the experimenter finger touching each fingertip of the participant, and 4) a threat to break the embodiment illusion, namely the experimenter grabs the cup of water and pours it only on the surrogate hand.

3.2.1.4 Measures

For the first time, to the knowledge of the authors, PD was used as a psychophysiological measure of SoE. As a baseline to compare the novel measure, we recorded SCR and HR as additional implicit measures. Furthermore, we used a reduced version (two items for each subscale: appearance, multisensory, response, embodiment, ownership) of the SoE questionnaire from [198], as an explicit measure. Participants evaluated each item using a Likert Scale from *1* (*completely disagree*) to *7* (*completely agree*).

3.2.2 Results

For PD analysis, we applied a Hampel filter to re- move the outliers, we used the convergence to cover missing data, and we considered the mean of the left and right pupils. A two samples t-test did not report a significant difference between groups for the three physiological measures (PD, SCR and HR). We did find a significant difference on all subscales of the questionnaire between the group experiencing supportive and suppressive conditions (Appearance t_{31} = 5,57, $p < .001$, M supportive = 5.48,

Figure 3.1: On the left, the experimenter's perspective during the experiment. The four pictures on the right represents the participant's view during the embodiment experience and, in order, the four phases of stimulation.

M suppressive = 3.80; Multisensory t_{31} = 6,43, $p < .001$, M supportive = 5.93, M suppressive = 3.76; Response t_{31} = 4,03, $p < .001$, M supportive = 4.85, M suppressive = 3.51; Embodiment t_{31} = 6,23, $p < .001$, M supportive = 5.53, M suppressive = 3.78; Ownership t_{31} = 6,19, $p < .001$, M supportive = 5.86, M suppressive = 4.03). For the suppressive group, an independent t-test reported a significant difference between the mean of PD during the embodiment experience (M = 4.68mm) and the threat (M = 4.57mm)(t_{15} = 3.22, p = .006). No effects on HR and SCR were found in this group. For the supportive condition, we found a significant difference between the mean of HR during the embodiment experience (M = 77,41bpm) and the threat (M = 74,43)(t_{15} = 3.21, p = .006). No effects of SCR and PD were found in this group.

3.3 User Study 2

The outcomes of the initial study were inconclusive, yet they sparked our curiosity. Consequently, we designed a second user study, replicating the setup from the first study, to investigate the hypothesized direct impact of Sense of Embodiment (SoE) on Pupil Dilation (PD), through a Rubber Hand Illusion. This study specifically involved the introduction of emotional stimuli to examine the intricate dynamics at play, and it also explored the effect of individual differences on the SoE.

3.3.1 Method

We investigated, in a between group design, if individuals with high kinesthetic intelligence (experimental group) are more resilient to feeling embodied with a surrogate compared to individuals with average kinesthetic intelligence (control group) [78]. We identified the experimental group in dancers and gymnasts who practice the discipline at a competitive level.

3.3.1.1 Participants

16 of 26 right-handed participants (between 20 and 37 years old, 9 women and 17 men) were sampled from the staff and student body (control group), while 10 were sampled from dance / gymnast associations (experimental group). The ethics committee of the University of Twente approved the study (RP 2020-132).

3.3.1.2 Setup and Procedure

The experiment lasted 20 minutes. Participants were asked to read and sign the consent form. Then, they were instructed to take a rest position (the same described in user study 1), by placing their right hand on the table as indicated by markers, and the left hand on their left leg. We asked them to look at their right hand for the duration of the embodiment experience. Participants wore a blue latex glove on the right hand and an HTC VIVE Pro head-mounted headset (HMD). The HMD has an integrated eye tracker, which we used to measure and record PD. After the eye tracker calibration, a video was displayed in the HMD and the experiment started. The video consisted of a pre-recorded embodiment experience, in which participants experienced the first person perspective (1PP) of a surrogate (in this case a confederate), in the same room and position in which the participants were. The video was recorded using a ZED mini stereo camera. Participants observed stimuli and tasks administered to the surrogate's hand, while the same stimuli and tasks were administered to their real hand. The video and HMD data were managed and collected with the Unity platform and game engine.

3.3.1.3 Stimuli and Tasks

The stimuli and the tasks were administered in the following order: 1) cross-modal congruency task: we designed two variants (see Figure 3.2): i) in the first case, participants watched the tip of a pen touching the top of their right hand. Only three times out of six the participants' hand was actually touched by the pen. ii) Participants saw the tip of a pen and a brush alternatively touching the hand displayed in the video. Randomizing the order of the stimuli, only three times out of six the visuo-tactile information was congruent (i.e. the participant saw and was touched by the same object). 2) Linking dots: It was designed to test the sense of agency and self-location of the participants in an active task. We realized two drawings with numbers inside black dots and a small red dot in the center of all of them. We placed a tablet with the drawing in the same position as the one that they watched in the video. They were asked to link the dots in ascending order, sliding their right index on the tablet from one dot to the other. The experimenter had the role of placing the finger in the proper initial position and telling the participants when to move from one dot to the next. 3) Threat: it is common to threaten the surrogate to break the illusion of embodiment and to assess the level of SoE through physiological measures [72, 270, 272]. We primed the participants with scissors placed both on the real table and on the one displayed in the video. The participant watched the scissors being grabbed and then used to hit the table next to the surrogate hand.

3.3.1.4 Measure

We collected PD, as an implicit measure, in combination with the SoE questionnaire (the same used is User Study 1) from [198], as explicit measure.

3.3.2 Results

For PD analysis, we applied a Hampel filter to re- move the outliers, we used the convergence to cover missing data, and we considered the mean of the left and right pupils. A t-test of two samples indicated that there was no significantly lower mean of the PD in the control group, which was expected to experience a higher SoE (M = 3.87mm), than the experimental group (M = 3.85mm) (t_{20} = 0.47, p = .642). The questionnaire responses also did not report significantly higher mean scores for either of the three embodiment components. However, within the control group, an independent t-test reported a significantly larger mean of the PD at the moment of threat, when participants were expected to experience a higher SoE (M = 3.69 mm) than during the first part of the embodiment illusion (M = 3.87mm) (t_{14} = 4.52, p< .001), i.e. a positive correlation between SoE and PD in the presence of an emotional stimulus. However, for the experimental group, we did not find a significant difference in pupil size between threat (M = 3.98mm) and the first part of the embodiment illusion (M = 3.85mm)(t_6 = 1.36, p = .223).

3.4 User Study 3

Finally, we examined the hypothesized indirect effect of Sense of the SoE on PD. This was achieved by incorporating pupil diameter as a metric within a telerobotics setup, allowing us to explore the nuanced relationship between SoE and PD in this context. Compared to a VR or AR setting, a telerobotics setup demands elevated cognitive workload and attention. Learning to tele-manipulate the device and engage with the remote environment requires heightened mental engagement. Thus, this setup provided a pertinent context to test the indirect effect in the absence of emotional stimuli, coupled with the challenges posed by a demanding task.

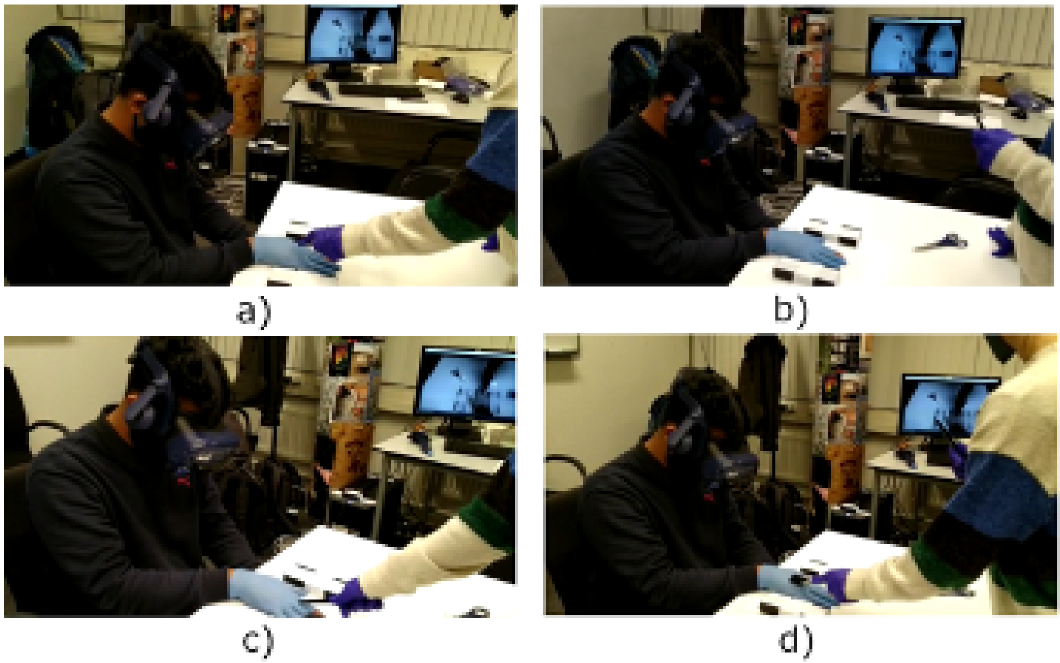


Figure 3.2: Pictures a) and b) represent, respectively, congruent and incongruent stimuli during the first variant of cross-modal congruency task. Pictures c) and d) represent, respectively, congruent and incongruent stimuli during the second variant of cross-modal congruency task. The screen in the picture was displaying the video of the embodiment experience, that was a reference for the experimenter to provide the stimuli.

3.4.1 Method

We explored the relationship between SoE and task performance, learning effect, and cognitive workload. We manipulated two embodiment conditions experienced by two groups: one group experienced sensory cues that support embodiment, while the other group experienced sensory cues that suppress embodiment. Each group had to face the experiment task at two levels: one with their own hand and another with a robotic surrogate.

3.4.1.1 Participants

28 right-handed participants (16 women and 12 men, between 19 and 49 years old) were recruited from the TNO participant pool. Participants were paid 30 € and their travel costs were reimbursed. The participants were divided into two groups: 15 participants experienced the supportive condition, while 13 participants were the suppressive one. The ethics committee of TNO approved the study (RP 2021-088).

3.4.1.2 Procedure and Task

The experiment lasted 45 minutes. Participants were asked to read and sign the consent form. Then, they performed a peg-in-hole task with only two distant horizontal holes. They had 90 seconds to place the peg in the holes as many times as they could.

They had to repeat the task six times in total, three times by using their own hand, and three times by using the robotic surrogate. Half of the participants completed the three tests using their own hand first, while the other half did so using the robotic surrogate first. After each trial, they had to complete two questionnaires.

3.4.1.3 Setup and Material

The teleoperation setup consisted of a telemanipulator, a haptic control interface, and a visual telepresence system. The telemanipulator was Shadow Hand Lite, equipped with 3D force sensors on its fingertips, mounted on the flange of a KUKA IIWA 7 serial link robot. The haptic control interface was developed by SenseGlove DK1, which tracks finger movements at 11 degrees of freedom and can provide passive force feedback on each finger. The movements of the operator's wrist in space are recorded by an HTC VIVE tracker mounted on the SenseGlove. The visual system consists of a ZED mini stereovision camera with a HTC VIVE Pro Eye that relays the visuals to the operator while also collecting gaze and PD. The setup was slightly different between the two conditions. For the supportive condition, the ZED mini was placed to provide a 1PP, and the operators received tactile feedback just at the moment they grasped and released the peg. In the suppressive condition, instead, the ZED mini provided a 3PP by facing the operators (i.e., a mirrored perspective of the workspace). In addition, they had to wear two thick gloves during the accomplishment of the task with their own hand and, while accomplishing the task using the robotic surrogate, the tactile feedback was continuous from the moment they grasped the peg until they released it (see Figures 3.3 and 3.4 for an overview of the setup). In both conditions at both levels, the participants had to wear the HMD during the experiment.

3.4.1.4 Measures

As an implicit measures, we collected pupil dilation, and task performance. As explicit measures, we used a reduced version of the SoE questionnaire from [198] (we addressed two items for each embodiment component: ownership, agency, and self-location)and we administered the cognitive workload questionnaire from [112].

3.4.2 Results

For PD analysis, we applied a Hampel filter to re- move the outliers, we used the convergence to cover missing data, and we considered the mean of the left and right pupils. For the supportive condition, we observed a significantly lower mean pupil dilation when participants completed the task with their own hand (M = 4.39mm) than with the robotic surrogate (M = 4.68mm) (t_{14} = 4.08, p-value = .001), the responses of the embodiment questionnaire showed the same effect (ownership t_{14} = 4.44, $p < .001$, M human hand = 5.89, M robotic hand = 3.92; Agency t14 = 9.95, $p < .001$, M human hand = 6.73, M robotic hand = 3.69; Self-location t_{14} = 3.59, p = .003, M human hand = 4.08, M robotic hand = 3.11). Of the cognitive workload questionnaire, only the subscale *mental workload* (M human = 2.40, M robot = 4.16)(t_{14} = 3.73, p = .002) showed an effect. For the suppressive condition, we observed a significantly lower mean pupil dilation when participants completed

Figure 3.3: Frames extracted from the ZED mini recordings during the supportive condition. On the left, the participant's view during the human hand level. On the right, the participant's view during the robotic surrogate level.

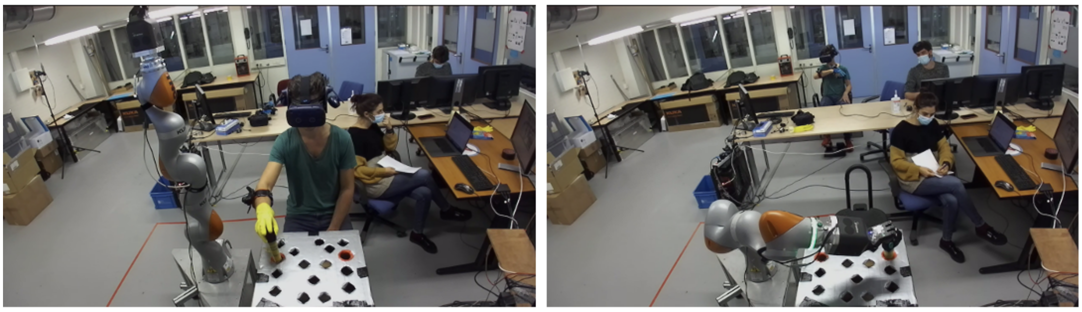


Figure 3.4: Frames extracted from the ZED mini recordings during the suppressive condition. On the left, the participant's view during the human hand level. On the right, the participant's view during the robotic surrogate level.

the task with their own hand (M = 3.68mm) and with the robotic surrogate (M = 4.13mm) (t_{12} = 4.07, p-value =.002), the responses of the embodiment questionnaire showed the same effect (ownership t_{12} = 7.55, $p < .001$, M human hand = 5.70, robotic hand = 2.74; Agency t_{12} = 6.47, $p < .001$, M human hand = 5.89, robotic hand = 3.06; Self-location t_{12} = 2.75, p = .018, M human hand = 4.17, robotic hand = 3.17). The cognitive workload questionnaire also showed the same effect while using their own hand (M = 19.97) and the robotic surrogate (M = 23.59)(t_{14} = 2.30, p = .04).

3.5 General Discussion

Based on the results of User Studies 1 and 2, where emotional stimuli, such as a threat to the surrogate, were introduced, a positive correlation between Pupil Diameter (PD) and Sense of Embodiment (SoE) was observed. On the basis of our results, we can partially accept H1 (we expect a positive correlation between PD and SoE in the presence of emotional stimuli like a threat to the surrogate). This finding aligns with the Rubber Hand Illusion literature, where a similar effect is reported, namely no effect in the absence of emotional stimuli, but a significant effect when the surrogate hand is under threat [71, 72, 184, 201**?**]. However, while User Study 2 is coherent with our hypothesis, in User Study 1 we found an effect only in the suppressive group (while we would have expected the opposite).

User Studies 1 and 3 observed that PD tended to be greater for participants experiencing conditions designed to provide low SoE compared to high SoE. Although User Study 2 did not confirm this effect due to an uniform supportive embodiment illusion in both groups. This aligns with the predicted indirect effect of SoE on PD through workload, as indicated in the literature [116, 184]. The negative correlation is expected when individuals face a task with a certain amount of workload, and this was supported by the responses to the cognitive workload questionnaire in User Study 3, particularly in the suppressive condition, in which participants had to struggle to complete the task with their own hand and even more with the robotic surrogate. However, in User study 1, even if the PD of the supportive group was smaller than the suppressive one, we did not find a significantly lower mean. Due to these results, we partially accept H2 (we expect a negative correlation between PD and SoE in the absence of emotional stimuli).
These results do not appear to be perfectly consistent among the three studies, and they raise some doubts and limitations. For example, in User Study 1, we found a PD effect under threat in the suppressive condition, but not in the supportive condition as expected. The potential restriction of pupil dilation under strong SoE raises the consideration of a floor effect. In situations of heightened embodiment, where pupils are constricted, PD may reach a limit, limiting its sensitivity to variations in high SoE. The conclusion suggests that PD is sensitive to nuanced changes in SoE, particularly when SoE is not overwhelmingly high. This sensitivity can be valuable for detecting subtle variations in emotional responses. However, this interpretation underscores that PD may not offer universal insights across all levels of SoE. Instead, it appears to be most informative in specific conditions, such as weak SoE.

3.6 Conclusions

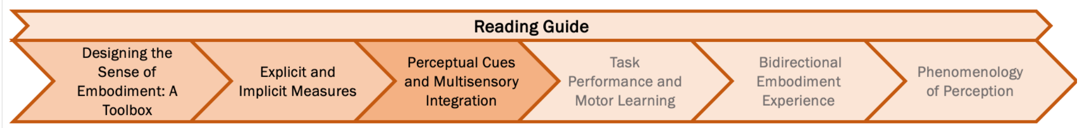
The need for further investigation is emphasized, especially concerning the versatility and the level of sensitivity of PD as a measure of embodiment. Despite the need for further investigation and the acknowledged limitations, PD is considered a suitable and promising measure of the direct and indirect effects of SoE. Future directions could involve exploring other eye recordings that may be informative, such as blink rate patterns [207], which were demonstrated to be a valid measure of individual engagement. In summary, the studies provide nuanced insights into the complex relationship between PD, SoE, emotional stimuli, and workload. While confirming some expected correlations, the results also highlight the need for ongoing exploration and refinement of methodologies to better understand the intricacies of these physiological responses in the context of embodiment and emotional experiences.

4

The Relative Contribution of Five Key Perceptual Cues and their Interaction to the Sense of Embodiment

Perception precedes reality.

Andy Warhol



Abstract

A range of perceptual cues drive the Sense of Embodiment (SoE) with an external object, such as a virtual arm that looks like one's own or can be controlled like one's own. Since most experiments test one or two cues at a time, it is difficult to establish their relative contribution and possible interaction. This work aims to investigate the importance of five key perceptual cues (field of view, visuo-proprioceptive synchronicity, tactile feedback, visual humanlikeness, and connectedness) and their potential interaction in a single, full factorial experiment. Participants touched a target dot, which changed position after a hit, for one minute in a virtual environment seen through a head-mounted display. The movements of the arm and hand of the participants were mapped to a virtual arm and hand. All perceptual cues had two levels: SoE supportive and SoE suppressive. 28 participants completed the task in all possible combinations. We recorded Task Performance and self-ratings of Sense of Ownership and Sense of Agency. The results showed that visuoproprioceptive synchronicity had the largest effect on all three measures. The relative importance of the remaining four cues differed for the three dependent variables. The cues did not have significant interactive

effects. We conclude that when designing an interface for maximum supportive embodiment, visuo-proprioceptive synchronicity is the most important perceptual cue. The extent to which other supportive cues can improve the embodiment experience depends on the considered variables, but generally the more other supportive cues can be added, the better.

4.1 Perceptual Cues and Their Role in the Embodiment Experience

The relation between SoE components and how to disentangle them is still an open debate. Most of the literature reports a strong correlation between the sense of ownership and agency ([93, 205, 252]). The two senses have common spatiotemporal constraints of the integration processes, and they affect each other at a level that is very difficult to disentangle their effects on the embodiment experience. However, there is also evidence that these two components involve different cognitive processes ([124, 240, 251]). These cognitive and neuroscientific studies sustain that although both components appear to be phenomenally uniform and strongly correlated, they are complex cross-modal phenomena of heterogeneous functional and representational levels.

Understanding SoE is particularly important if SoE determines performance in teleoperation. Teleoperation systems have been developed to allow human operators to perform complex tasks in unpredictable or hazardous environments, such as inspecting deep-sea and space structures, mining operations, and minimally invasive surgery ([65, 191, 228]). Teleoperation aims to replicate human manipulative skills and dexterity at a remote workplace over an arbitrary distance and on an arbitrary scale ([185]). If operators feel as if they are present in the remote or virtual world, this can increase control, improve task performance, and reduce cognitive workload ([217]). The idea is that operators should ideally have the (illusory) experience that avatar body and hands are their own body and hands, not or hardly noticing that the operation is being mediated, increasing the transparency of the teleoperation system ([38]). The role of SoE in the impact of teleoperation performance ([185]) gained attention in the last decade, when studies on the illusion and experience of embodiment started to be designed and developed ([69, 180, 201]). Some studies try to demonstrate that a high level of SoE can improve the performance of telepresence and teleoperation tasks ([167, 217, 220]). The study by Sanchez-Vives and Slater (2005) supported the investigation of telepresence beyond only the domain of computer science and other technologically oriented disciplines, but also as a mainstream part of neuroscience. They maintain that studies on perception, way-finding, self-representation, and sense of self will also contribute to the understanding of telepresence. Moreover, they stated that the concept of presence is sufficiently similar to consciousness that it may help transform research within that domain. This is indeed the direction this line of research is heading. Schiefer et al. (2015) wanted to assess the effect of sensory feedback on task performance in people with limb loss. Sensory feedback from peripheral nerve stimulation improved object discrimination and manipulation, sense of embodiment, and confidence. An embodiment survey showed an improved sense of integration of the prosthesis into self-body image with sensory feedback. Even Marasco et al. (2018) tested the importance of sensory feedback. The authors developed an automated neural-machine interface that vibrates the muscles used to control prosthetic hands. This system stimulates the kinesthetic sense in amputees, allowing them to control the movements of the prosthetic hand in the absence of visual feedback and increasing their sense of agency. This approach was a promising strategy to improve motor performance.

We are interested in the key perceptual cues that cause SoE and affect task performance. Mostly using the classic Rubber Hand Illusion paradigm ([29]), several cues have been identified that affect SoE and task performance. However, it is unclear what the relative importance is of, for example, visuotactile and visuomotor synchronicity between the real and virtual arm ([133]), likeness of the fake hand ([107]), the position of the real limb with respect to the virtual one ([250]), viewing mode (direct view, virtual reality) and point of view ([232]) on the strength of embodiment illusion. In the present study, we use a pointing task with a virtual hand to experimentally explore the effect of five important perceptual cues, selected from the literature review by Toet et al. (2020) on SoE and task performance: field of view, visuo-proprioceptive synchronicity, tactile feedback, visual human likeness, and connectedness. To the best of our knowledge, testing the effect of five perceptual cues in a single experiment, with the result allowing direct comparison and exploration of interactions, has not been done before. More cues have been reported to affect SoE ([246]), but including more cues would have made the experimental design and potential interactions too complex and the duration of the experiment too long. The cues were also selected so that they were compatible and independently modifiable within the task at hand. Below is a description and background of the selected cues.

i) The *field of view* is the open and observable area that individuals can see without head or eye movements, directly or through an optical device ([84]), such as a VR headset. Normally, humans have a slightly over 210-degree forward-facing horizontal arc of their visual field, i.e. without eye movements (with eye movements included, it is slightly larger). It was shown that a reduced field of view affects the sense of agency and can cause movement impairments ([266]).

ii) *Visuo-proprioceptive synchronicity* refers to the synchronicity between visual and proprioceptive cues detected by the operator. Since proprioception and visual information are normally aligned in time and space, the synchronicity of the two could break the embodiment illusion ([140, 144]). However, even if visuo-proprioceptive synchronicity seems to improve subjective embodiment perception, the importance and weight of this cue on the sense of embodiment is still unclear ([41]).

iii) *Tactile feedback* refers to the availability of tactile information. In teleoperation and VR settings, operators lack natural tactile feedback because their hands are not actually touching any object, and the current haptic technology is still limited. Tactile feedback can be provided artificially using very small forces or cues (such as vibration) that are mostly only felt through mechanoreceptors in the skin ([143]). The effects of tactile feedback and visuo-tactile synchronization on SoE were extensively studied and explored. Currently, there is still an open debate between who supports the importance of this perceptual cue ([88]) and who considers it not essential to the embodiment experience ([144]).

iv) Arm *visual humanlikeness* refers to the humanlikeness and appearance fidelity of embodied objects, such as a robotic hand. The more the object that is controlled is similar to the real body, the greater will be the operator's sense of embodiment ([176, 229]). Several studies have been conducted on the effect of avatar anthropomorphism on self-identification and likeness of the user or operator ([129, 145]), leading to the design of anthropomorphic systems in, e.g., robotics, virtual reality, mixed reality systems to increase user experience ([154, 179]).

Finally, v) *connectedness* of the arm to the body refers to the perception that the embodied object is an extension of the operators' body and (visually) connected to it as a continuum of the their own body ([158]). In Perez-Marcos et al. (2012), the authors investigate the importance of four factors on the SoE in a virtual rubber hand illusion: visuo-tactile synchronicity while stroking the virtual and the real arms, body continuity, alignment between the real and virtual arms, and the distance between them. The results show that the subjective illusion of ownership over the virtual arm and the time to evoke this illusion are strongly affected not only by synchronous visuo-tactile stimulation but also by connectivity of the virtual arm with the rest of the virtual body. In our study, these perceptual cues were presented at two levels: one to support the SoE (referred to as SoE supportive) and one to suppress or break the SoE (referred to as SoE suppressive) in the context of controlling a virtual arm. While there is an extensive body of literature on the effects of each of these single cues on embodiment, their relative contribution and potential interaction are still unknown. By including all five cues in a full-factorial design, we can measure the relative contribution of each to the SoE and task performance and test for possible interactions between cues.

4

4.1.1 Research questions and hypotheses

This study was led by three research questions (RQ):) What is the ranking of the perceptual cues for two tested SoE components (sense of ownership and agency) and task performance, and is the order consistent over the different dependent variables recorded to estimate the SoE? 2) Are there interaction effects between the perceptual cues? Is a simple additive model sufficient or do we need a more complex model to test the effect of the perceptual cues together? 3) Are SoE and task performance related? If so, this may imply that a higher SoE leads to higher task performance. Associated with these RQs are the following hypotheses (H):

H1a) We hypothesize that perceptual cues have a different weight in affecting the sense of ownership, the sense of agency, and the performance of tasks. Specifically, H1b) the sense of ownership will be mainly affected by the visual human-likeness, the visuo-proprioceptive synchronicity, the field of view, and the connectedness, which are the cues most strongly connected to the appearance and veridicity of the external device in relation to the human operator's perception of the remote body and environment. H1c) The sense of agency will be mainly affected by visuo-proprioceptive synchronicity, tactile feedback, and field of view, as this component is related to the mimicking of actions by the external device and the feeling of control. Finally, H1d) the task performance will be directly affected by the visuo-proprioceptive synchronicity and the field of view, since this component is related to the efficiency and effectiveness in accomplishing the task and indirectly affected by all the other cues, namely through increasing SoE.

Regarding RQ2, we expect that H2) we will not find a significant interaction between the perceptual cues.

Finally, in H3) we expect that the SoE and task performance are related.

4.2 Method

4.2.1 Participants

28 right handed participants (16 females and 12 males, between 19 and 49 years old) were recruited from the pool of TNO participants. The sample size was determined based on similar previous studies found in the literature ([167, 233, 234, 251]). Because the questionnaire was administered in its original language (English), participants could only join if they could read, speak, and understand English. The study was approved by the TNO Institutional Review Board (reference number: 2020-012). Participants were paid 30€ and their travel costs were reimbursed.

4.2.2 Materials and task

Participants viewed a virtual scene through a head mounted display (HMD), the HTC Vive. The HTC Vive offers a 110 ° field of view, a maximum refresh rate of 90 frames per second, and a combined resolution of 2160 × 1200 pixels (1080 × 1200 pixels per eye). A Vive Tracker was placed on the floor to determine the center of the half-circle range, and another was strapped to the right wrist of the participant. The experiment was run on a Lenovo Legion T730-28ICO 90JF with a GEFORCE RTX 2080 Super graphics card and an Intel Core i9 processor. The project was created in Unity 2019.2.17f1 and Visual Studio 2019. The scene was visualized using SteamVR 1.15.19 and the SteamVR Unity Plugin 2.6.1. The 'VR Hands and FP Arms Pack' from NatureManufacture was used for the arm and hand. The Rootmotion 'Final IK' package was used to allow arm segments to move naturally. To receive tactile feedback, participants used the Elitac Tactile Display, which is a glove that contains tactors to provide tactile stimulation to different parts of the hand. During the experiment, we activated one tactor placed on the tip of the right index finger for tactile feedback, with no offset at an intensity of 10 out of 15 on a logarithmic scale and a duration of 200 ms. The virtual environment consisted of a white grid with 9 dots, of which 8 were black and one was red. Participants were asked to touch the red dot. Upon touching, the red dot turned black and a random other black dot turned red. Participants were asked to touch as many red dots as they could in one minute.

4.2.2.1 Questionnaire

We administered a shortened version of the Embodiment questionnaire, in English, from Gonzalez-Franco & Peck (2018) to measure the sense of ownership and the sense of agency. Because participants had to complete the questionnaire 32 times, we reduced the number of questions from 16 to 5. We removed questions that were somewhat repetitive and addressed similar aspects of the sense of ownership and agency. We kept the questions that were easy to understand and had a clear relation to our experimental setup. For the sense of ownership, we administered the following questions: 1) I felt as if the virtual hand was my hand; 2) It seemed as if the virtual hand replaced my real hand; 3) It seemed as if the touch I felt was caused by the virtual hand touching the virtual target; 4) At some point it felt that the virtual hand resembled my own hand. For the sense of agency, we administered the following question: 5) It felt like I could control the virtual hand as if it were my own hand. Participants rated sense of ownership and agency on a 7-point Likert scale (from 1 = strongly disagree, to 7 = strongly agree). The order of the questions was randomized for each trial.

4.2.3 Design

Each perceptual cue had two levels: 1) supportive, in which we set the perceptual cue so that strong SoE was expected; 2) suppressive, in which we set the perceptual cue so that weak SoE was expected. The supportive and suppressive settings of the perceptual cues were as follows: 1) the field of view in the supportive condition allowed participants to have a human-like range of view (approximately 90 degrees temporally to central fixation, 50 degrees superiorly and nasally, and 60 degrees inferiorly); while in the suppressive condition, we narrowed the range and participants had the experience of observing the environment through a 14x4 cm casing (approximately 90 degrees temporally to central fixation, 25 degrees superiorly and nasally, and 30 degrees inferiorly). 2) Visuo-proprioceptive synchronicity was supportive when the movement of the real and virtual hand was congruent, and suppressive when we added a 20 ms delay[1] to both visual and tactile feedback. 3) the tactile feedback, in support conditions, was characterized by a vibration that the participants felt every time they touched the red point. To create the suppressive level of this cue, we removed the tactile feedback. 4) When the visual human-likeness was in its supportive condition, participants controlled a realistic human virtual hand; in the suppressive condition, they controlled a blue shiny virtual hand. Finally, 5) when the connectedness was in the supportive condition, the virtual hand was attached to a virtual arm. In the suppressive condition, instead, participants had to perform the task by manipulating a floating hand (see Figure 4.1 for an overview of the setup). Participants performed the 32 trials of the same task, one for each possible combination of perceptual cues, that is, trials $2(supportive/suppressive)^5(perceptualcues) = 32$, where one trial is a one minute point task. The order of the trials was randomized per participant.

[1]The delay was decided on the base of the pilot data: we tested the same task by adding 10ms, 20ms, and 40ms delay. While the 10ms delay could barely be perceived, the 40ms delay made task achievement almost impossible and too frustrating for the operator. Therefore, we opted for 20ms delay, since it made the task challenging but doable. However, even if the intentional delay was 20 ms, the real one might have been higher, due to the setup implementation: the system would wait for 20ms, start storing frames, and then show the participants the frame that was saved 20ms before. This process took some time (in the microsecond range). The trackers themselves and HMD have their own delays (in the millisecond range). Although the limitation is that the total delay was not completely computable, it was constant for all participants

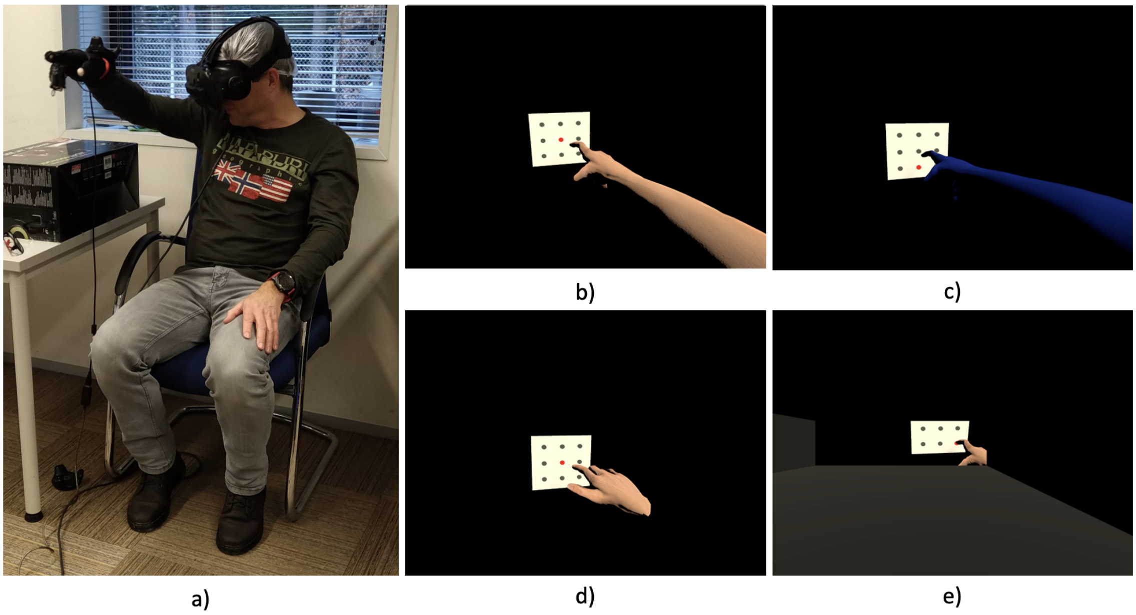


Figure 4.1: On the left, a) is a picture of a participant accomplishing the required task. On the right, frames extracted from the HMD recordings during different perceptual cues configurations. b) represents all the cues in the supportive condition; c) represents only the human-likeness in the suppressive condition; d) represents only the field of view in the suppressive condition; e) represents all the cues in the suppressive conditions.

4.2.4 Procedure

Participants were asked to sit in a chair, wear the sensory glove, the tracker on the wrist, and place their right arm on a support platform on their right. The supportive platform was used to make the task less challenging for the participants. Before asking to wear the VR headset, we instructed them on the procedure and the task. Next, we calibrated the position of the virtual hand and the target according to the height of the participants. During the calibration process, they experienced controlling the virtual hand. Participants performed the 32 1-minute trials. After each trial, they answered the short embodiment questionnaire. Participants continued to the next trial whenever they indicated they were ready. At the end of the experiment, we asked the participants for feedback on their experience and their opinion on the experiment.

4.2.5 Analysis

For each participant and each of the 32 trials, the sense of ownership was defined as the average score for the four sense of ownership questions. For the sense of agency we used the score on the sense of agency question. Task performance was defined as the number of touched red dots. To determine the classification of the five perceptual cues in order of importance with respect to their effect on the three dependent variables (RQ1) and to determine whether interactions occurred (RQ2), we used a linear mixed effects model, applying a pairwise comparison between the independent variables. The five perceptual cues (and their pairwise interactions) were treated as fixed effects, while the participants were treated as random effects. We used a step function to remove the non-significant (p-value > .05) interactions between the factors (perceptual cues) from the model. We removed, each time we updated the model, the interaction with the highest p-value, until all included perceptual cues had a significant effect on the dependent variable. This operation was done for each dependent variable (sense of ownership, sense of agency, and performance as defined by the number of touched dots) separately and resulted in a ranking of the cues in terms of t-statistics. To determine whether SoE and task performance are related (RQ3), we performed Pearson's correlations between sense of ownership, agency, and task performance.

4.3 Results

Figures 4.2, 4.3, and 4.4 show the sense of ownership, the sense of agency, and the performance of the task for each of the (supportive and suppressive) levels and each of the five cues. Figures 4.2 and 4.3 indicate that the manipulations worked for all perceptual cues in the expected direction: when the cues were in the supportive condition, SoE appeared to be higher and task performance appeared to improve compared to the suppressive condition. Figure 4.4 indicates that, except for visuo-proprioceptive synchronicity, the effects on task performance were small and less consistent.

The results of the linear mixed-effects model (Table 4.1) confirmed these impressions. A significant effect was found for all perceptual cues on the sense of ownership and agency, with the only exception of human likeness on the sense of agency ($t = -0.51$, $p = 0.61$). For task performance, a significant effect was found of the visuo-proprioceptive synchronicity ($t = -7.07$, $p < .001$) and connectedness ($t = -2.32$, $p = .02$).

In Table 4.2, we report the rank order of the perceptual cues based on the t-values, separately for the sense of ownership, the sense of agency and the performance of the task, and allowed to answer our first research question (What is the ranking of the perceptual cues for the sense of ownership, the sense of agency and the performance of the task, and is the order consistent with the different dependent variables recorded to estimate the SoE?). Table 4.2 indicates the relative importance of perceptual cues to obtain a high SoE experience and task performance. For all dependent variables, modulating visuo-proprioceptive synchronicity had the strongest effect.

The linear mixed-effects model indicated that none of the perceptual cues showed significant interaction effects on any dependent variable, answering our second research question as to whether there are interaction effects between the perceptual cues?). Table 4.3 indicates the order of removal of the interactive and main effects of the different models.

To answer our third research question (Are the SoE and task performance related?), we performed a Pearson's correlation analysis between the sense of ownership, agency, and task performance. SoE and task performance were found to be weakly, but significantly correlated for both sense of ownership ($r=.13$, $df=26$, $p<.001$) and sense of agency ($r=.18$, $df=26$, $p<.001$). A strong correlation was found between the sense of ownership and the sense of agency ($r= .75$, $df=26$, $p < .001$).

Dependent Variable	Independent Variable	df	std error	t-value	p-value
Sense of Ownership	Field of view	863	0.245	-2.959	.003
	Human likeness	863	0.245	-4.234	<.001
	Tactile feedback	863	0.245	-12.866	<.001
	Connectedness	863	0.245	3.251	.001
	Visuo proprioceptive synchronicity	863	0.245	-19.003	<.001
Sense of Agency	Field of view	864	0.074	-2.914	.004
	Human likeness	864	0.074	-0.511	.61
	Tactile feedback	864	0.074	-7.962	<.001
	Connectedness	864	0.074	2.433	.015
	Visuo proprioceptive synchronicity	864	0.074	-26.709	<.001
Task Performance	Field of view	865	1.470	-1.911	.06
	Human likeness	865	1.470	0.784	.43
	Tactile feedback	865	1.470	1.082	.28
	Connectedness	865	1.470	-2.320	0.02
	Visuo proprioceptive synchronicity	865	1.470	-7.067	<.001

Table 4.1: Overview of the pairwise comparison between the levels supportive and suppressive for the perceptual cues for sense of ownership, sense of agency, and task performance.

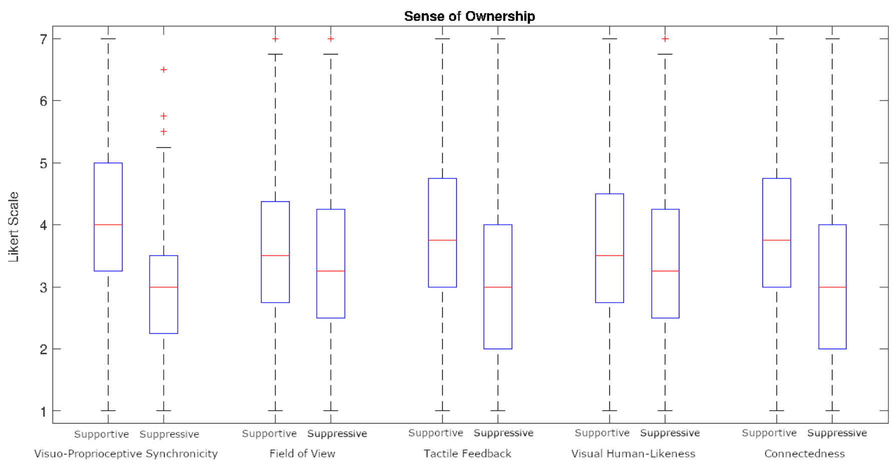


Figure 4.2: The sense of ownership scores for all the perceptual cues in each condition, supportive and suppressive. Legend: On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the $+$ symbol.

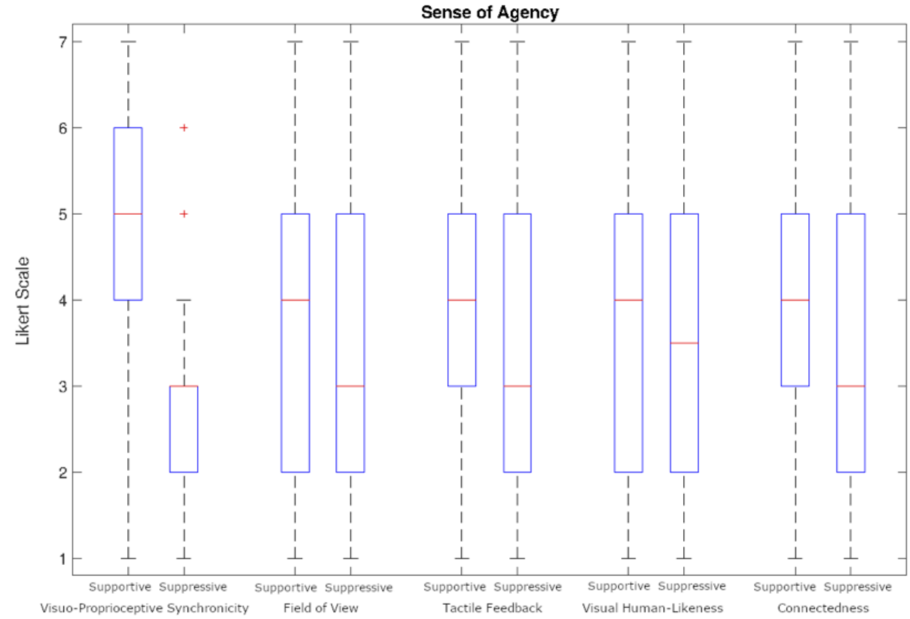


Figure 4.3: The sense of agency scores for all the perceptual cues in each condition, supportive and suppressive. Legend: On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the $+$ symbol.

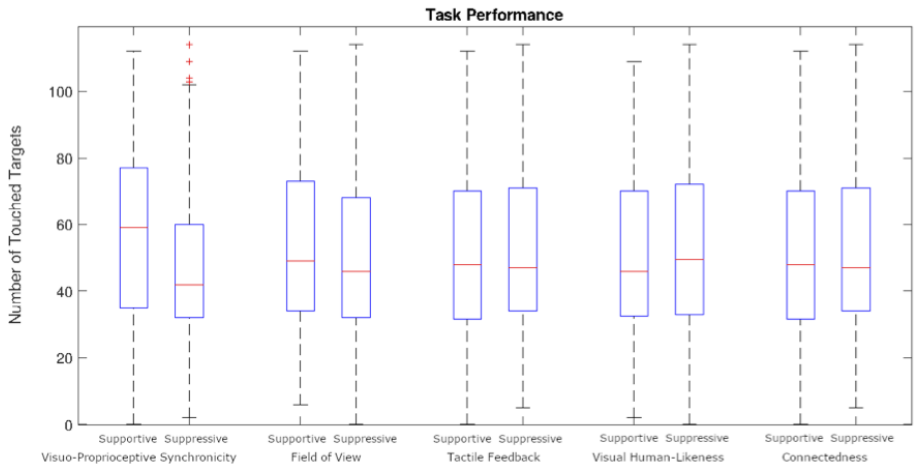


Figure 4.4: The task performance scores for all the perceptual cues in each condition, supportive and suppressive. Legend: On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the $+$ symbol.

4

Ranking	Sense of Ownership	Sense of Agency	Task Performance
1	Visuo-proprioceptive synchronicity	Visuo-proprioceptive synchronicity	Visuo-proprioceptive synchronicity
2	Tactile feedback	Tactile feedback	Connectedness
3	Human-likeness	Field of view	Field of view (no effect)
4	Connectedness	Connectedness	Tactile feedback (no effect)
5	Field of view	Human-likeness (no effect)	Human-likeness (no effect)

Table 4.2: The ranking (based on t-values) of the perceptual cues for sense of ownership, sense of agency, and task performance.

Dependent Variable	Interaction and Main Effect	Removal Order
Ownership	field of view - visuo-proprioceptive synchronicity	1
	field of view - connectedness	2
Agency	field of view - human-likeness	1
	field of view - tactile feedback	2
	human likeness - connectedness	3
	field of view - visuo-proprioceptive synchronicity	4
	human-likeness - tactile feedback	5
	field of view - connectedness	6
	human-likeness - visuo-proprioceptive synchronicity	7
	tactile feedback - visuo-proprioceptive synchronicity	8
	tactile feedback - connectedness	9
	connectedness - visuo-proprioceptive synchronicity	10
	human-likeness	11
Task Performance	connectedness - visuo-proprioceptive synchronicity	1
	human-likeness - tactile feedback	2
	field of view - connectedness	3
	field of view - human-likeness	4
	tactile feedback - visuo-proprioceptive synchronicity	5
	human-likeness - connectedness	6
	tactile feedback - connectedness	7
	field of view - tactile feedback	8
	human-likeness - visuo-proprioceptive synchronicity	9
	field of view - visuo-proprioceptive synchronicity	10
	field of view	11
	tactile feedback	12
	human-likeness	13

Table 4.3: The order of removal of the interaction and main effects from the different models.

4.4 Discussion

Our study resulted in rank orders of perceptual cue importance for the sense of ownership, sense of agency, and task performance. We found significant effects of the perceptual cues on each dependent variable, with some exceptions (RQ1). We did not observe any significant interaction effect among the independent variables, supporting the hypothesis that a linear, additive model can adequately describe the combined results of the five perceptual cues (RQ2). Furthermore, we found a weak but significant correlation between SoE and task performance (RQ3).

To discuss our findings in detail, on the basis of the results reported in Table 4.2, we accepted H1a (perceptual signals have a different weight in affecting the sense of ownership, agency and task performance). We observed a different effect of each perceptual cue on the dependent variables. H1b (the perceptual cues that mostly affect the sense of ownership are visual human likeness, visuo-proprioceptive synchronicity, field of view, and connectedness), H1c (the perceptual cues that mostly affect the sense of agency are visuo-proprioceptive synchronicity, tactile feedback, and field of view) and H1d (the perceptual cues that mostly affect task performance are visuo-proprioceptive synchronicity and the field of view) were partially accepted. We found a significant effect of all perceptual cues on the sense of ownership and agency, with the only exception of the human-likeness on the sense of agency (as expected). However, the weight (based on the t value) that each cue had in affecting the dependent variable was slightly different than hypothesized: visuo-proprioceptive synchronicity was the cue that affected all dependent variables most. As for task performance, we found a significant effect of visuo-proprioceptive synchronicity and connectedness, but not of field of view.

We did not find an interaction effect among the independent variables, and the additive model that we adopted allowed us to test the independent variables. Based on the result, we accept both H2a (we do not expect an interaction between the perceptual cues) and H2b (an additive model is enough to test this combination of perceptual cues together).

Finally, our results were consistent with H3 (SoE and task performance are related). The correlations between the dependent variables were significant, although the correlation between the SoE components and the performance of the task was weak.

Visuo-proprioceptive synchronicity was the cue that most strongly affected the sense of ownership and agency. When this cue was in the suppressive condition, participants had more difficulties accomplishing the task and experiencing embodiment, as also reported in other studies ([62, 136]).

However, while visuoproprioceptive synchronicity has been found to be the most important of the perceptual cues tested, note that a weak SoE was still obtained in the suppressive condition (median sense of ownership score of 3, and a median sense of agency score of 2.5). For both SoE measures, visuo-proprioceptive synchronicity is followed by tactile feedback, demonstrating the importance of the information provided by the tactile sense in establishing SoE. Then, the ranking differentiates, showing a difference in the sense of ownership and agency. For the sense of ownership, the third rank is occupied by human-likeness, which is a coherent result with respect to our initial hypotheses, considering that the sense of ownership focuses on the sense of self-attribution of the external embodiment. Instead, for the sense of agency, we found the field of view to be more important. This can be explained by the fact that in the supressive condition the field size was only 14x4cm. This made control of the virtual arm and hand more complicated and reduced the possibility of observing the movement of the virtual arm, reducing the level of perceived agency. Ranking fourth for both sense of ownership and agency is connectedness. Connectedness provides more (visual) information on arm posture and position, supporting the experience of the external arm and hand as attached to one's own shoulder. This may have facilitated the perception of the joints in space and the control of the virtual hand. For the sense of ownership, ranking last, but still significant, is the field of view. The field of view under suppressive conditions degraded the perception of the environment but only had a small effect on the perception of embodiment, probably due to the point of view that was not changed. The first person perspective helps the operator to have an immersive perception of the embodiment, especially when it is realized using a VR headset ([235]). For the sense of agency, ranking last and non-significant is human-likeness. The finding that human-likeness did not affect sense of agency, but did have a significant effect on sense of ownership is in line with previous studies ([93, 240, 251, 252]). Pyasik et al. (2018) argue that the individual spatio-temporal constraints for the integration of sensory-related cues, which are unconscious, are common to both the sense of ownership and the sense of agency, whereas their subjective and conscious experience would rely on additional processes specific for each sense. In this experiment, we assessed SoE using a questionnaire ([198]) in this experiment. Although intended to measure perceptual experience, we cannot rule out the possibility that questionnaire data include (cognitive) bias. Combining questionnaires with implicit measures such as skin conductance response ([201]), heart rate ([235]), or pupil dilation ([78]), can make the results more robust and provide more information on the discrepancies in sense of ownership and agency.

For task performance, there are only two cues that affected it, firstly, visuo-proprioceptive synchronicity and secondly, connectedness, where performance was better in the suppressive condition than in the supportive condition. The effect of visuoproprioceptive synchronicity was consistent with participants always reporting difficulties in performing the task, while this cue was presented under suppressive conditions. If anything, we would have expected connectedness to increase proprioceptive information, therewith supporting task performance. The opposite finding may have been caused by a less cluttered display when only the disconnected hand is presented rather than the entireeee arm. Field of view, tactile feedback, and human likeness did not have a significant effect on task performance. The lack of an effect of field of view could be explained by the fact that, even if it could affect the SoE, especially the sense of agency as reported in Wenk et al. (2021), the reduction of this cue in the suppressive condition did not hamper task execution. This could also be related to both the design of the task, which was simple to accomplish, and the small workspace in which participants had to operate the virtual arm.

Although there is evidence that sense of ownership and agency involve different cognitive processes ([124, 240, 251]), in correspondence to most of the literature, we found a strong correlation between sense of ownership of body and sense of agency ([93, 205, 252]). The exact relation between the sense that one's body is one's own (body ownership) and the sense that one controls one's own bodily actions (agency) has been the focus of much speculation, but is unclear. Tsakiris et al. (2010) discuss two models to describe the relationship between the sense of ownership and sense of agency. First, an additive model, in which agency and body ownership are strongly related, because the ability to control actions is a powerful cue to body ownership; plus possible additional sub-components unique to ownership and agency. An alternative independence model sustains that agency and body ownership are qualitatively different experiences triggered by different inputs and recruiting distinct brain networks. a network of sensorimotor transformations and motor control, and a set of hetero-modal association cortices implicated in various cognitive functions. We still do not know the exact functions and contributions of these brain regions to the sense of agency. We found that the correlations between task performance and sense of ownership and between task performance and agency are significant but weak, with an explained variance below 5%. Previous studies reported that a higher SoE resulted in a much larger increase in task performance ([167, 217, 220]). The tasks in those studies required a major interaction with the environment, which was usually better characterized than ours. An interesting possibility could be that a high SoE reduces the cognitive workload of the operator and has a larger effect when the task demands are high, for example,when operating an avatar or teleoperation system that is more complexly designed and that can achieve more complicated tasks as is the case in studies. Following this line of reasoning, a higher SoE could also affect the learning speed: reducing the cognitive workload of the operator may speed up learning to perform the task. This prediction could be tested in a more complex task with novice users.

4.4.1 Future works

In future work, we want to redesign the task in order to include the assessment of the sense of self-location, to have a complete picture of the SoE. Moreover, we want to extend the perceptual cues to test, such as head movement control and point of view. To include these variables, we need to redesign the setup and the experiment. To circumvent possible relations between task performance and SoE as subjectively reported through post hoc knowledge of your own performance, implicit (unconscious) measures of SoE would be of great value ([261]). To this end, adding physiological measures such as skin conductance, heart rate, and pupil dilation may be of interest. Finally, and as mentioned above, we are interested in investigating how SoE may affect learning of a different and more complex task. However, in this case, the findings of Brouwer et al. (2014) indicate that using physiology to monitor learning is not as straightforward as one might expect.

4.4.2 Conclusions

Our full factorial experiment resulted in a rank order of five different perceptual cues with respect to their effect on the sense of ownership, the sense of agency, and the performance of tasks. Different rankings were found, but visuo-proprioceptive synchronicity affected all three outcome measures most strongly. We did not observe an interaction effect between the perceptual cues, and we found a weak relation between the level of SoE and task performance. These findings can help to decide on the factors to optimize in a system to achieve a high sense of ownership, agency, and task performance.

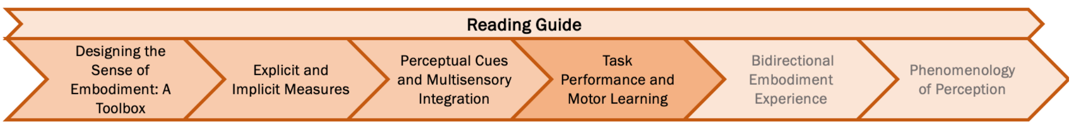
5

The Effect of Sense of Embodiment on Task Performance and Motor Learning

We are what we repeatedly do. Excellence, therefore, is not an act but a habit.

Aristotle

5



Abstract

In this chapter, we transition from the theoretical and experimental groundwork of manipulating and measuring Sense of Embodiment (SoE) to addressing a fundamental question: What is the purpose of optimizing the SoE in a teleoperation system? This exploration centers on investigating the potential positive effects of SoE on motor adaptation, the acceleration of motor learning, and the potential enhancement of task performance. The chapter delves into this investigation by focusing on two critical research questions (RQs) posed in the Introduction: RQ8) What is the effect of SoE on task performance in a perceptual-motor task? RQ9) What is the effect of SoE on the asymptote of the learning curve in a perceptual-motor task? Drawing insights from the existing literature, the hypothesis emerges that enhancing SoE yields positive effects, not only on task performance (H1) but also on the overall embodiment experience (H2). An additional layer of exploration is introduced through an exploratory research question: Are these results consistent across diverse scenarios and tasks? The study design encompasses three distinct user studies, each set in different applications and featuring various avatars, yet all anchored in similar tasks,

specifically a modified peg-in-hole task: 1) in the first user study, participants operated a robotic arm with a human-like hand as end-effector, and they were required to perform a classic peg-in-hole task. 2) In user study 2, the task is transformed into a variation that we called "peg-on-button", wherein participants use a robotic arm with a gripper as the end-effector to press a lit button. 3) User study 3 is implemented in virtual reality (VR), participants manipulate a virtual hand holding a controller to perform a reaching task and a proprioceptive judgment task. Across all three studies, a consistent pattern emerges: a setup that fosters embodiment has a positive impact on motor learning and adaptation, resulting in improved task performance. A supportive setup also reduces the perception of the surrogate as a mere mediator between the operator and the remote environment, especially when contrasted with a setup that suppresses embodiment. The positive effects on motor learning and task performance advocate for the incorporation of embodiment-supportive designs in teleoperational setups. However, the nuanced relationship between SoE and long-term task performance prompts a call for further exploration and consideration of various factors influencing teleoperation outcomes across diverse scenarios and tasks.

5.1 Introduction

One of the main reasons that brought to investigate SoE in teleoperation, was to observe the effects of its manipulation on task performance. In the literature, we could not find studies approaching this effect in the teleoperation context, therefore we decided to first understand how to manipulate it in a teleoperation setup, and then to observe the effects of this manipulation in different teleoperation scenarios. If high SoE can improve task performance, increase the learning rate, and decrease the motor adaptation time, then it becomes also relevant to optimize it in a teleoperation setup, and of give more importance to the design stage of a system. Our investigation was led by one main prediction that we tested: if SoE has a positive effect on motor adaptation, - i.e. faster learning when SoE is higher, and it decreases the perception of the surrogate as mediator; as a consequence, it also has an effect on task performance. In other words, higher SoE leads to increased the learning speed compared to lower SoE, and the final performance plateau after learning is dependent on the level of embodiment.

In this chapter, we present three user studies aimed at better understanding the concept of SoE and its effect on task performance and motor learning in four different teleoperation setups. We report how task performance and motor learning are affected by different levels of SoE. SoE levels are manipulated by varying perceptual cues known to affect SoE [74, 84, 210], as also reported in Chapter 4 of this thesis. Perceptual cue manipulation consists of influencing the multisensory integration of external stimuli obtained by controlling an avatar in order to improve or degrade the level of embodiment of the operator. This can be obtained by changing the field or point of view of the camera used in the surrogate presented to the operator, by activating or deactivating the tactile feedback, and by choosing a certain haptic device rather than another. The experiment presented in Chapter 4 was realized as a benchmark to understand the importance of five selected perceptual cues and the effect of their potential interaction with SoE and task performance. Based on that, we designed the conditions and setup of all user studies. The design and setup of Experiment 1 was previously presented and described in Chapter 3. However, while in that chapter we presented the assessment of pupil dilation as an implicit measure of embodiment, here we present data relative to task performance, and motor adaptation and learning, where motor adaptation refers to flexibility in learning new movements, but can also be used to determine whether some operators can generate a motor pattern to which they become used [264]. Repeated adaptation can lead to learning a new and more permanent motor calibration. This type of learning is likely to be an important method for making long-term improvements in operators' movement patterns. The conclusion on SoE and task performance would then depend on the moment of testing: during the learning phase, there may be a positive effect, but once learning is completed, this effect disappears. Therefore, in this chapter, we investigate not only the (final) level of performance but also the motor adaptation process. For what concerns the supporting literature related to the relationship between motor adaptation and SoE, we are testing this assumption in the presented studies for the first time. However, there are studies suggesting and exploring explicit and implicit motor learning strategies and how they can affect motor adaptation in different tasks and motor conditions [171, 242]. Our plan encompasses the development of experimental methods to measure both the explicit and implicit components of the SoE phenomenon, employing approaches that encompass both empirical and subjective assessments. Studies have shown that implicit and explicit processes underlie human motor skill learning and each of them has its own peculiar properties and contributions. For example, implicit processes such as sensory prediction and reward prediction error-based learning seem invariant and stationary, while explicit processes such as awareness and strategy synthesis appear to be stationary. Furthermore, in [172], the authors developed techniques that isolate these processes and that could naturally be extended to common SoE paradigms. Second, we experiment with different contextual features that could affect SoE, such as visual and dynamic similarity with the human limb. In our virtual reality (VR) environment and physics engine, this can be easily updated by proposing different tools/avatars with different kinematic properties. The main differences between the three user studies concern 1) the humanlikeness of the surrogate, 2) the level of immersion and participation of the embodiment experience, and 3) the way in which the SoE is manipulated. This allowed us to test the effect of SoE in different contexts of application. To make the task performance comparable, the studies have in common that the operators were required to manipulate a right arm, and they were asked to accomplish a similar task (variations of a peg-in-hole) with quantifiable performance metrics, i.e. a task that returns count data, and a similar level of difficulty. Experiment 1 involved a robotic arm with a mechanically complex, humanoid robotic hand attached to it (see Figures 5.2 and 5.3) and operators had to perform a peg-in-hole task. In Experiment 2, operators controlled a robotic arm with a peg as an end-effector (see Figure 5.8) and did a peg-on-buttons task, namely they had to press the button that lighted up among the six buttons in the remote workspace. In Experiment 3, participants were required to touch the red dot every time they spotted one on a 5x4 grid of black dots. The diagrams represented in Figure 5.1 summarizes the design of the three user studies. Starting from the research questions presented in the Introduction - RQ8) What is the effect of SoE on motor adaptation in a perceptual-motor task? RQ9) What is the effect of SoE on the asymptote of the learning curve (i.e. the performance level after the learning curve reached a plateau) in a perceptual motor task? - We hypothesize that H1) SoE has a positive effect on task performance, i.e. faster learning when SoE is higher, and H2) it improves the overall embodiment experience and decreases the perception of the surrogate as mediator. As a side issue, we also formulated an explorative research question: RQE) Are these results robust over different tele-robotics contexts and tasks (i.e. the three experiments)?

5.2 Prior Works

In teleoperation, some of the most investigated aspects are the effects of perceptual cues manipulation and time delay on SoE and task performance. In the three experiments presented in this chapter we manipulated visuo-proprioceptive information, visuo-tactile information, point of view, and the visual-likeness of the surrogate.
More specifically, time delay is one of the main issues encountered and studied in teleoperation systems, particularly in telerobotics. For instance, [15, 74, 122, 260] investigated to what extent an operator can deal with and adapt to time delay, and the impact of time delay on SoE. Overall, results show that time delay affects the integration of proprioceptive information and multisensory stimuli. As a consequence, the level of SoE decreases and operators need some time to adapt to the situation. It is still unclear what the tolerance threshold is (if one exists) of adaption to delayed signals in a teleoperation scenario. This threshold should point to the maximum limit of delay that an operator can handle to be neverthless capable of manipulating the surrogate. For what concerns the relation to task performance, studies demonstrated that it can be improved by conditions that support embodiment [167, 217, 220]. When the operator is strongly embodied with the surrogate, the perception of the surrogate as mediator is lower [38]. In other words, the operators do not perceive the surrogate as a third party object, or external tool, that mediates the interaction between them and the remote environment, but they feel embodied and in full control of the surrogate as if it was part of their own body.
Starting from the literature, it is assumed that different levels of SoE through a surrogate, obtained by the manipulation of several perceptual cues, will have an effect on the task performance. In the case when operators feel strongly embodied, task performance would improve compared to situations in which the embodiment level is weak. However, these results cannot be generalized, since they are mostly limited to VR studies, or they are conducted in the prosthetic field with participants who experienced an upper limb loss; another aspect is that they usually focus on cognitive load and not on motor performance [167, 217, 220, 262]. Moreover, there is still a debate if high SoE really leads to better performance. The effect of SoE on task performance has not been widely replicated and there are also studies that found no effect or just different advantages in manipulating certain perceptual cues [62, 74, 101, 137]. An alternative explanation for the inconsistent results could be that SoE is not directly

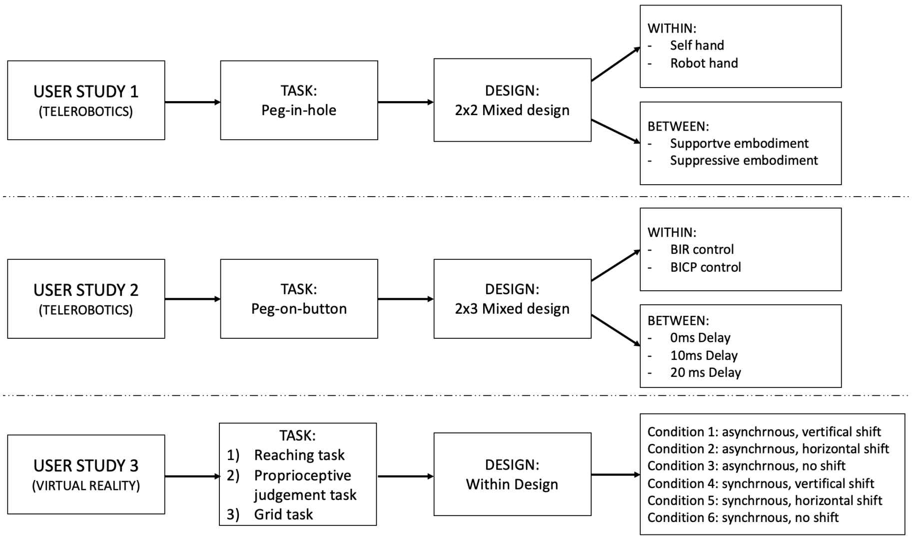


Figure 5.1: The diagram summarizes, for each user study, the setup, the task, the design, and conditions.

related to the final level of performance, but results in faster learning and a steeper learning curve, and differences in motor adaptation. These differences can be due to the context of application (VR, social robotics, field robotics, industrial robotics, or others), the level of complexity and the kind of task, and the different kinds of avatars.

5

5.3 User Study 1

5.3.1 Method

We explored the effect of SoE on task performance and the motor adaptation with two independent variables: SoE level (supportive, suppressive) which was a between subjects variable and *hand* (self, robot) which was a within subject variable. Of the two SoE level groups, the "supportive" group experienced perceptual cues that support embodiment, while the "suppressive" group experienced perceptual cues that suppress embodiment. Each group did the task under two levels of Hand: 1) "self" using their own hand (either with or without the gloves), and 2) "robot" using a robotic surrogate. We set the perceptual cue such that strong (supportive) or weak (suppressive) SoE was expected, based on previous studies [74]. Particularly, we manipulated the perspective (1PP for the supportive condition, and 3PP for the suppressive one), haptic feedback (on for the supportive and off for the suppressive).

5.3.1.1 Participants

Twenty-eight right handed participants (16 females and 12 males, between 19 and 49 years old) participated voluntarily. Participants were paid 30 € and their travel costs were reimbursed. Participants were divided into two groups: 15 participants experienced what we call the supportive condition, while 13 participants the suppressive one (see next section). The ethics committee of TNO approved the study (RP 2020-012).

5.3.1.2 Setup and Materials

The teleoperation setup consisted of a telemanipulator, a haptic control interface and a visual telepresence system. The telemanipulator was the Shadow Hand Lite, equipped with 3D force sensors on its fingertips, mounted on the flange of a KUKA IIWA 7 serial link robot. The haptic control interface was realized by the haptic glove SenseGlove DK1, which tracks finger movements in 11 degrees of freedom and can provide passive force feedback on each finger. The movements of the operator's wrist in space are recorded by an HTC VIVE tracker mounted on the SenseGlove. The visual system consists of a ZED mini stereovision camera with a HTC VIVE Pro Eye relaying the visuals to the operator, while also collecting eye gaze and pupil diameter. The HTC Vive offers a 110°field of view, a maximum refresh rate of 90 frames per second and a combined resolution of 2160 × 1200 pixels (1080 × 1200 pixels per eye). The setup was slightly different between the two conditions. For the supportive condition, the ZED mini was placed to provide a first-person perspective (1PP), and the operators received the tactile feedback (vibration) just in the moment in which they grasped and released the peg. In the suppressive condition, instead, the ZED mini was providing a third-person perspective (3PP) by facing the operators (i.e., a mirrored perspective of the workspace). The chosen 3PP was challenging, but the purpose was to observe if the participants' performance while using their own hand and the surrogate hand would have been drastically affected in both conditions. Moreover, they had to wear two thick gloves during the task accomplishment with their own hand and, while accomplishing the task using the robotic surrogate, the tactile feedback was continuous from the moment in which they grasped the peg until they released it (see Figures 5.2 and 5.3 for an overview of the setup). In both conditions at both levels, participants had to wear the HMD during the experiment.

5

5.3.1.3 Task

Participants had to repeat a peg-in-hole task 6 times in total, three times by their own hand and three times by using the robotic surrogate. After each trial, they had to fill a customized version of the questionnaire on embodiment from [198]. Half of the participants accomplished the three trials using their own hand first, while the other half by using the robotic surrogate first (see Figures 3.3 and 3.4).

5.3.1.4 Procedure

Participants were asked to fill out the consent form and then they were given detailed instructions about the experiment and the tasks. Participants were also instructed that if they were unsure how the task worked they could ask the experimenter for help. The experiment lasted 45 minutes. Participants were asked to do a peg-in-hole task. The grid, used as workspace, had 16 holes. However, participants were required to focus just on two of them marked in red. The two holes had a distance of 32cm. Participants had 90 seconds to place the peg in the holes as many times as they could.

Figure 5.2: Frames extracted from the ZED mini recordings during the supportive conditions of User Study 1. On the left, the participant's view during the self level. On the right, the participant's view during the robot level.

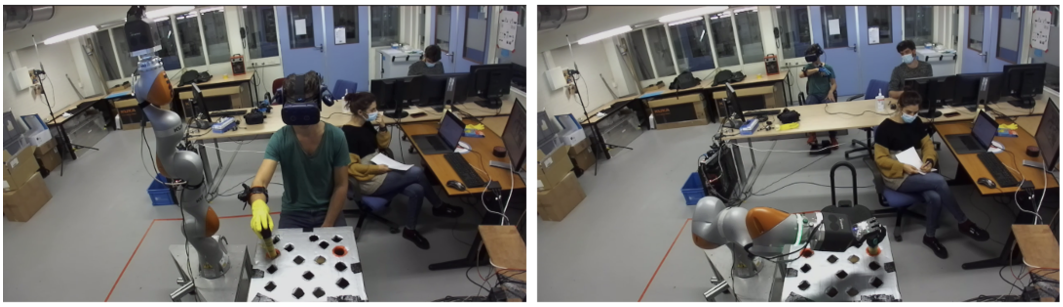


Figure 5.3: Frames extracted from the ZED mini recordings during the suppressive conditions of User Study 1. On the left, the participant's view during the self level. On the right, the participant's view during the robot level.

When participants accomplished the task with their own hand, they were sitting in front of the grid, and they could directly interact with it, but they were observing it through the HMD. When they were accomplishing the task using the robotic arm, they were sitting at a safety distance from the grid and they used the robotic device to interact with it.

5.3.1.5 Measures

To measure SoE, we adopted a reduced version of the embodiment questionnaire from [198]. Participants were asked to assess two items for each embodiment component (ownership, agency, and self-location) using a Likert-scale from 1 (*strongly disagree*) to 7 (*strongly agree*). Moreover, motor adaptation was determined on the basis of task performance over trials. SoE components were measured as the average score for the items addressing each variable in the questionnaire. Task performance was defined by how many times participants managed to insert the peg in the two holes for each trial of 90 seconds.

5.3.2 Results

We implemented a 2x2 mixed-design ANOVA model, with one dependent variable within subjects (self and robot hand) and one dependent variable between subjects with two levels (supportive and suppressive embodiment). We observed the effect of this manipulation on four dependent variables: the SoE components (the sense of ownership, agency, self-location), and task performance. To determine the motor learning among trials we applied a one-way ANOVA. We found a significantly positive trend in the mean scores of task performance in the supportive conditions among trials (see Table 5.1), at both levels. However, we did not encounter the same results in the suppressive condition at both levels. Moreover, we observed a significantly different score for the self level compared to the robot hand, in both conditions and for almost all the evaluated items and task performance (see Figures 5.4, 5.5, 5.6, 5.7 and Table 5.3). The self level resulted in higher scores than the robot one.

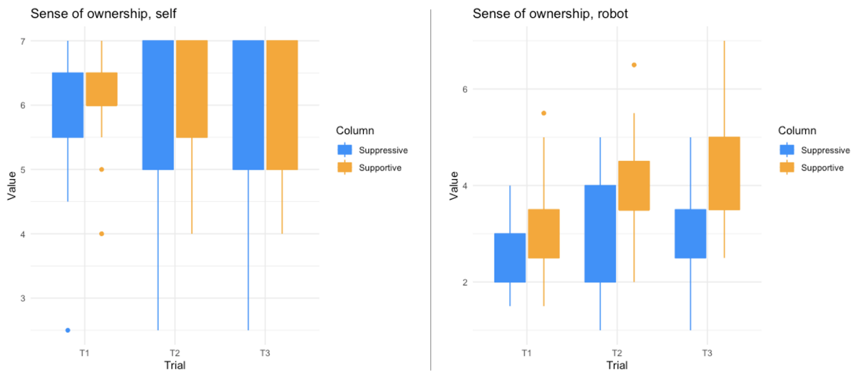


Figure 5.4: Sense of Ownership over trials between the supportive and suppressive conditions. On the left, participants were performing the task with their own hand, on the right with the robot hand.

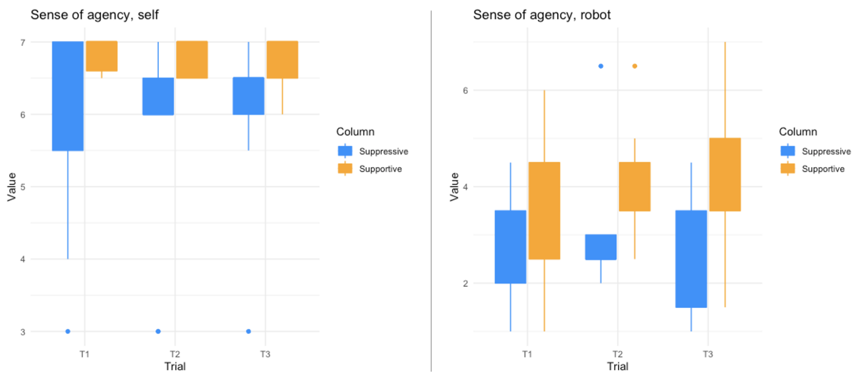


Figure 5.5: Sense of Agency over trials between the supportive and suppressive conditions. On the left, participants were performing the task with their own hand, on the right with the robot hand.

In the supportive condition at the self level, sense of ownership and task performance were found to be significantly correlated during the second trial (r=.61, df=13, p=.016) and we found a trend during the third trial (r=.49, df=13, p=.065). For what concerns the sense of agency, we found a trend during the first trial (r=.49, df=13, p=.063), a significant correlation during the second trial (r=.62, df=13, p=.01), and a trend during the third trial (r=.48, df=13, p=.07). Finally, we found

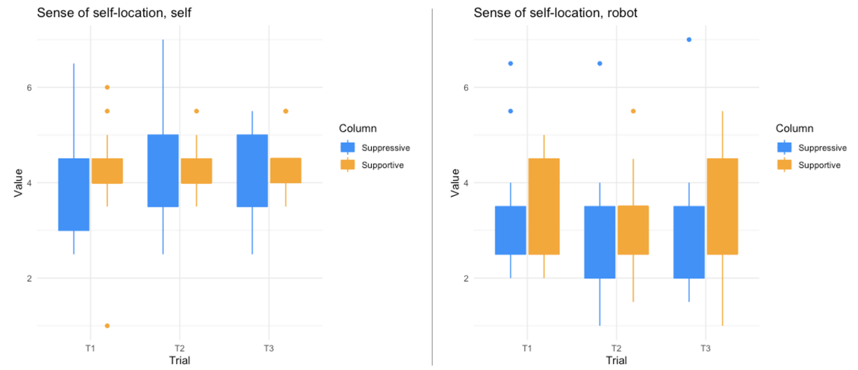


Figure 5.6: Sense of Self-Location over trials between the supportive and suppressive conditions. On the left, participants were performing the task with their own hand, on the right with the robot hand.

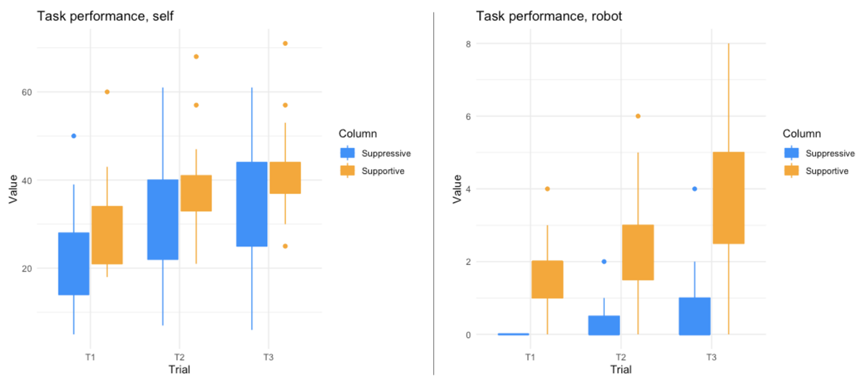


Figure 5.7: Task performance over trials between the supportive and suppressive conditions. On the left, participants were performing the task with their own hand, on the right with the robot hand.

a significantly negative correlation between sense of self-location and task performance during the first trial (r=-.58, df=13, p=.023), the second trial (r=-.79, df=13, p<.001), and a trend during the third (r=-.50, df=13, p=.06).

Moreover, we looked for correlation among the embodiment components. We found a weak correlation between sense of ownership and agency only during the third trial (r=.57, df=13, p=.02). There was a correlation between sense of ownership and self-location during the first (r=.57, df=13, p=.02) and the third (r=.73, df=13, p=.002) trials. Finally, there was a correlation between sense of agency and self-location during the second (r=.61, df=13, p=.01) and third (r=.59, df=13, p=.02) trials.

In the suppressive condition at the self level, we did not find a significant correlation or a trend between any SoE component and task performance. We only found a correlation between sense of agency and ownership only during the first (r=.62, df=11, p=.02) and third trials (r=.69, df=11, p=.009).

In the supportive condition at the robot level, we did not find a significant correlation or a trend between any SoE component and task performance. We found a correlation between sense of ownership and agency during the first (r=.78, df=13, p<.001), second (r=.87, df=13, p<.001), and third (r=.71, df=13, p=.003) trials; we also found a correlation between sense of ownership and self-location, but only during the third trial (r=.61, df=13, p=.02).

In the suppressive condition at the robot level, we did not find a significant correlation or a trend between any SoE component and task performance. We found a correlation between sense of ownership and agency during the second (r=.66, df=11, p=.013) and third (r=.81, df=11, p<.001) trials. Moreover, we found a correlation between agency and self-location during the first (r=.54, df=11, p=.05) and second (r=.82, df=11, p<.001) trials, and a weak trend during the third one (r=.51, df=11, p=.07).

To summarize, within subjects, when participants where using their self hand, they evaluated the SoE components significantly higher compared to the robot hand level, and they also performed significantly better. However, among the three trials, we found a significant improvement only in task performance at both levels and conditions, but not for what concerns the evaluation of the embodiment components. Finally, between subjects, at the self hand level, we found that participants attributed a significantly higher score only to the sense of agency, while performing the task in the supportive condition than in the suppressive one. At the robot level, instead, when participants were in the supportive condition, they evaluated the SoE components significantly higher compared to the suppressive condition, and also the task performance resulted significantly better.

Value	Level	Hand type	df	SD	F-value	p-value
Ownership	Supportive	Self	42	1.06	0.30	.74
Agency	Supportive	Self	42	0.41	0.30	.74
Self-location	Supportive	Self	42	1.01	.04	0.96
Task Performance	Supportive	Self	42	11.73	3.67	.03
Ownership	Suppressive	Self	36	1.37	0.09	.91
Agency	Suppressive	Self	36	1.22	0.120	.89
Self-location	Suppressive	Self	36	1.10	0.59	.56
Task Performance	Suppressive	Self	36	142.42	2.45	.10
Ownership	Supportive	Robot	42	1.42	2.37	.10
Agency	Supportive	Robot	42	1.44	1.37	.27
Self-location	Supportive	Robot	42	1.19	0.06	.94
Task Performance	Supportive	Robot	42	1.70	4.72	.01
Ownership	Suppressive	Robot	36	1.10	0.77	.47
Agency	Suppressive	Robot	36	1.11	0.10	.91
Self-location	Suppressive	Robot	36	1.41	0.43	.65
Task Performance	Suppressive	Robot	36	0.78	2.58	.09

Table 5.1: Overview of the One-way ANOVA results of Sense of Embodiment components and Task Performance, over time (from trial 1 to trial 3) in both conditions and levels.

Value	Hand type	Suppressive	Supportive	df	SD	t-value	p-value
Ownership	Self	5.70	5.89	26	1.07	0.45	.65
Agency	Self	5.87	6.73	26	0.77	2.97	.01
Self-location	Self	4.17	4.08	26	1.05	0.22	.82
Task Performance	Self	36.11	2.62	26	12.63	1.60	.12
Ownership	Robot	2.74	3.92	26	1.42	2.19	<.001
Agency	Robot	3.06	3.69	26	1.18	1.40	<.001
Self-location	Robot	3.17	3.11	26	1.31	0.11	.02
Task Performance	Robot	2.62	0.36	26	0.90	6.62	<.001

Table 5.2: Overview of the dependent t-test results of the Sense of Embodiment components and Task Performance between subjects (suppressive - supportive) at both levels (self - robot hand.

Value	Level	M Self	M Robot	df	SD	t-value	p-value
Ownership	Supportive	5.89	3.99	14	1.71	4.44	<.001
Agency	Supportive	6.73	3.69	14	1.18	9.95	<.001
Self-location	Supportive	4.08	3.11	14	1.04	3.59	.003
Task Performance	Supportive	36.11	2.62	14	11.21	11.57	<.001
Ownership	Suppressive	5.70	2.74	12	1.41	7.55	<.001
Agency	Suppressive	5.866	3.06	12	1.56	6.47	<.001
Self-location	Suppressive	4.17	3.17	12	1.31	2.75	.018
Task Performance	Suppressive	28.46	0.36	12	13.71	7.39	<.001

Table 5.3: Overview of the dependent t-test results of the Sense of Embodiment components and task performance within subjects (self - robot hand) in both groups (suppressive - supportive).

5.4 User Study 2

In the findings gleaned from User Study 1, a noteworthy advancement in task performance was evident during trials where the embodiment was intentionally designed to be supportive. To further probe the dynamics of teleoperation, we aimed to determine whether the sense of embodiment could be manipulated not only through perceptual cues but also at the level of the control system architecture. In this instance, the experimental setting shifted to an industrial scenario context, featuring an avatar less visually human-like than in the previous experiment—a robotic arm with an attached peg serving as the end-effector.
To explore this, we introduced a bilateral impedance reflection (BIR) control system, strategically designed to address time delay challenges in teleoperation scenarios. This was set in contrast to a bilateral impedance control with a passivity layer (BICP), the prevailing system architecture in use in the literature. The significance of this study lies in its demonstration that embodiment can be intentionally shaped at various levels within the system setup and architecture. Furthermore, it contributes novel insights into the intricate relationship between Sense of Embodiment (SoE) and task performance, underscoring the malleability of embodiment across different teleoper-

ation configurations.

5.4.1 Method

Force feedback generally has a positive effect on the telemanipulation experience and sense of embodiment of the operator. However, systems with force feedback are vulnerable to time delays, reducing their transparency and stability. The BIR was implemented to deal with delays compared to a more unstable BICP. Under delays conditions, the BIR represented the supportive embodiment condition compared to the BICP that represented the suppressive one. Three time delay groups (0, 10, and 20 ms one-way delay) of 10 participants each executed the task with both controllers.

5.4.1.1 Participants

30 right-handed participants were sampled from staff and students, between 20 and 30 years old, 8 females and 22 males, and they were equally distributed in three groups. The ethics committee of the University of Twente approved user study 2 (RP 2021-110).

5.4.1.2 Setup and Materials

We implement a classical bilateral impedance controller with passivity layers (BICP) and a bi-directional impedance reflection controller (BIR). In this method, the impedances of the operator and the environment are estimated and reflected back to the remote robot and haptic interface, respectively. A trajectory predictor is added to compensate for the delayed motion and we implemented a rigid impedance model. The operator side of the experimental setup can be seen in Figure 5.8. On the local side, there was a Haption Virtuose haptic device, that can measure the positions of the operator and apply force feedback in 6 degrees of freedom. The operator could see the remote environment via a screen, through a webcam added on site in order to provide a first person perspective on the environment. Participants had to operate the Franka Emika arm that was positioned on the right side of the interaction platform, from the operator's perspective. The setup at the remote side can be seen in Figure 5.8. It consisted of a workspace with six 3D printed blocks with a blue button on the top, equally distanced in rows, and 4 white 3D printed blocks with a blue button on the side facing each other. A led was attached to each button. These buttons and LED's were attached to an Arduino MEGA to manage the LEDs power on or off. On the operator and remote side, two computers were both running the control software. A local network was used to connect them. ROS was used as middleware, which also handled the communication between the two computers. The messages between the local and remote site was differently delayed for each condition.

5.4.1.3 Task

We asked to accomplish a task that we called *Peg-on-button*: six cubes with a button on the top were placed in the work-space in a grid. Participants had to press the button

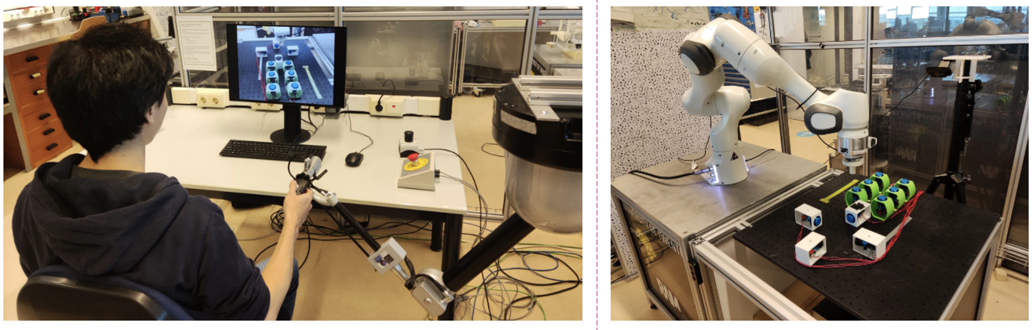


Figure 5.8: From User Study 2, on the left, the operator's side of the experimental setup, via a screen the operator can see the remote site and control the robot using the Haption Virtuose. On the right, the remote side of the experimental setup, where the Franka Emika arm is controlled equipped with a force sensor.

that enlightened. After they pressed the button, another random button highlighted. The task ended after two minutes.

5.4.1.4 Procedure

Participants were asked to fill out the consent form and then they were given detailed instructions about the experiment and the tasks. Participants were also instructed that if they were unsure how the task worked they could ask the experimenter for help. The experiment duration was approximately 30 minutes. After the accomplishment of the task with one controller, participants were required to fill in the questionnaire. Finally, they were asked to repeat the same task with the other controller and to fill in the questionnaire again.

5.4.1.5 Measures

SoE components were measured through the evaluation of the items presented in a customized version of the embodiment questionnaire from [198], the same adopted in user study 1. For each participant and each set of tasks, sense of ownership, agency and self-location were defined as the average score for the items addressing those variables in the questionnaire. Task performance was defined as the number of correct buttons touched during the peg-on-button task. Instead of comparing raw completion times, we considered measuring efficiency metrics, such as the number of correct actions or outputs per unit of time. This approach accounts for variations in time and focuses on the effectiveness of completing the task regardless of the delays.

5.4.2 Results

We implemented a 2x3 mixed-design ANOVA model, with one independent variable within subjects (BIR and BICP controls) and one independent variable between subjects with three levels (0ms, 10ms and 20ms delay). We observed the effect of this manipulation on four dependent variables: the SoE components (the sense of ownership, agency, self-location), and task performance. We found a significantly higher

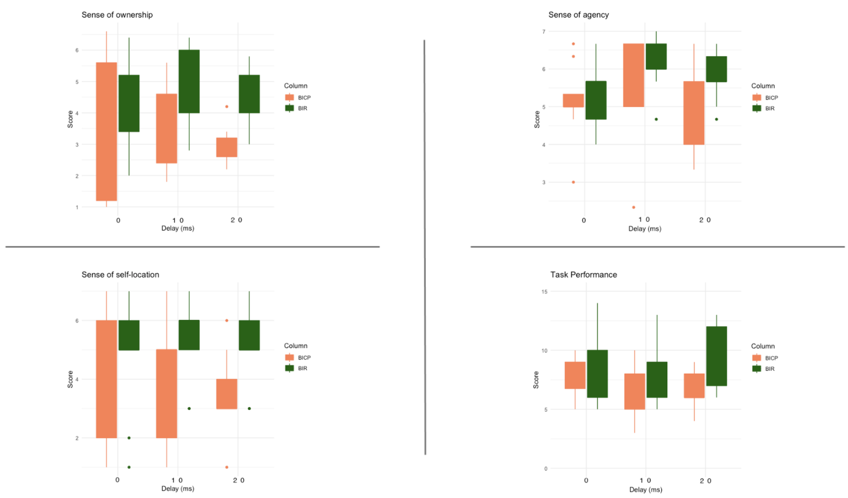


Figure 5.9: Plots of the embodiment components scores and task performance in the three delay conditions (0ms, 10ms, 20ms) compared for both controllers (BICP and BIR).

mean score of the BIR control over the BICP for the embodiment and for the task performance in the delayed conditions (see Figure 5.9 and Table 5.4). There was a significant difference between the two controls in ownership ($F(2, 25) = 40.77$, $p < .001$), agency ($F(2,25) = 5.61$, $p = .03$), self-location ($F(2, 25) = 20.61$, $p < .001$), and task performance ($F(2, 25) = 9.31$, $p = .005$), but not a significant differences across the three groups who experienced different time delay conditions. We also found a significant interaction between controls and delay ($F(2, 25) = 7.67$, $p = .003$) in task performance. Following up this interaction indicated that there was no significant difference between controls at the 0 delay condition, but the difference was significant in the other two conditions. The mean scores of the BICP decreased over time.

We did not find a significant correlation neither among SoE components, nor between SoE components and task performance, in none of the delay conditions.

The results indicate that in case of delay, for almost all the evaluated items (i.e. the components of the embodiment and the performance of the task), the operators preferred the BIR over the BICP control and achieved a better performance in the peg-on-button task.

Measure	Mean Square	Partial η^2	F-value	p-value
Ownership	18.55	.62	40.77	<.001
Agency	4.76	.183	5.61	.03
Self-location	20.61	.42	18.17	<.001
Task Performance	12.92	.380	7.67	.003

Table 5.4: Overview of the pairwise comparison between the BIR and BICP controllers for the sense of embodiment (represented by the sub-scales: ownership, agency, and self-location), and the task performance (peg-on-button was the only designed to check and compare task performance).

5.5 User Study 3

In User Studies 1 and 2, a consistent enhancement in task performance was noted when the embodiment was intentionally designed to provide support. Acknowledging and addressing the limitations identified in a telerobotics setup, we initiated a third user study in Augmented Reality (AR): 1) to better support the assumption that the SoE improves task performance; 2) to explore the role of the mediator; and 3) to support a within subjects design to investigate the SoE in teleoperation scenarios, sustained by the idea that the SoE does not drastically drop in absence of emotional stimuli. This study aims to closely examine the factors influencing task performance in an active Rubber Hand Illusion (RHI) scenario, without the added complexity of telemanipulating a robotic device.

5.5.1 Method

We performed a within-subjects design user study in VR, in which we manipulated visuo-proprioceptive information (by adding and removing delay) and the virtual avatar's hand location (shifted right, forward, or in the same position of the operator's real hand). We assessed SoE through the combination of surveys and proprioceptive measures, as explicit and implicit measures, respectively.

5.5.1.1 Participants

We recruited 30 participants (17 females and 13 males, between 18 and 28 years old) from the student pool. The sample size was determined on the basis of previous similar studies that we found in the literature [167, 233, 234, 251]. The study was approved by the Institutional Review Board of Princeton University (reference number: IRB#14912). Participants received institutional credits for their participation.

5.5.1.2 Setup and Materials

Participants viewed a virtual scene through a head-mounted display (HMD), the HTC Vive. The HTC Vive offers a 110 ° field of view, a maximum refresh rate of 90 frames per second, and a combined resolution of 2160 × 1200 pixels (1080 × 1200 pixels per eye). The scene consisted of controlling a virtual floating right hand covered by a glove and holding a Vive controller, in a first person player perspective. The main component of the virtual environment to which participants interacted, consisted of a light gray desk with, depending on the task, either five dots of different colors or a 5x4 grid. The virtual desk was calibrated to be at the same height of a real table that was placed in front of the participants in the experiment room while performing the tasks. This was done to create a mixed haptic feedback between the virtual desk and the real desk at which the participants were sitting during the experiment (see Fig. 5.10 for an overview of the setup and the VR environment). The project was created in Unity 2019.2.17f1 and Visual Studio 2019. The scene was visualized using SteamVR 1.15.19 and the SteamVR Unity Plugin 2.6.1., which also provided the hand model holding the Vive controller. The latter was used to control the virtual hand that

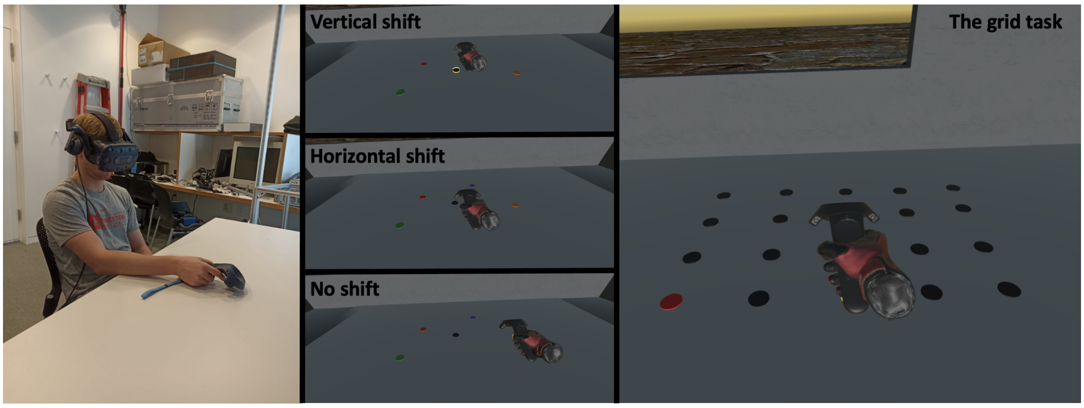


Figure 5.10: On the left, the operator performing User Study 3, in the middle we can observe the three manipulated hand shifts, and on the right we can observe extracted frame from the grid task.

was either vertically or horizontally shifted by 5cm, or the position corresponded to the self-one hand of the operator.

5.5.1.3 Tasks

Participants were instructed to perform both a *reaching task* and a *proprioceptive judgment task*. For the first task, participants were asked to touch with the controller hold both in their hand and in the controlled avatar's virtual hand each dot that they observed on the virtual table in a random order. They were required to repeat the task two times, and the second time they could not see the virtual hand while manipulating it. Then, participants were asked to grab the controller with the left hand, and to accomplish the proprioceptive judgment task by using the right hand to find the five targets below the real desk. They were asked to point to the target with the right index. Then, they were required to identify the position of their index finger below the real table with the top of the controller placed above the table. Next, participants were required to accomplish the *grid task*. Participants were asked to touch the red dot every time they spotted one in a 5x4 grid of black dots. When they touched the red dot, it became black and another random one became red. The task lasted 5 minutes. After this task, they were asked to perform the blind reaching task and the judgment task. In this way we could measure the proprioceptive drift in the same condition and between conditions, and we could also observe the effect of the SoE manipulation on task performance.

5.5.1.4 Procedure

Participants were asked to fill the consent form and then they were given detailed instructions about the experiment and the tasks. Participants were also instructed that if they were unsure how the task worked they could ask the experimenter for help. The experiment duration was approximately 75 minutes. For the baseline, participants accomplished the reaching task by seeing the virtual hand, and then they repeated it without seeing the virtual hand. Later, they were asked to accomplished the judgment task. Following, they accomplished the grid task, and finally they were required to accomplish again the reaching task and the judgment task without seeing the virtual hand. In between each condition, participants were required to fill a survey. The conditions were presented in a random order and participants had to experience each condition, therefore they repeated the set of tasks six times.

5.5.1.5 Measures

To measure SoE, we adopted a reduced version of the embodiment questionnaire from [198] (the same presented in the previous studies). Participants were asked to assess items on ownership, agency, and self-location. Moreover, we introduced four questions to assess the level of perception of the mediator. The items were evaluated using a Likert-scale from 1 (*strongly disagree*) to 7 (*strongly agree*): 1) I was so immersed that I forgot I was experiencing the virtual avatar and environment through a setup; 2) the experiment setup allows to perceive every single stimulus as I would perceive it with my own body; 3) I feel I went through a process of motor adaptation, and that I improved my motor skills over the virtual avatar; 4) every sensory and physical prediction on the interaction between the virtual avatar and the environment was correct. As for the previous studies, task performance was determined measuring efficiency metrics, such as the number of correct actions or outputs per unit of time. This approach, as reported above, accounts for variations in time and focuses on the effectiveness of completing the task regardless of the delays.

5.5.2 Results

We conducted a 2x3 ANOVA to examine the effect of delay and spatial mismatch on six dependent variables: the SoE components (sense of ownership, agency, and self-location), the mediator perception, the task performance, and the proprioceptive information of the participants in both a reaching task and a distance judgment task. We manipulated two independent variables: the synchronicity of the virtual hand response (synchronous, asynchronous), and the virtual hand shift with respect to the real hand position of the participant (right shift, forward shift, no shift).

Starting from the qualitative data, the results of the analysis for the sense of ownership show a significant effect of conditions on the outcome variable ($F(5, 29) = 3.67$, $p = .003$). These results suggest that there is a significant difference of condition on SoE score. A Tukey's HSD post-hoc test was conducted to determine which conditions differed significantly from each other. Post-hoc comparisons revealed that the sense of ownership in the most supportive condition of the SoE SNS (synchronous with no hand shift) was evaluated significantly better than in the most suppressive conditions of the SoE, namely ASV (asynchronous, forward shift) ($p = .02$) and ASO ((asynchronous, right shift) ($p = .04$), while the difference between the SSV condition (synchronous, forward shift) and ASV approached statistical significance ($p = .07$). No other pairwise comparisons reaches statistical significance.

The analysis for the sense of agency show a significant effect of condition on the outcome variable ($F(5, 29) = 2.66$, $p = .02$). These results suggest that there is a significant difference between the conditions effect in terms of the outcome variable.

The Tukey's HSD post-hoc comparisons revealed that the sense of agency in the SNS condition was evaluated significantly better than in ASV ($p = .02$), ASO ($p = .02$), while the difference between the ANS (asynchronous, no shift) condition approached statistical significance ($p = .06$). No other pairwise comparisons reached statistical significance.
The sense of self-location did not show a significant effect of conditions on the outcome variable ($F(5, 29) = 1.20$, $p = .36$).
Finally, for what concerns the mediator perception, the analysis shows a significant effect of condition on the outcome variable ($F(5, 29) = 3.63$, $p = .004$). These results suggest that there is a significant difference between the conditions effect in terms of the outcome variable. The Tukey's HSD post-hoc comparisons revealed that the setup in the SNS condition was perceived significantly less than in ASV ($p = .01$), ASO ($p = .02$), while the difference between the ANS condition approached statistical significance ($p = .07$). No other pairwise comparisons reached statistical significance.

The quantitative data referred to the task performance during the grid task, and the proprioceptive data collected during the reaching task and the distance judgment task. Starting from the task performance, an ANOVA was conducted to examine the effect of the conditions manipulation on the task performance. The analysis revealed a significant main effect of the manipulation on the dependent variable ($F(5, 29) = 17.2$, $p < .001$). A Tukey's HSD post-hoc test was conducted to further investigate the nature of these differences. The SSV condition displayed a significantly higher mean on the dependent variable compared to the asynchronous conditions ASV ($p < .001$), ASO ($p < .001$), and ANS ($p = .004$), indicating a substantial difference between these two conditions. Similarly, the SSO condition demonstrated a significantly higher task performance compared to the ASV ($p < .001$), ASO ($p < .001$), and ANS ($p < .001$) conditions, indicating a notable distinction also between these ones. The SNS condition exhibited the highest mean on the dependent variable compared to ASV, ASO, and ANS, with a significant difference of $p < .001$ for the three comparisons. To summarize, participants performed significantly better in all the synchronous conditions, while the hand shift did not play a role.

An ANOVA was conducted to examine the effect of the conditions manipulation on the proprioceptive information of participants. In the reaching task, we compared the distance differences: the Euclidean distance between the baseline points reached by the participants while seeing the virtual hand (BVH) and the baseline points reached by the participants while not seeing the virtual hand (BNVH), BVH and the points reached by the participants while not seeing the virtual hand at the end of the experiment session (NVH), and BNVH and NVH. We averaged the distances for the five targets, in order to get one unique value of the distance. We found a significant effect on the outcome variable only in the synchronous conditions when the hand was shifted either to the right or forward. Particularly, for the SSV condition ($F(5, 29) = 19.04$, $p < .001$) and SSO condition ($F(5, 29) = 11.69$, $p < .001$), we conducted a Tukey's HSD post-hoc test to further investigate the nature of these distance differences. In both cases, the BNVH-BVH demonstrated a significantly smaller difference compared to both the BVH-BNVH ($p < .001$) and BVH-NH ($p < .001$).

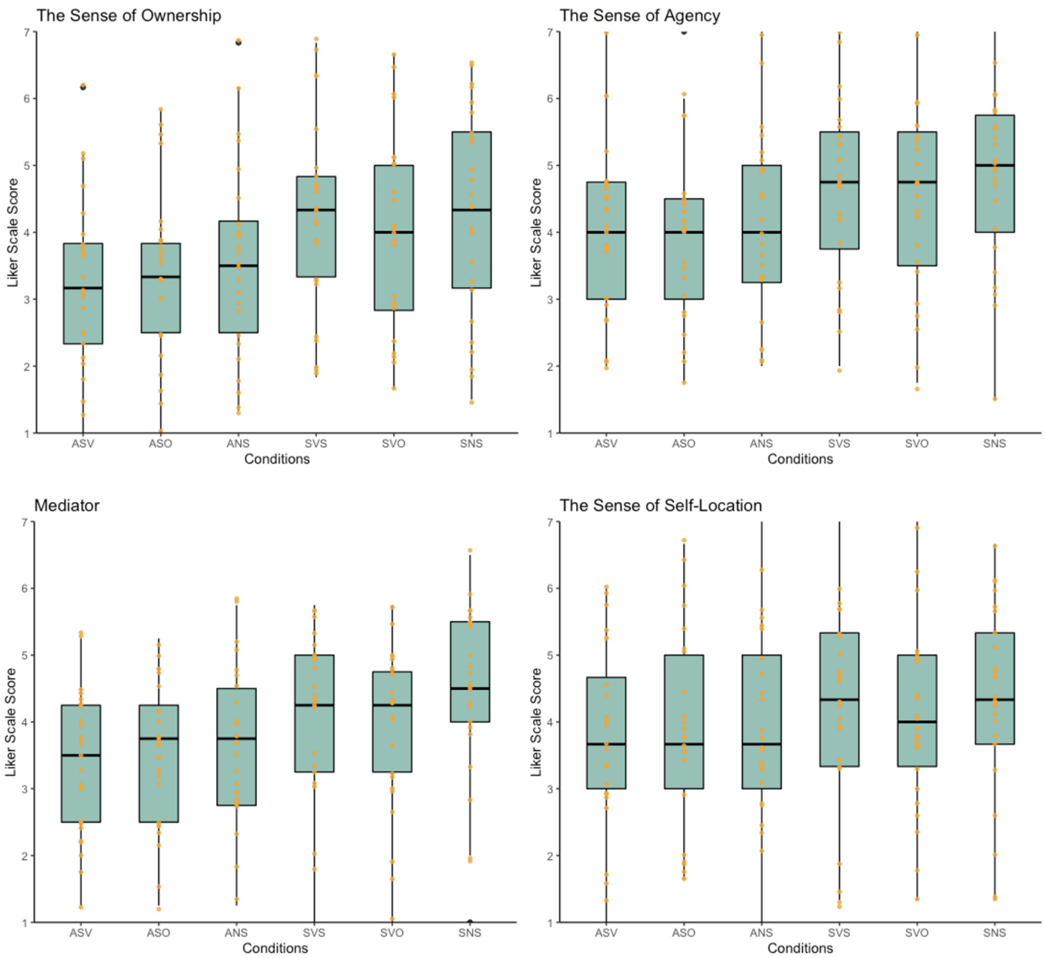


Figure 5.11: The plots represent the evaluation of the sense of ownership, agency, self-location and mediator among the six conditions. *Legend*: ASV = Asynchronous, shifted vertically; ASO = Asynchronous, shifted horizontally; ANS = Asynchronous, no shift; SSV = Synchronous, shifted vertically; SSO = Synchronous, shifted horizontally; SNS = Synchronous, no shift.

Finally, for the distance judgment task, we compared the Euclidean distance between the baseline points reached by the participants at the beginning of the experiment session (DJB) and the points reached by the participants at the end of the experiment session (DJ) among conditions. We averaged the distances for the five targets, in order to get one unique value of the distance. The analysis revealed a significant main effect of the dependent variables manipulation on the independent variable ($F(5, 29) = 4.46$, $p < .001$). A Tukey's HSD post-hoc test revealed that the distance from the baseline was much higher in the ASV condition compared to all the other conditions ($p < .001$). We did not find other significant effect among the other conditions.

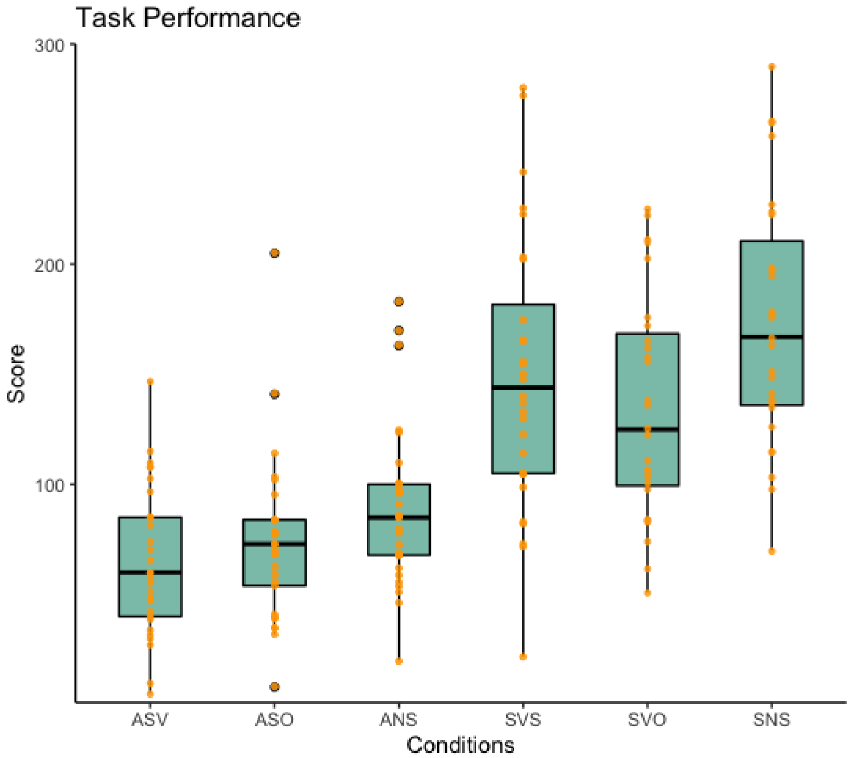


Figure 5.12: The plot represents the task performance among the six conditions. *Legend*: ASV = Asynchronous, shifted vertically; ASO = Asynchronous, shifted horizontally; ANS = Asynchronous, no shift; SSV = Synchronous, shifted vertically; SSO = Synchronous, shifted horizontally; SNS = Synchronous, no shift.

5.6 General Discussion

We hypothesized a positive effect of SoE on task performance (faster learning when SoE is higher)(H1). On the basis of the results, we accept H1. In study 1, Figure 5.7 and Table 5.1 show a better task performance in the condition in which the perceptual cues support a high SoE, i.e. when the level of embodiment is designed to be high, we observe a positive learning effect and a significant better performance over time. In the condition designed to provide a low level of SoE, instead, we observe a weak and non-significant learning effect over trials. Moreover, between subjects, we observed a significant higher task performance of the supportive embodiment group compared to the suppressive one while performing the task with the robot hand. In study 2, we observed that by increasing the delay, participants perceived the proposed BIR control as providing higher SoE than the BICP control (see Figure 5.9). Moreover, we can also observe a significant better performance by using the BIR in delayed conditions. In study 3, we observed that the more the condition was designed to support the SoE, the better the resulting task performance was.

However, the interpretation of our results does not appear to be so straightforward. Some observations brought us to the need the design an ultimate user study to test if the operators can achieve the same task performance in both a supportive and suppressive embodiment, with the hypothesis that in the latter case they would

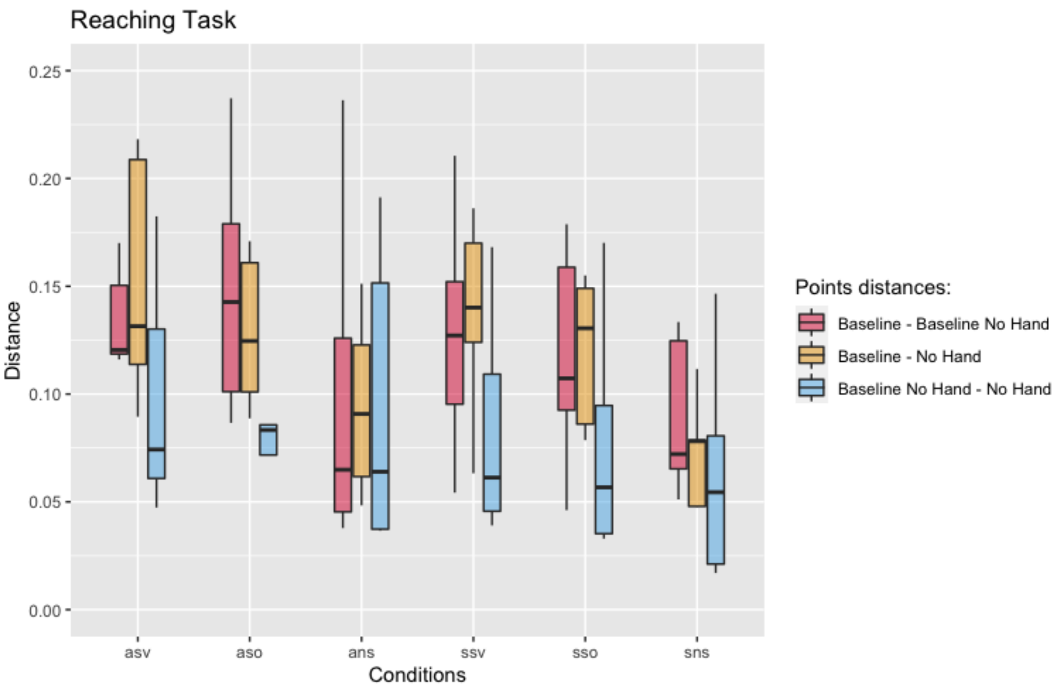


Figure 5.13: The plot represents the distances between the points reached by participants among the three phases of data collection for the reaching task. We average the distances of the five targets to get one unique distance value. Each bar, as reported in the legend on the right of the plot, represents a point difference: the difference between the point reached during the baseline when the hand was visible and the point reached during the baseline when the hand was not visible; the difference between the point reached during the baseline when the hand was visible and the point reached when the hand was not visible after the grid task at the end of the experiment session; finally, the distances between the point reached during the baseline when the hand was not visible and the point reached when the hand was not visible at the end of the experiment session. This set of data is reported for each condition. *Legend*: ASV = Asynchronous, shifted vertically; ASO = Asynchronous, shifted horizontally; ANS = Asynchronous, no shift; SSV = Synchronous, shifted vertically; SSO = Synchronous, shifted horizontally; SNS = Synchronous, no shift.

need more time to adapt to the system. Particularly, in user study 1, the most significant correlations were found when perceptual cues support achieving a high SoE. We found significant positive correlations between SoE components and task performance, but no correlation was found when perceptual cues suppressed SoE. However, in case of a full dependency between SoE and task performance, we would have expected to find a negative correlation in the suppressive embodiment condition. In fact, in the suppressive condition, even if the learning effect is not significant over the three trials, we can observe a trend that suggests an increasing performance over time (see Figure 5.7. Moreover, between subjects, we did not find a significant difference between the task performance of the supportive and suppressive groups while performing the task with the self hand. One explanation could be that while using their own body, even if the perceptual cues are negatively affected, the learning rate is already so high that people reach the same plateau in task performance that they

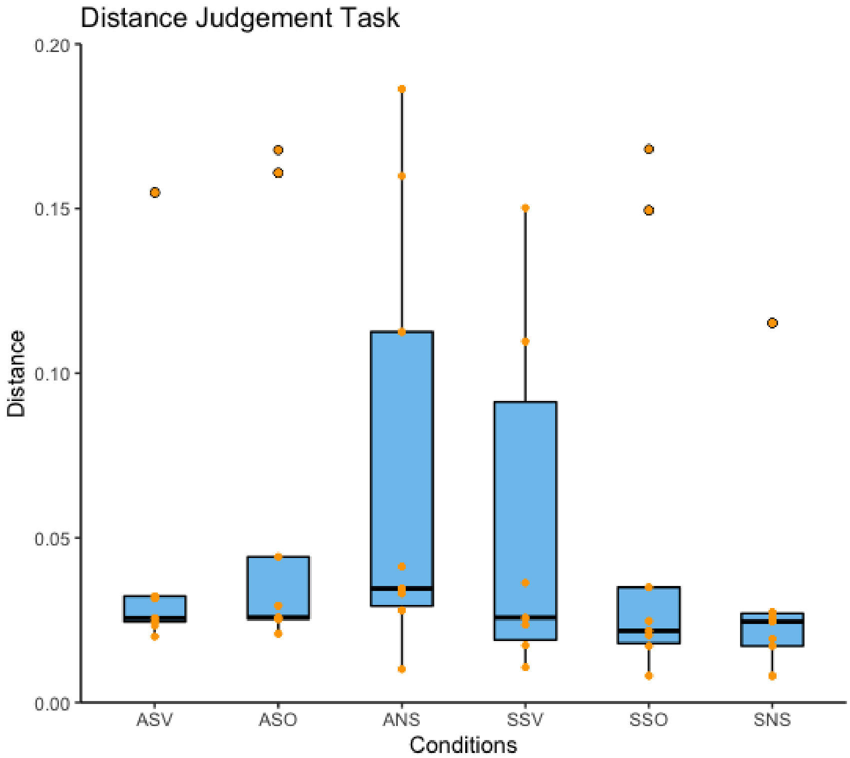


Figure 5.14: The plot represents the distances between the points reached by participants between the baseline and at the end of the experiment session for the distance judgment task. We average the distances of the five targets to get one unique distance value. These data are reported for each condition. *Legend*: ASV = Asynchrnous, shifted vertically; ASO = Asynchrnous, shifted orizontally; ANS = Asynchrnous, no shift; SSV = Synchrnous, shifted vertically; SSO = Synchrnous, shifted orizontally; SNS = Synchrnous, no shift.

would reach in the most supportive condition (i.e., using their own body without any mediator). For what concerns the observations in the other user studies, in user study 2, we did not find a significant correlation between the scores of the embodiment components and task performance. In study 3, we found a significant higher task performance for the most supportive embodiment conditions compared to the least supportive ones. However, these results were not that strong and consistent for the qualitative measures of embodiment, even if we could observe a consistent trend: the scores attributed to assess the SoE components and the mediator perception were higher for the conditions supporting embodiment.

This brings us to H2, namely a positive effect of SoE on the overall operator's experience (i.e., an increasing evaluation of the embodiment components and a decreasing perception of the surrogate as a mediator). On the basis of our results, we can only partially accept H2. Starting from user study 1, we could not observe a significant different among the scores attributed to the SoE components between subjects in the supportive and suppressive conditions at the self hand level, and neither at the robot hand level. The only exceptions are the sense of agency at the human level, which was scored significantly higher by the supportive condition group, and the sense of ownership at the self hand level, that was scored significantly higher by the supportive condition group. At the robotic hand level in the supportive group, we can also observe an increasing feeling of ownership and agency over the robot surrogate over trials within subjects (see Figure 5.4 and 5.5), but the differences in scores are not significant. Another interesting observation was that the correlation between self-location and task performance was negative, while participants performed the task in the self hand condition. This is probably due to the setup: participants had to accomplish the task while experiencing the workspace through a teleoperation setup even when they were using their own hand. Therefore, while they could perceive their own body as usual, their visual perception and proprioception were affected by the setup. Over trials, they improved their task performance but they poorly scored sense of self-location items. For what concerns user study 2, in the delayed conditions, namely in the conditions designed to test the if our controller (BIR) could improve the SoE experience and task performance, we observe a significant higher score of the embodiment components while using the BIR than the BICP controller. Finally, in user study 3 we could observe that the condition designed as the most supportive one of the SoE was scored significantly higher than the most suppressive ones for all the independent variables (sense of ownership, agency, and mediator), with the only exception of self-location that did not report any significant effect of the manipulations.

To answer the explorative RQ (Are these results robust over different scenarios and tasks?), SoE experience and task performance seem setup and task dependent. However, there are some common observations, such as a better task performance while using a setup designed to create a supportive embodiment experience compared to a suppressive embodiment one. Across experiments, through participants' feedback and our observations, we could conclude that the context of application and the setup play a role in the SoE experience. For example, based on the presented literature, we assumed that a VR setup would have been more immersive than a telerobotics one, and one of the reasons is the easiness of use of the former compared to the latter. For example, in the study 3, in which the usage of the setup was pretty easy to learn compared to the studies 1 and 2, we expected a more immersive experience that would have created a faster and stronger connection between the operator and the avatar, and it would have decreased the perception of the least as mediator. As consequence, operators would have been less aware of their own body and the proprioceptive system would have managed to efficiently integrate the information provided by the surrogate, and to update the body schema. However, what happened was that participants were mostly focusing on using the avatar as a tool and they spent less time focusing on the entire embodiment experience. We learnt that also the task and the goal also have an effect on the operators' embodiment experience. There is evidence that there is a difference between a classic embodiment illusion, such as the RHI, in which the participants have a passive experience and focus all their attention on the perceived stimuli and the avatar, compared to an active experience such as in a teleoperation setup in which a big part of the attention span is on the task accomplishment and in the motor learning experience. In those cases the priority is given to the task, and the SoE experience becomes secondary. The proprioceptive data of the User Study 3 allowed us to support this assumption. Indeed, we could observe that even when we manipulated the visuo-proprioceptive information and the avatar's hand location, the participants did not updated their body schema. They just used the hand as a tool to accomplish the task, and they were facilitated in the SoE supportive conditions compared to the suppressive ones. Looking at the differences of the reached points in the space for the reaching task, we could observe that the smallest difference was between the baseline without seeing the avatar's hand and the same task repeated at the end of the session, after the grid task in which they got used to the hand (see Figure 5.13. This brought us to the conclusion that our manipulation, even if it had an effect on task performance, was not enough to affect the body schema of the participants. In the distance judgment task, the observations lead us to the same conclusions. Indeed, Figure 5.14 shows that there was not a significant difference among conditions and between the baseline and the task repeated at the end of the experiment session. We could just observe a bigger variance for the asynchronous condition in which the avatar's hand was not shifted, and for the synchronous condition in which the avatar's hand was shifted forward. Indeed, these are the only two conditions that challenged the participants in the processing of the proprioceptive information coming from the virtual environment, while the others were too exasperated for body schema, or they mostly matched participants' hand location and motor response.

5.7 Conclusions, Limitations, and Future Works

The findings suggest that SoE directly influences task performance, impacting motor learning and diminishing the perception of the avatar as a mediator. A supportive embodiment appears to enhance the operator's learning experience within a setup, potentially encouraging sustained usage over time. However, the intricate relationship between SoE, task performance, and motor learning in teleoperation necessitates further exploration. Notably, we observed a positive trend in task performance even in suppressive conditions of embodiment, even though it did not reach statistical significance. Extending the task duration and increasing the number of trials would yield more observations, providing a deeper understanding of how the manipulation of embodiment affects task performance. Crucially, this extended investigation could clarify whether operators in suppressive conditions can eventually attain the same level of performance as those in supportive conditions, given additional training and adaptation time.

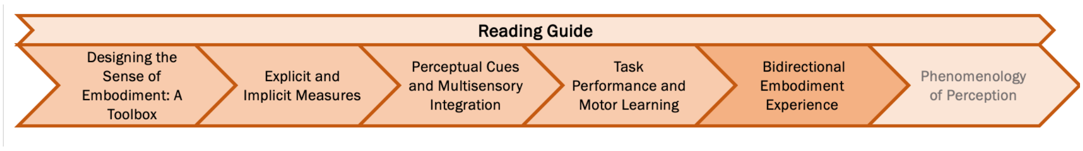
In future works, we aim to empirically examine the proposed explanation put forth in the discussion, i.e. the concept of the surrogate being perceived as a mediator, which introduce a friction in the achievement of a completely immersive embodiment experience. Furthermore, we intend to investigate the impact of manipulating the SoE on operators' motor learning over an extended series of trials, continuing until operators reach a performance plateau. This exploration will be conducted across diverse contexts of application to discern any context-specific variations. Lastly, we aspire to delve deeper into the definition of SoE, considering both its subjective and objective dimensions to provide a more comprehensive understanding of this crucial aspect in our research domain.

6

An Inner Perspective on Embodiment: the Multimodal EchoBorg, a Human Operated by a Machine

My life's work has been to prompt others and been forgotten. Remember that night when Christian came to your balcony? That moment sums up my life. While I was below in the shadows, others climbed up to kiss the sweet rose.

Edmond Rostand, Cyrano de Bergerac



Abstract

In this chapter, we present a different perspective on the concept and sense of embodiment. What if a human becomes the embodied surrogate? We present a Multimodal Echoborg interface to explore the effect of different embodiments of an Embodied Conversational Agent (ECA) in an interaction. We compared an interaction where the ECA was embodied as a virtual human (VH) with one where it was embodied as an Echoborg, i.e., a person whose actions are covertly controlled by a dialogue system. The Echoborg in our study not only shadowed the speech output of the dialogue system but also its non-verbal actions. The interactions were structured as a debate

between three participants on an ethical dilemma. First, we collected a corpus of debate sessions with three humans debaters. This is what we used as a baseline to design and implement our ECAs. For the experiment, we designed two debate conditions. In one the participant interacted with two ECAs both embodied by virtual humans). In the other the participant interacted with one ECA embodied by a VH and the other by an Echoborg. Our results show that a human embodiment of the ECA overall scores better on perceived social attributes of the ECA. In many other respects the Echoborg scores as poorly as the VH except *copresence*.

6.1 The Multimodal EchoBorg

Many HCI researchers aim to create an 'artificial social entity' that is as human-like as possible, in both the (non)verbal behavior it exhibits and in the way its body looks. The term artificial social entities can cover a broad spectrum of technical artifacts, ranging from chatbots to virtual characters to physical social robots. In this work we focus specifically on *Embodied Conversational Agents* (ECAs). Researchers developing ECAs frequently use the Wizard of Oz (WOz, [53]) method to design and evaluate the ECA. A human operator performs the tasks of one or more components of the system that are not (yet) implemented. The person interacting with the system is tricked into believing that they are interacting with an autonomous artificial system, but in reality there is 'another person behind the curtain'. One can also imagine the complete opposite: a user is talking to a person of flesh and blood, whose decisions of what to say are made by an autonomously operating piece of software. Corti and Gillespie [47, 96] introduced this WOz variant with the term *EchoBorg* (EB): a person that speaks out the utterance generated by a chatbot. This type of illusion, where a person's utterances are fully determined by a third person, was first investigated by Milgram [177] under the name *cyranic illusion*. The name refers to the French classic play *Cyrano de Bergerac* by Edmond Rostand, where the unattractive but eloquent Cyrano covertly provides the attractive Christian with the words to woo the beautiful Roxane, by whispering the right words into Christians ears from a balcony while Christian is on a date with Roxane. Milgram found that the combination of the two persons is perceived as one identity, which he named a *Cyranoid*. He investigated how different the two identities involved in the Cyranoid could be before the illusion breaks down, for instance by a child determining the utterances of an adult. Corti et al. [47] were able to maintain the cyranic illusion, even when a chatbot determines the utterances of a human the resulting EB is perceived as one identity. Confronted with a human embodiment, a user initially has no reason to question whether this person is controlled by a system. Thus, with an EB it is possible to study the 'mind' of a conversational agent without potential bias evoked by user expectations of the capabilities of an artificial agent. A user might think "it's a machine, so it won't understand me" and as such might not display, for example, conversational repair behaviour [48]. The apparatus, or *cyranic interfaces*, used by Corti and Gillespie (and before that by Milgram) are limited to the speech modality. In this paper, we present a cyranic interface for multimodal echoborgs, extending the speech-only EB method to allow for multimodal behaviour shadowing. The Multimodal EchoBorg

(MEB) consist of an ECA system that dictates the speech, non-verbal back-channels, gaze and gestures of the human through a specialized interface. Using an MEB it is possible to study how all behaviours that are traditionally generated for a Virtual Human (VH) embodiment are perceived when users do not expect an artificial mind as they are interacting with a real person. We performed a study in which we compare the same interactions with an ECA that was either embodied as a VH or an MEB, both controlled by the same system. We examine the effect of the embodiment on the user perception of the agent in terms of concepts that are often used when evaluating artificial agents (i.e., animacy, anthropomorphism, intelligence) and the perception of the overall experience of the interaction.

In the next section, we discuss some of the literature that looks at the perception of different embodiments. Next, we describe the MEB set-up, followed by the first exploratory study.

6.2 Relationship between Perception and Embodiment

Humans interacting with others can quickly form an impression about the skills, personalities, and attitudes of others toward others. These impressions can be based on just a few seconds of observation of the other's appearance and (non)verbal behavior such as facial expressions and gestures [6, 10, 40, 151]. The effects of virtual human behavior on the perception of agent personality and interpersonal attitudes have been investigated in perceptual studies (properties of gestures [181, 236] with language [64, 182] on personality, posture [187] on emotion, gaze, and proxemic behaviors on interpersonal attitudes [139]), as well as in studies focusing on impression shaped during first encounters with virtual characters [39].

6

In addition to the appearance (e.g., hair color, height), the fact that the MEB is physically embodied by a human makes it different from the VH on a screen. Li [153] discusses studies that investigate the experience of interacting with physically co-present social robots, telepresence robots, and virtual agents. He concludes that *"robots were more persuasive and perceived more positively when physically present in a user's environment than when digitally displayed on a screen either as a video feed of the same robot or as a virtual character analog"* [153, p25]. Also in human-human communication, the shape and representation of interlocutors affect how humans respond to and perceive each other. In Bailenson et al. [17], participants engaged in a technology-mediated interaction at various levels of *behavioural and form realism*, including voice only, video conference and through simple virtual polygon avatars. The reported levels of perceived copresence and self-disclosure were affected by those conditions. For example, both verbally and nonverbally, people disclosed more information to avatars that were low in realism. One fundamental aspect of the (M)EB is that users are (at least initially) lead to believe that they are talking with an autonomous human instead of with a machine. This is called the perceived level of agency and is known to be an important predictor of how mediated social interactions are carried out. In social games, experiences are affected by beliefs about the agency of other players, and whether or not they are physically copresent. Research consistently finds that the belief that another player is human (positively) affects various aspects of experience [155, 265], such as engagement, flow, presence, enjoyment and physiological arousal. This has also been investigated from a neuroscientific perspective: Katsyri et al. [131] found that in a first-person video game, winning versus losing activates the brain's reward circuit differently depending on the belief on whether the opponent was human or computer controlled. In conclusion, a lot of evidence points towards a human, physically present interaction partner positively affects the engagement, arousal, and interactant's traits perception, over a VH on a screen.

One work that addresses the difference between how humans and agents are treated differently is that of De Melo and Gratch [58]. They propose a believability benchmark, which, according to them, requires "people, in a specific social situation, to act with the virtual agent in the same manner as they would with a real human". Based on previous research (e.g., [27, 28]), they claim that the higher the attributions of mind people make, the more likely machines are to pass the benchmark of believability. Empirical evidence suggests that, compared to VHs, humans are treated more favorably in most contexts by default. The authors' theory is that this is due to the expectations we have of the other's *mind* and *experience*. Agents need to employ additional strategies and actively display capabilities to sway the user's perception of the agent in these dimensions if they seek to match a human in *believability*.

Most of the work discussed so far addresses unilateral constructs such as the flow of the experience or perceived traits of others. However, in (mediated) social interactions, there are also bilateral constructs that emerge between the interlocutors. For example, [22] have investigated *coupling* in human-agent interactions, the bilateral impact that each interlocutor has on the other's behavior, making the interaction a dynamic and mutual flow. According to this, both MEB and VH should exhibit the same amount of interactivity. However, we may expect that a MEB is still favored in these constructs over a VH given the overall different expectations that humans have towards other humans versus from machines.

Summarizing, there is some evidence that a human (embodiment) would be favored in a number of ways over a screen-based VH embodiment - based on the physicality of the human, and based on the implied belief that a human is an autonomous conscious entity, unlike an (apparent) machine such as a VH.

6.2.1 How will the MEB be perceived?

Concepts and findings in the domain of mediated social interaction help us to understand how the interaction with an ECA embodied by a MEB might be perceived differently from the same interaction with an ECA embodied by a VH. However, given the hybrid nature of the (M)EB (mind of a machine, body of a human), the prior work does not allow for direct predictions in this regard. In previous work on EBs, the non-EB condition featured textual interfaces rather than alternative (artificial) embodiments [46, 47, 48], and as such does not provide insights on how an (M)EB might perform when compared to other embodied agents. For our present work, we compare two conversational agent embodiments with a representation of a real or virtual body, pulling the compared conditions more alike. Note that our approach is not intended as a definitive study on the effect of embodiment on conversational agent perception, but intended as a first exploration of how an ECA embodied by a

MEB is perceived in the dimensions relevant for our community and how sensitive the conventional measures are in this setup.

From the point of view of the methodology, we referred to Corti et. al [46] as benchmark. They analysed the adjectives participants attributed to the respective conversational partner. Participants used adjectives that are of artificial or inhuman nature ("mechanical", "computer", "robotic") to describe their interaction partner when interacting with the text interface, while used adjectives of a human nature ("shy", "awkward", "autistic") to describe the EB. Instead of asking participants to freely attribute adjectives, we administered them the commonly used Godspeed Questionnaire Series (GQS) [19] for evaluation of artificial agents. It uses semantic differential scales to cover similar concepts. These concepts are anthropomorphism, animacy, likeability, and perceived intelligence (and perceived safety, as a concept specific to robots). Given that these concepts in the GQS have a directionality from machine-like (low) to human-like (high), we expect that a human embodiment for our ECA, as achieved with the MEB, would be rated more favorably on these concepts.

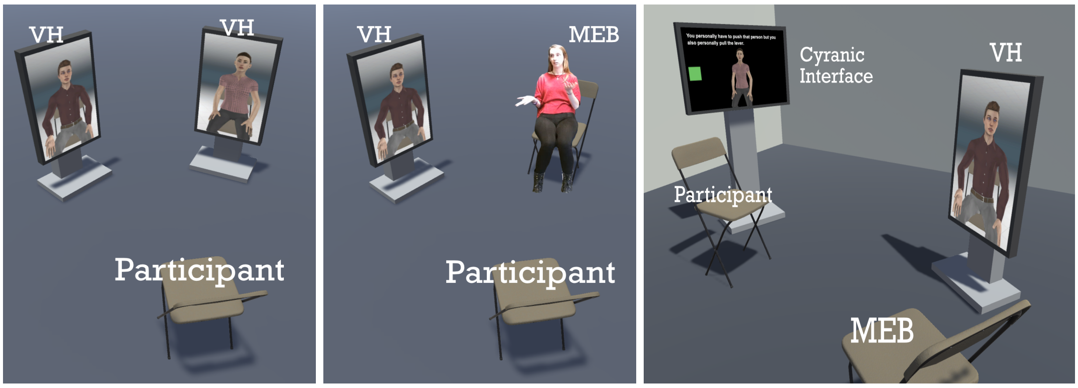
Hypothesis 1 Participants will rate a MEB higher than a VH embodied conversational agent on the key concepts: anthropomorphism, animacy, likeability, and perceived intelligence.

The discussed literature demonstrates that experiences are more engaging when participants believe they are interacting with a human than when they are interacting with a machine, even if the behaviour of the other players is otherwise equal [131, 155, 265]. This depends on the bias that humans expect more relevant social actions from other humans [192]. Based on this, we would expect that the overall engagement and flow of the interaction, as well as the emotional experience and reaction, would be better experience when interacting with ECAs embodied by the MEB, rather than a VH. To rate those aspects, we administered the Game Experience Questionnaire (GEQ) [121] for the engagement and flow, and the Self-Assessment Manikin (SAM) [30], for the emotional response.

6

Hypothesis 2 The quality of the interaction with the ECA, as reflected in constructs such as flow, arousal and engagement (as measured by the GEQ and SAM), will be rated more positively by participants when the ECA is embodied by the MEB.

In regards to the bilateral constructs such as coupling [22], it is more difficult to make a prediction. Coupling implies an evolving equilibrium among the interlocutors. It is the capability to compensate disturbances in the interaction by evolving it. This is why it is highly complex to reproduce when employing virtual agents, since it implies that they should manage to face unexpected stimuli and situations. On the basis of the coupling concept, participants should perceive the same amount of interactivity from both human and VH embodiments. Therefore, the discourse flow and engagement should be at virtual agents level for both embodiments. However, on the basis of the reported literature, we could also assume that a MEB is favored over a VH, given the different expectations and bias that humans have from other humans and from machines, that could alter the interaction perception.



(a) (b) (c)

Figure 6.1: 3D illustration of the debater placement in (a) HAA and (b) HEA conditions. The echoborg view in (c) shows how the screen with the cyranic interface was positioned, behind the participant.

Hypothesis 3 On measures regarding the bilateral relationship between the ECA and participant during the interaction (as reflected by the coupling instrument [22]), the ECA will score higher when embodied by the MEB.

6.3 A Cyranic Interface for Multimodal EchoBorgs

We designed a novel apparatus that allows for multimodal behaviour shadowing, namely speech, gestures, nonverbal back-channels, and gaze. A human shadower receives instructions on what to say and which non-verbal behaviours to display from an ECA system through the *cyranic interface* (visible in Figure 6.1(c)).

The components of the ECA For behavior realization and planning, we employ the Articulated Social Agent Platform (ASAP) realizer [259]. For rendering the Virtual ECA embodiment on screen as well as for the cyranic display, we employ the ASAP Unity bridge [138]. The dialogue scenario is modeled using the Dialogue Game Execution Platform (DGEP) [147]. For dialogue management, we use the Flipper Dialogue Engine [258].

The Cyranic Interface The instructions for a human confederate shadowing the ECA were provided in the following way. With respect to speech, we displayed the output of our dialogue system, to be uttered by the MEB, on a screen (hidden from the participants) akin to a teleprompter. In our case, the text shown on the teleprompter was the direct output of our dialogue system, which would otherwise be spoken out by the ECA using a text-to-speech (TTS) system. We explored the alternative to play audio of the utterances through hidden earpieces, more similar to the conventional speech shadowing. However, it appeared to be very difficult to shadow a TTS voice. Moreover, while the utterance selection of the system is dynamic, the ECA utterances in our user study were pre-scripted. After a bit of practice, our MEB became familiar with the utterances, and managed to shadow the speech fluently.

A simple ECA gaze behaviour model sufficed as we envisioned a triadic interaction. Therefore, we could keep the interface for gaze shadowing simple: there is a green highlight at the left or right half of the screen, indicating whether gaze should be directed to the conversation partner on the left or right (from the perspective of the MEB).

The Echoborg was also instructed to back-channel at certain times while listening. Our ECA system only includes a single type of back-channel, head nods. In the MEB interface, these behaviors are signaled by (discretely) flashing the word *nod* on the screen.

When it comes to gestures, shadowing motion and poses are challenging for the MEB. Lexical instructions for gestures are difficult to translate into fluent and animate motions that retain the semantic connection with the words uttered. As an alternative, we decided to show the motions on an animated copy of the ECA, rendered on the screen behind the participant. While ad-hoc mimicking remained difficult, we observed a learning effect, as with the speech shadowing. Because the speech and gestures generated by the system were the same for each utterance, our MEB was able to learn the speech and gestures produced by our system and was able to shadow with similar 'size' and 'stroke' from seeing the animation only in peripheral vision.

6.4 Exploratory User Study

Unlike previous work on EBs with unscripted dialogues [47], we modeled the dialogue scenario more strictly for our ECA. Besides the increased experimental control when comparing the interactions between embodiments, it also simplifies the complexity of the overall system.

We modeled an ethical debate-like scenario, with a *moderator* and two opposing debaters, the *proponent* and *opponent* discussing different variations of the *Trolley Dilemma* [243]. It is an ethical dilemma questioning about whether to sacrifice one person to save a larger number. The scenario is a runaway trolley barreling down the railway tracks. On the tracks there are five people tied up and unable to move, and the trolley is heading straight for them. A person is standing in the train, next to a lever. Pulling the lever, the trolley will switch to a different set of tracks. However, there is one person on the side track. The proponent is asked to argue for pulling the lever, while the opponent is asked to argue for staying passive. The moderator's role is to open and manage the discussion and to introduce the dilemma and its variants to the debaters before yielding the floor to them for their arguments.

6.4.1 Modelling ECA & Dialogue

To model this scenario and inform the design of utterances and gestures for the ECAs, all roles (debaters and moderator) are modeled from an only-human debate corpus. We recorded audio and video, and we transcribed the dialogues. We also measured some of the interaction experience and interlocutor perception that were also used in the user study later on.

In total, we recorded 6 triads (2 females, 16 males). From the transcriptions, the arguments used to defend the two debaters' positions (pulling the lever/being pas-

Table 6.1: The key arguments, that we classified, of the *Trolley Dilemma* debate and the moral questions that describe them.

Key arguments	Moral question
Fate	Can the fate decide for the life of human beings?
Numeric	Human life is a qualitative or quantitative matter?
Economic	Is it better to save more lives because they are a greater resource for the society?
Responsibility	If we make the choice of pulling the lever, do we become responsible for a murder?
Inaction	Can 'inaction' be considered as 'action'?

sive) were categorized by type of argument (see Table 6.1) and selected for our ECA to use as utterances. In a small survey (20 participants in SurveyMonkey), external raters ranked the selected utterances for the different arguments based on their ability to convince. This allowed us to select balanced arguments for both the proponent and opponent. For the non-verbal behaviors of the ECA, we have consulted the video recordings from the corpus of the selected utterances and presented them to an actor. The actor acted out these utterances wearing a MoCap suit. This yielded full-body gesture animations for each utterance. The MoCap recordings and selected utterances were then combined and linked to the dialogue move. As mentioned in Section 6.3, our ECA system uses the DGEP dialogue argumentation framework. DGEP uses the concept of *dialogue moves*, namely the schematic representations of a single move in a dialogue, its reply, and the connections to the argument structure.

6.4.2 Experiment Design

Participants were assigned to one of two conditions: *Human-Agent-Agent* (HAA) or *Human-Echoborg-Agent* (HEA). Participants were always assigned to the role of the moderator, while the debaters (*proponent* and *opponent*) were always acted out by our ECAs. In both HAA and HEA, the opponent was always embodied by the VH. In HEA, the proponent was embodied by the MEB, while in HAA, the proponent was also embodied by a VH. We call this between-subject variable *proponent embodiment*. For those participants assigned to the HEA condition, it is also interesting to compare their ratings of the VH embodiment of the opponent versus the MEB proponent embodiment. This is a within-subject variable which we refer to as *debater embodiment*.

6.4.3 Materials & Apparatus

The moderator and the two debaters are positioned in a triangle (see Figures 6.1(a) and 6.1(b)). VHs were shown on large TV screens in portrait mode. When the proponent was embodied by the MEB, that screen was replaced by a chair for the MEB to sit on. For the MEB's cyranic interface, a large screen was placed behind and out of sight of the participant, facing the MEB (see Figure 6.1(c)). Due to the fact that there were other screens in the experiment room, participants did not get suspicious about seeing the screen behind their chair while entering the room. Furthermore, all screens were turned off, or appeared to be turned off when participants entered the room. Therefore, they could not see the agent on the screen.

The moderators received cue-cards to guide the debate through the different variants (in any order). The cue-cards represented utterance hints that participants could rephrase and use in the order that they preferred while interacting with the two debaters. This allows the participant to participate in the interaction without affecting the conversation in an unpredictable way.

The detection of when the participant is speaking and which movements their utterances represent is done secretly by the experimenter in a WOz fashion [53].

6.4.4 Multimodal EchoBorg Training

We recruited an experienced actress from the student body to act as the MEB in this user study (see Figure 6.1(b)). Following a number of training sessions on the debate with the researchers, she became familiar with the scenarios and behaviours. Although not systematically quantifying the accuracy of shadowing, comparing recordings of MEB behaviors with VH behaviors showed that the actress was able to shadow speech and gestures reliably.

6.4.5 Participants

The Ethics board of the University of Twente approved the user study. In total, 36 participants were sampled from university staff and student body, between 19 and 46 years old, 16 women and 20 men, and the number of participants was equally distributed between conditions.

6.4.6 Measures

We selected several existing questionnaires measuring interaction experience and interlocutor perception that are commonly used in the IVA community, as discussed in Section 6.2.1. Therefore, we used the GQS to address the first hypothesis, which concerns the effect of the appearance, and the virtual or physical presence of the embodiment on the human interlocutor's perception. To address the second hypothesis, related to the effect of embodiment on interaction experience, participants completed the Game Experience Questionnaire (GEQ) [121] and the Self-Assessment Manual (SAM) [30]. Finally, to address the third hypothesis, we measured the dynamic *coupling* between the participants and the ECA embodied by both the VH and the MEB using the questionnaire from [22].

~

6.5 Results

We conducted a one-way ANOVA on the effect of the between subject variable "proponent embodiment" on each of the sub-scales of the questionnaires and dimensions described above. Two of the GQS sub-scales showed significant effects: animacy (F(1,34)= 5.834, p= 0.021, $\eta^2 p$=0.146) and anthropomorphism (F(1,34)= 20.061,

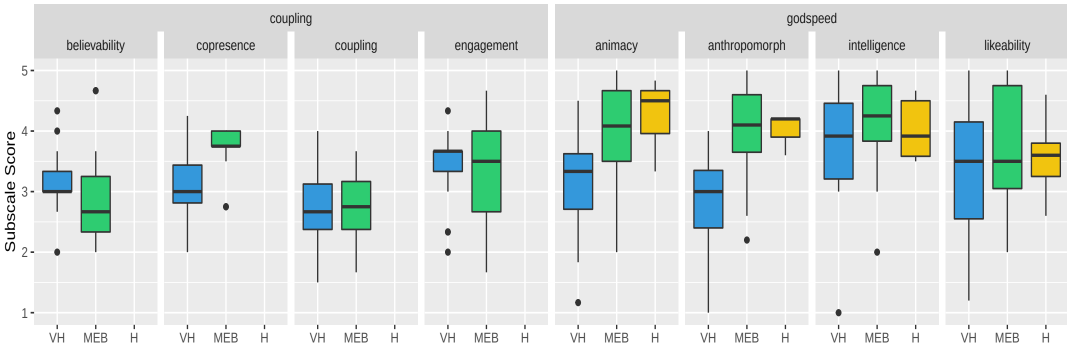


Figure 6.2: Moderator scores attributed to the proponent debater embodiments (between subjects), also showing the moderator scores for the proponent in the Human-only pre-study corpus.

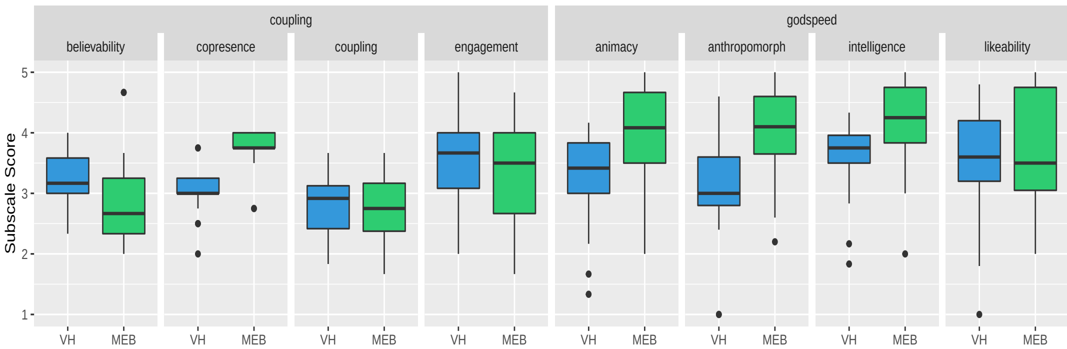


Figure 6.3: Moderator scores attributed to the different debaters (within subject) in the HEA setting (Virtual Human acting as Opponent, EchoBorg acting as Proponent).

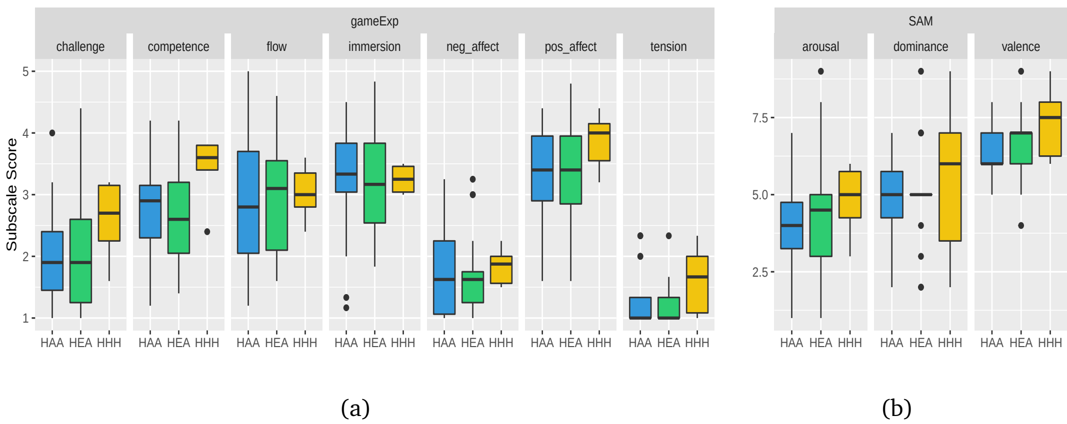


Figure 6.4: Scores on moderator game experience (a) and SAM self-report scales (b) between the different combinations of debaters (HAA = only agents, HEA = Multimodal EchoBorg as proponent, HHH = human only pre-study corpus).

Table 6.2: Statistics of pairwise comparisons. The top half showing the comparisons of scores attributed to the proponent debater embodiment (VH or MEB) between subject. The bottom half showing the comparisons of scores attributed to the proponent (VH) and the opponent (MEB) within the HEA condition.

	scale	subscale	contrast	estimate	SE	df	t-value	p-value
between	godspeed	animacy	VH - MEB	-0.722	0.299	34	-2.415	0.02
between	godspeed	anthropomorphism	VH - MEB	-1.233	0.275	34	-4.479	$<.001$
between	coupling	copresence	VH - MEB	-0.639	0.155	34	-4.113	$<.001$
within	godspeed	anthropomorphism	VH - MEB	-1.022	0.293	17	-3.491	0.003
within	godspeed	intelligence	VH - MEB	-0.537	0.258	17	-2.079	0.05
within	coupling	copresence	VH - MEB	-0.667	0.133	17	-5.030	$<.001$

p<0.001, η^2p=0.371). Post-hoc tests show that the proponent was rated higher on animacy and anthropomorphism, when embodied by the MEB. On the co-presence sub-scale of the coupling questionnaire, we found a significant effect of the "proponent embodiment" between configurations (F(1,34)= 16.920, p<0.001, η^2p=0.332). Post-hoc tests revealed that the proponent was rated higher on co-presence, when embodied by the MEB. Since participants within the Human-EchoBorg-Agent (HEA) condition (n = 18) interacted with both an MEB and a VH embodiment, we conducted an ANOVA on the effects of the within subject variable "debater embodiment" on those sub-scales that measure attributes of the individual debaters. Again, two sub-scales showed (near) significant effects: anthropomorphism (F(1,17)= 12.190, p= 0.003, η^2p=0.418) and perceived intelligence (F(1,17)= 4.322, p= 0.053, η^2p=0.203). There were no other significant effects of "proponent"and "debater embodiment" found on any other sub-scales. Statistics for the between and within post-hoc tests are reported in Table 6.2, and response distributions are visualized in Figures 6.2, 6.3 and 6.4.

6.6 Discussion

Reiterating, we compared participants' perception of a traditional VH embodiment with a MEB embodiment, while both had the same conversational agent ('mind') determining the utterances and behaviour they display during a debate. We examined the participants' perception of the agent in terms of concepts that are often used when evaluating artificial agents, and participants' perception of the overall experience of the interaction.

6.6.1 Comparing the Multimodal EchoBorg and Virtual Human embodiments

Looking at the hypotheses, we observe the following. We only partially support our first hypothesis, that the MEB is perceived as more favorably than the VH on perceived agent traits: while the MEB does score higher on the Godspeed instrument sub-scales that measure the anthropomorphism (both in the within and between comparison) and animacy (between subjects), there is only near-significance for the intelligence in the within comparison, and no difference in likability ratings. These results suggest that interaction is more than just appearance. Our interpretation here is that only measures that relate to the outer appearance of the embodiment seem to be favored by the human embodiment, while it fails to lead participants into (falsely) overestimating traits that are related more to the behavior of the conversation partner, i.e. the intelligence and likability.

Our second hypothesis, that the quality of the interaction with an MEB will be rated more positively than with a VH, is rejected. We had speculated that whenever there is another human involved, even though it displays the same limited behaviors and interactivity as displayed on the virtual embodiment, the interaction would be perceived as more engaging and interesting. This appears not to be the case, as interactions featuring the MEB were not rated more positive than those only featuring VH embodiments. Together with the observation in regard to the first hypothesis, this may lead us to assume that any initial expectations favoring a human embodiment are overruled by the limited perceived mind during the interaction.

Finally, our third hypothesis concerns how participants perceived their bilateral relationship with the ECA. We hypothesized that the MEB would be rated more favorably because the human appearance evokes the expectation of a human level of interactivity. On the basis of our results, we reject this hypothesis. Looking in more detail at the sub-scales, coupling with the debater, engagement and believability did not score significantly higher for the MEB. Only the co-presence sub-scale the MEB was rated significantly higher. This is a measure that might be more influenced by the physicality of the embodiment rather than by the displayed behaviour. Thus, a human embodiment might not create a better relationship between a user and ECA, but might evoke higher feelings of co-presence.

In an attempt to find alternative explanations, we may consider works such as that of Nowak and Biocca, who found that more anthropomorphic embodiments of agents (or in their case avatars) might “set up higher expectations that lead to reduced [co-]presence when these expectations were not met” [188, p481]. Initially, a MEB will set up the highest expectations, while the limited capabilities of our conversational agent very likely meant that the MEB was not be able to meet those expectations during the interaction.

It is also important to consider one other limitation of this study, namely the sample size. Its small dimension could under power the statistical significance of the results. We need to replicate the experiment with a larger population. However, the present study still shows a possible methodology and how sensitive the conventional measures are in a setup like this. The reported results are not the main contributions, but they provide an overview on the effects that the MEB can have on a human interactant in this preliminary version.

6.6.2 Comparisons with a Human-only experience

Next we explore the scores of the different ECA embodiments with the scores collected in the all-human corpus recording sessions. We find that the scale ratings attributed to human proponents, for the GQS, are quite similar to those attributed to the MEB proponents on all sub-scales (see Figure 6.2). Although we expected this for the perceived animacy and anthropomorphism subscales, we also expected humans to receive much higher ratings on intelligence and likeability, based on the coupling concept [22]. The experience in the pre-study corpus, in fact, was more open and interactive than the experiment sessions. Due to the fact that all the interactants were participants, there was no virtual agent limiting the conversation or creating bias. Instead, we find that the levels are similar to both the VH and the MEB ratings. For intelligence, an explanation may be a ceiling effect, with medians and upper quartiles concentrating around the 4-5 point level of the sub-scale. For the GEQ, comparing the responses of moderators from the human-only pre-study corpus to the responses in the experiment sessions, we see a different trend from the debater perception rating discussed before (see Figure 6.4(a)). The experience from the human-only session scores seem much higher in terms of perceived challenge, competence, positive affect and tension when compared to the experiment sessions. Similarly on the SAM-instrument, the ratings on arousal and valence seem somewhat higher (on dominance we have a high variance in the responses, but the median level is also higher). Thus, perhaps the increased interactivity of the human debaters informed these measures - which would further support that the limiting factor for the MEB scores is based on the limitations of the ECA system controlling the MEB, which the human embodiment could not hide. Alternatively, human corpus recording sessions had different rules and featured a less structured debate, which may also have affected the game experience scores. During those debate sessions, social dynamics and unexpected stimuli were more common. On the basis of the literature, this probably contributed to increase the level of attention, arousal, and engagement.

6.6.3 An evaluation and inner perspective of the Multimodal EchoBorg

A contribution of this work is the first implementation of a Multimodal EchoBorg apparatus for ECA systems. To understand the limitations and how to improve it in the future, we asked the participants, at the end of the experiment, to provide feedback on the MEB interlocutor before we revealed that the actress who acted as the MEB was a confederate. All participants reported that after more or less five minutes of conversation, they perceived the interlocutor as awkward. They provided different explanations for this behavior: some participants thought that the interlocutor was shy, others thought that the interlocutor had some mild form of mental disorders, only two participants understood that she was a confederate and she was acting. We also asked the actress that acted as the MEB to fill out a self-report after each session. She reported deviations from the behaviour that the ECA asked her to perform. Specifically, she reported how, when and why she deviated and in which modality she deviated (listening behaviour, speech, gestures). We compared her reported deviations with recordings of her actual behaviour to check if her perception was consistent with the real experience. The actress never reported deviations in listening behavior. She reported most deviations in speech, citing as reason:

i) *"I thought that was the sentence I had to say but instead I said it faster."*; ii) *"I tried not to look at the screen because I felt that the participant might notice something was happening behind him."*; iii) *"The participant wanted to speak and I had to speak over him."*. Concerning the gestures, the actress reported that it was not always easy to shadow the gestures from the interface, for example: *"I had the impulse to follow my own reaction to what I was saying"*. From the recordings, we observed that the majority of deviations happened during the gesture shadowing, while we observed only very small variations in the speech shadowing and no variations in the listening behaviors. Thus, the actresses self-reported deviations and the observed deviations were not in line, suggesting that the actress was perhaps less aware of her performance on gesture shadowing. Perhaps integrating an automatic feedback mechanism for shadowing behavior in a future MEB setup could improve the quality of shadowing.

6.7 Conclusion and Future Works

We explored how the embodiment of an ECA influences the perception of the interaction using an upgraded version of the EchoBorg method, the Multimodal EchoBorg. We present our first experiences employing the EB method in ECA research. From a practical standpoint, we have built an apparatus for multimodal shadowing, and gained insights into how it can be employed with a confederate in an experimental setting. From the user study, we have obtained a first overview of the biases that may occur when replacing the embodiment of a VH with a real human, keeping all other aspects of the ECA behavior the same. In summary, the results from our study do not support our initial assumption that an experience with an MEB would always be rated favorably over the same interaction with a VH based on the belief that (one of) the actor(s) was a human. Instead, we see that the limited artificial mind may shine through more than expected, limiting such favorable ratings.

We acknowledge a number of limitations of the present work. First and foremost, the sample size was relatively small for ANOVA with post-hoc tests. We reported significant results; however, the study also has a possibly inflated test power due to the procedure used. Furthermore, the study design lacks a counterbalance between debater role and gender, and the analysis of comparisons between and between subjects in this way may have inflated statistical power. Future studies are necessary and may benefit from a different study design. For example, a dyadic interaction scenario with a strict between subject design is more suitable for a more rigorous investigation of the MEB when studying perception biases. Furthermore, metrics for the shadowing performance of the MEB need to be defined and measured for control purposes. The next important step to understanding if and how we can benefit from the MEB method for ECA development is to look more at how the user treats the MEB, perhaps with a similar methodology to that used in [48]. Additionally, there are possibilities to improve the MEB interface further, allowing for more accurate shadowing in even more modalities using, for example, visual overlays in covert AR glasses, or perhaps haptic displays that provide information for motion in different bodyparts.

In fact, we recognize that in our study, the MEB was potentially overreliant on a priori knowledge of the dialogue. She was able to practice her performance, as in large parts the speech and the accompanying gestures were fully deterministic. For a future *production* MEB system, it should also be possible to realize dynamic, spontaneous behaviors. Additionally, not all MEB behaviors could be controlled (e.g., non-verbal leakage). There may even be systematic biases that are not controlled for,

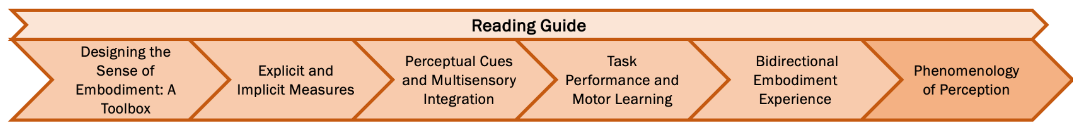
for example in the MEB's gaze behavior, due to the use of the MEB interface.

Reflecting on Rostand's play *Cyrano de Bergerac*, the moral of the story was that Roxane was attracted to Christian's body but ultimately fell in love with Cyrano's mind: a feat not likely repeated by our MEB, as our ECA *indeed* turned out to be not as smart as it looked.

7

A Multidisciplinary Intersection: the Sense of Embodiment from Philosophy to Technology

> If you don't know where you are going any road can take you there.
>
> *Lewis Carroll, Alice in Wonderland*



7

Abstract

The aim of this chapter is to report a different disciplinary perspective on the SoE concepts. Up until now, we have reported the point of view of empirical sciences such as computer science, robotics, neuroscience, cognitive science, psychology, and design. Following, we present the philosophical view on this broad topic. We especially focus on the work of the philosopher Merleau-Ponty: *Phenomenology of Perception*.

7.1 The Sense of Embodiment at 360 degrees

A fascinating aspect of SoE and its relation to teleoperation is the multidisciplinarity that its investigation involves. The concept of SoE is faced and studied in different research contexts, from philosophy [61, 165, 209], neuroscience [26, 33, 89], computer science [2, 132, 160], robotics [159, 246, 269], and is usually mediated by cognitive science [74, 75, 199]. All these fields investigate SoE from different, possibly complementary perspectives. From a philosophical point of view, the concept of embodiment was observed and acknowledged before any attempt at scientific investigation in the human brain. If we look at the Hindu scripture Bhagavad Gita [32] (the dating of this text is uncertain and spans a period ranging from the V century to the II century BCE), an embodied philosophy was already conceptualized and presented, i.e. the idea that the essence of an individual can be transferred, because the body is mutual and just a container. Referring to more recent works, the philosopher Maurice Merleau-Ponty in his Phenomenology of Perception [174] approaches the dimension of experience referring to the relation between the lived body and the phenomenal intersensory world. He argues that the body is not just a mere biological or physical unit, but it structures one's situation and experience within the world, supporting the idea of a dynamic body schema. Nowadays, we still lack evidence to fully understand the SoE phenomenon. We are still missing a proper definition of the attributes of embodiment theory. However, scientific disciplines such as neuroscience and cognitive science are currently building the theoretical foundations of SoE.

Although neuroscience provides empirical evidence for this phenomenon, philosophy allows its description. This would lead us to think that there is a perfect integration of these two fields: philosophy is used to describe what we observed from neuroscientific evidence. However, philosophical theories of embodiment were defined before experimental neuroscientific evidence. The concept of embodiment was observed and acknowledged before it was possible to study it in the human brain. Today, we still lack evidence to fully understand the phenomena of SoE and consciousness. To make a well-known comparison, the same happened with the general relativity theory of Albert Einstein. Initially, it seemed to offer little potential for experimental testing, as most of its assertions were on an astronomical scale. Its mathematics seemed fully understandable to a small number of people. It took almost 50 years, around 1960, for physics and astronomy to become central to physics [73]. New mathematical techniques made its concepts more easily visualized. As astronomical phenomena were discovered, such as quasars and the first black holes, the theory explained their attributes, and their measurements confirmed the theory.

The main issue related to supporting such intuition with philosophical theories, without neuroscientific experimental tests and evidence, is the speculation that arises from it. When it comes to defining, understanding, and explaining a certain phenomenon, if it is not possible to observe and comprehend it with transparency, then all researchers begin to create their own terminology to describe it and their own methodology to measure it. This pushes away the possibility of a standardization and a common integration of all future evidence on the topic. Moreover, this can bring about the generation of fictional words that do not describe the phenomenon itself, but just define a concept, or an abstraction, generated by the human mind. The historian and philosopher Yuval Noha Harari perfectly explains the concept of fictional realities [111]. He describes the difference between the imaginative capacities of humans and non-human animals. While animals and humans experience 'objective reality' (e.g., trees) and 'subjective reality' (e.g., feeling pain), only humans experience 'intersubjective realities' (e.g., human rights, corporations, money, nation). Harari also refers to intersubjective realities as fictional realities. The term 'entities' is used to describe what is generated within the intersubjective sphere.

In [59], we found a very interesting parallelism between the *phenomenology* of Merleau-Ponty and the concepts of body schema and image, and how they affect motor learning. In this chapter, we try to project the philosophical view of the philosopher on the embodied cognition theory on each sense of the embodiment component (ownership, agency, and self-location). In particular, the goal will be to define the self, its consciousness, and its relationship with the environment. These aspects will be defined as interrelated and impossible to disentangle. In this way, each aspect is represented as an important contributor to the development of SoE in an individual.

7.2 The Philosophical point of view

In the XX century, the philosopher Merleau-Ponty, on the basis of Edmund Husserl philosophy [148], described phenomenology as the study of the essence, such as the essence of perception and consciousness. However, *phenomenology* is a philosophy that collocates essences in existence, and states that human beings and the world can just be understood on the basis of their factuality. He underlines the importance of the structures of the factual world, beyond the pure logic structures. We move from a consciousness of intents to an existential dimension engaged in the real world. In the last stages of his career, Husserls describes phenomenology as a science of the *universal how of being* that already exists in the world. This science aims at investigating the transcendental correlation subject-object, and its goal is to understand the typical operations of the constituent subject. This is called *transcendental epoché*, and it is a behavior that aims to find intentional modalities relative to the subject-world of correlation. The epoché tries to reflect the entire concreteness of life, and its intermediate and final determinations, which do not reveal something human, neither something about the soul, neither about the psychic life, nor about the real psychophysics of human beings. Everything is part of the phenomenon of the world as a constituted pole. The epoché makes explicit the constitutive activity of the self, in relation to the phenomena of the world. Martin Heidegger presents the problematic of the constitution, namely *the being in the world*, the stoicity of the living spirit, the actual existence, and the impossibility of a pure transcendental subject. Then, we arrive at Merleau-Ponty, who criticizes the transcendental aspects. He insists especially on the reduction to the world of life. Phenomenology is a philosophy that considers that the world exists before thought. The aim is to find a relation with the world to define and attribute a philosophical statute. This philosophy aims at being not only a reliable and exact science, but also a summary of space, time, and the world in the way in which they are lived. Merleau-Ponty presents a phenomenological reduction, thanks to which we can have access to the concrete being of the world, in a pre-theoretical dimension. He overpasses the categories of subject and object by recognizing their mutual implication.

7.2.0.1 The Sense of Ownership

The philosopher describes all the existence of the body as intentional, not just the consciousness of the individual. Willfulness is no anymore a peculiarity of consciousness, but it concerns all dimensions of existence. The existence is defined as the embodied consciousness, through the analysis of the perception of our own body. Merleau-Ponty extends the willfulness of consciousness to the entire body, integrating the moment of existence. The inherence to a body located in the world is necessary for the transcendental subject. The philosopher states that the subject cannot perceive the world without a body, therefore it needs to be embodied; however, at the same time, the objects and the world have some intrinsic properties and their own aseity. For what concerns the multisensory integration, in the primitive perception, it is impossible to distinguish among senses, and the world presents itself as impossible to disentangle. Below the perceptional level, there is another perceptual substrate that is characterized by its own aseity. Merlau-Ponty describes an active character of perception, and he recognizes how the body changes through the willfulness. The subject embodies the object, by giving a meaning to it. The relationship that is established between the perceiver and the perceiver is only intentional, and they are not two independent entities. There is a chiasmatic and dynamic relationship between subject and object, they do not coincide, but they are strictly related.

Mearleau-Ponty also revised the concept of subjectivity that he describes as connected to the body, such as an active power of meaning, willfulness activity always involved in the world. Physiological and psychic process cannot be dichotomously distinguished. The embodied self, who lives in first person, goes beyond the distinction between the inside-self and for-the-self, while the embodied object, which is the topic of the classic physiology, is a representation of the body seen from the outside. For Merleau-Ponty, the phenomenal body is never completely defined as an object, because it is the perspective from which every other point of view is derived. The subject of perception is time and history, because it is always willfulness, a tendency towards a new direction, openness to new possibilities. The philosopher states:

> The world is all inside, and I am all outside myself.

Merleau-Ponty does not see consciousness as something that has inside its own self, such as "consciousness of". Instead, the consciousness represents the "being attached" of the subject to the world through which just in a second moment the consciousness will have the possibility to join it. The philosopher puts his attention to the embodied subject that is located in a world that constitutes the personal substrate of each singular experience. The pragmatic subject is conditioned by the perception of the historical temporality, but it can transform it, because it can perpetuate the temporal development and enriches itself with new actions and purposes. Perceptual syntheses, being temporal, represent the possibility of the subject to transform the world.

7.2.0.2 The Sense of Agency

In the vision of Merleau-Ponty, the subject is not a transcendental consciousness such as "I think", but it is an embodied subject related to a world that becomes "I can", i.e.

a subject that has agency, immersed in the flow of time and that exists just in relation to the world. The phenomenology is an attempt to describe and reveal the ambiguity of the inter-world in which consciousness and existence, reflection and perception, the body and the object properties fill their distance and show their mutual relation. The philosopher describes an existential dimension, that is concrete and pragmatic, in which between "I can" (a subject that is embodied in a body and related to an environment) and the world there is not a relationship of founder-founded, but a dialectic relationship, or mutual foundation. By dialectic, it is meant the communication and interconnection of two poles - in this case, subject-object. Starting from Gestalt psychology, the philosopher formulates that the perceived objects are not the mere result of the combination of atomic sensory data, but they are made of lived experience and emotional tonality. The organization of the sensory field derives from the oriented subject activity and from the dynamic of the relationship between figure-background, which is considered the minimal unit of perception. Each human being is a whole organic entity in which the spirit is embodied in a body and the body has a spiritual meaning. The subject does not coincide with a pure subjective epistematics, but it becomes 'behavior', and a mutual integration between the self and the environment. The consciousness wears thin and reveals its relationship with the perceivable. Perception is not just a receptive moment, it is an active process of object constitution. Therefore, the world is not simply passively experienced by the subject who perceives, but it is actively structured through the synthesis of each singular property of the objects. Merleau-Ponty sees the epoché as a way to get access to the phenomenal field and to the concept of intentionality. He reveals a subject that discovers the sense of belonging to a body and, through it, the subject discovers the sense of belonging to a world. The body is not just the association of different parts that are mechanically put together, and it is neither an instrument of consciousness. The body is everything, structural and organic, it is a complex system, like a bow of the inner reality and the outward appearance. The physiological and the psychic are both integrated in the intentional movement of *being-in-the-world*.

7.2.0.3 The Sense of Self-Location

In the vision of Merleau-Ponty, temporality creates on one hand the possibility to define the self and constitute ourselves through the explication of the meaning that belongs to us since we are subjects located in the world and, from the other side, to go beyond the self by opening to something that is other than ourselves. The subject is present to oneself only through the spreading out in the multitude. The body and the object are linked, the connection is dynamic and can be compared only to the connection that exists among the parts of our own body. There is an implication relationship between the subject who perceives and the object that is perceived. We cannot say "I perceive", but "it is perceived in me". The world is an open unit that is familiar and can be modified. The temporal subject is connected to the world dimension, which has its own meaning, and which plays an active role in the perception. Human beings reveal the meaning of the world that the world itself presents to them.

The meaning of the world cannot be meant as an absolute spirit; it needs to be meant as a meaning that it is generated by the interaction of the experiences of the subject with other subjects and their relationship with the world. The self and the other coexist, and they are part of an inter-embodiment experience. The embodied subject is characterized by its openness to the world. In this view, the existence of an other outside the self is guaranteed by the structure of being-in-the-world.

7.3 Integrating the Tech perspective

We provided a different perspective on embodiment, being aware that this vision represents one of the possible formal definitions and explanations of what is happening when we try to embody an operator in an avatar. Merleau-Ponty argued that perception is not just a mental process, but is also closely tied to the body and the environment in which we exist. He believed that the body and the world are intertwined and that they cannot be understood separately. Merleau-Ponty's view on embodiment can be extended to the relationship between humans and technology. Starting from his vision, we can argue that technology is not something external to the body but is instead an extension of the body. Technology is an embodiment of human intentionality and creativity. As such, technology is not a tool for humans to use but rather is an integral part of human existence and identity. Technology is not neutral, but rather a reflection of the values, beliefs, and goals of the society in which it is created. Thus, the relationship between humans and technology is not one of mastery or domination but rather one of mutual shaping and influence. Humans shape technology, but technology also shapes humans and their perception of the world. This perspective on embodiment and technology has important implications for how we think about the role of technology in society. Instead of seeing technology as a means to an end, we could see it as a reflection of our values and beliefs. Our relationship with technology is not one of mastery, but is rather one of mutual influence and shaping. Overall, this view on embodiment and technology provides a unique perspective on the relationship between humans and technology. It highlights the importance of considering the embodied nature of human experience and the ways in which technology is integrated into our lives and shapes our perception of the world.

8 Discussion

8.1 Investigating Sense of Embodiment: What are we looking for?

The attempt of this thesis is to present first an overview, and then new insights on the concept of the Sense of Embodiment in relation to teleoperation. We present a collection of several studies designed to answer the main research questions listed in the introduction and to try to provide new insights.

To answer RQ1 (How can we define a standard framework or guidelines to design a teleoperation setup aimed at optimizing the embodiment experience of the operator?), in Chapter 2 we presented a toolbox to assess SoE, consisting of five steps: 1) application scenario; 2a) embodiment components involved, 2b) perceptual cues to optimize; 3a) surrogate, 3b) control device; 4) tasks; 5) assessment measures. The literature review presented led us to the conclusion that we lack a standard and well-supported definition of SoE. We integrated the previous well-known definitions into a unique one, highlighting, with respect to the previous definitions, the distinction between subjective and objective aspects of SoE, and underlining the importance of the role of the three components describing SoE, namely, the sense of ownership, agency, and self-location. The relation between the SoE components and how to disentangle them is still an open debate. Most of the literature reports a strong correlation between sense of ownership and agency ([93, 205, 252]). The two senses have common spatio-temporal constraints of the integration processes, namely they activate common brain areas and neural patterns, and they affect each other at a level that makes it very difficult to disentangle their effects on the embodiment experience. However, there is also evidence that these two components involve different cognitive processes ([124, 240, 251]). These components appear to be phenomenally uniform and strongly correlated, but they are complex crossmodal phenomena at heterogeneous functional and representational levels. This is reflected in both the assessment and the testing phase of the SoE. In evaluating, a combination of explicit and implicit measures provides the most complete picture of the phenomenon. Explicit measures assess the conscious embodiment experience of the individuals, while implicit measures measure intrinsic and unconscious changes in the SoE levels. In testing, tasks are often designed to test the three SoE components at the same time, without having clearly in mind which embodiment components one wants to address before designing or assessing an embodiment experience.

In Chapter 2, we introduce the concept of the *mediator* and its importance in the relationship between SoE and teleoperation. To answer RQ2 (To what extent does perception of the mediator affect the SoE experience and task performance? Does this concept play an important role in the teleoperation applications?), in the user study presented in Chapter 5 we administered a preliminary survey aimed at addressing the key aspects characterizing this concept. We are currently working on a final version of it after having observed its relevance. The results show a significant negative correlation between the mediator perception and the embodiment components, suggesting that when the perception of the set-up as mediator is closer to null, the level of transparency perceived by the operator while manipulating the avatar is higher, and this enhances the SoE experience. This means that the highest level of SoE is achieved when the operators feel so embodied with the avatar, and teleported to the remote environment, that they do not perceive the setup that they are using to teleoperate the system anymore, and they feel completely transported and immersed in the embodiment experience.

RQ3 (What different pieces of information about the embodiment are revealed, respectively, by the explicit and implicit measures?) and RQ4 (Can the current embodiment measures disentangle the embodiment components?) are mainly addressed in Chapter 3. With the presented studies, we confirm that while explicit approaches to measuring the SoE try to address its specific components, implicit measures may not exclusively reflect a specific SoE component. However, explicit approaches are subjective measures and depend on different factors of the user experience, certain biases, and language (e.g., see [123]). This means that they are mostly strongly affected by individual differences, making a comparison more challenging. This is why it is recommended to use a combination of explicit and implicit measures to assess SoE. We conclude that only explicit measures can attempt to disentangle the embodiment components, since implicit measures are too noisy and it is not possible to link the observed results to a specific component, since it is also challenging to design an experiment aimed at disentangling the components. However, it is important to keep in mind the limitations of the explicit measures and their strong variance.

An important aspect that we introduced in the previous chapters is the role of the perceptual cues and how they affect the SoE and task performance. The user study presented in Chapter 4 addresses RQ5 (to what extent do perceptual cues affect SoE components and task performance? Can we rank them in order of importance? And is the order consistent over the different SoE components and task performance?), RQ6 (Are there interaction effects between the perceptual cues? Is a simple additive model sufficient or do we need a more complex model to test the effect of the perceptual cues together?), and partially RQ7 (How are SoE and task performance related?), which we also address in Chapter 5.
In our study, we manipulated five relevant perceptual cues: 1) the field of view, 2) the human-likeness of the virtual arm, 3) the visuo-proprioceptive synchronicity, 4) the tactile feedback, and 5) the connectedeness of hand and arm, resulting in rank orders

of importance for the sense of ownership, sense of agency. The effect of perceptual cues on the dependent variables varied. We found a significant effect of all perceptual cues on the sense of ownership and agency, with the only exception of the human-likeness on the sense of agency (as expected). Visuo-proprioceptive synchronicity, as suggested by previous studies ([62, 136]), was always the most important cue, the rank of the other cues differed for the different embodiment components.

To answer RQ6, we did not observe any significant interaction effect between the independent variables, supporting the hypothesis that a linear additive model can adequately describe the combined results of the five perceptual cues.

Finally, we found that the SoE and task performance are related, although the correlation is weak. Task performance was only affected by two cues: the visuo-proprioceptive synchronicity and the connectedness, where performance was better in the suppressive rather than the supportive condition. When the visuo-proprioceptive synchronicity was in the suppressive condition, participants consistently reported difficulties in accomplishing the task. We would have expected connectedness to increase proprioceptive information, therewith supporting task performance. The opposite finding may have been caused by a less cluttered display when only the disconnected hand is presented rather than the whole arm. Field of view, tactile feedback, and human likeness did not have a significant effect on task performance. Previous studies reported that a higher SoE resulted in a much larger increase in task performance ([167, 217, 220]). An explanation could be that a high SoE reduces the cognitive workload of the operator and has a larger effect when the task demands are high, for instance, when operating an avatar or teleoperation system that is more complexly designed and that can achieve more complicated tasks, as is the case of the study presented in Chapter 4. On this basis, we hypothesized that a higher SoE could facilitate and speed up motor learning. We tested this prediction through the user studies presented in Chapter 5.

In Chapter 5 we deepen RQ7, and we also addressed RQ8 (What is the effect of SoE on task performance in a perceptual-motor task?) and RQ9 (What is the effect of SoE on the asymptote of learning in a perceptual motor task?). We hypothesized a positive effect of SoE on task performance (faster learning when SoE is higher) and on the overall embodiment experience. The results from the presented user studies show a faster motor learning, a better task performance and, on average, a decreasing perception of the surrogate as a mediator and a positive increase in perception of the SoE components in the conditions supporting the SoE compared to the suppressive ones. However, the results among the user studies are not so straightforward to interpret. There are some observations that raise more questions. For example, in all user studies, when we found a correlation between SoE components and task performance, it was rather weak. In the event of a dependency between SoE and task performance, we expected to find a negative correlation in the suppressive embodiment condition. As already discussed, in the suppressive condition, even if the learning effect is not significant over trials or among conditions manipulating the perceptual cues, we can observe a trend that suggests an increasing performance over time. These observations suggest the hypothesis that operators can achieve the same task performance in both a supportive and suppressive embodiment, but in the latter case, they would need more time to adapt to the system. However, this hypothesis would need further investigation in future works. A further explanation of these observations and results is that during the teleoperation of a robotic system, the internal model would begin to simulate the dynamical properties of the robot (or any other avatar or device), which could be incorporated (or replaced) with the current model of the body schema of the operator. [220] and [167] report that increasing SoE improves task performance. However, these studies are conducted in the prosthetic field and with participants who experienced upper limb loss. [217] reports similar results in a study conducted in VR. An explanation for our findings could be that the operator's perception of the surrogate as a mediator was higher compared to the studies in the prosthetic field reported in the literature. For what concerns studies in VR, the virtual avatar appears more similar and responsive to the operators' commands than a robotic arm, that has physical limitations. Moreover, the VR environment appears to be more immersive and it is usually perceived as an alternative reality. This means that operators experiencing a VR environment do not deal with processing images of their reality but differently rendered and compressed; they just get used to a new virtual environment. However, the appearance and level of control on the virtual workspace and the virtual avatar can affect the operators' perception of the transparency between them and the surrogate, moreover we observed that the task also plays a role in the avatar's perception. A definition of this observation is given in [76], in which the authors state SoE

> has both a subjective and an objective component. The subjective component is related to the individual update of the level of embodiment, that everyone performs intrinsically and unconsciously, mostly on the basis of perceptual cues. The objective one concerns the parameters, of the system or the experience (such as the setup, the context, and the tasks), that can be manipulated in order to increase or decrease the SoE. However, we are still unaware of the nature of all the parameters that can affect the SoE and their specific effects.

Finally, we focused on the diatomic interaction between the operator and the recipient, moving our attention to the recipient's experience. In Chapter 6 we attempt to answer RQ10 (What enhances the telepresence perception of a social entity?) and RQ11 (To what extent does the way in which the body and the mind of the operator are transferred to the remote avatar play a role in the recipient's experience and interaction?). Our conclusion here is that mind perception plays the most important role in affecting the interaction. This means that regardless of avatar appearance, if most of the characteristics traits of the operator are transferred to and rendered in the remote environment (such as voice or face), the likability of the interaction partner will not be affected even if the avatar appearance level is low. Therefore, a human embodiment might not create a better relationship between a recipient and an avatar representing an operator, but it might evoke higher feelings of copresence and engagement in the interaction experience.

Compared to the initial predicted model (see Figure 1.2) presented in the Introduction, we now report an updated and enriched version of it (see Figure 8.1), based on our investigations. The brain has to deal with the multisensory integration of the stimuli provided by the remote environment. On the basis of the perceptual cues experienced by the operator, some components will be enhanced more than others. Particularly, we investigated the effects of visual human-likeness, field of view, connectedness, visuo-proprioceptive synchronicity, and tactile feedback. The way in which these components are affected is dependent on the mediator, namely the setup realized to operate the system which can increase or decrease the transparency in perceiving the avatar. Individual differences can also play a role in the final perception of the SoE, as demonstrated in Chapter 3, in which we observed that the population with high kinesthetic intelligence is more resilient to the embodiment illusion, if we look at the pupil dilation data. Eventually, these factors define the final feeling of embodiment for the operator. We also observed, as reported in Chapter 5, that improving the SoE does not directly affect task performance, but it has an effect on motor learning and adaptation. We state that designing a system that supports the SoE can reduce the training time needed by an operator to learn how to use the system, it can bring to a significant rapid improvement of task performance, and facilitate the re-calibration of the body schema.

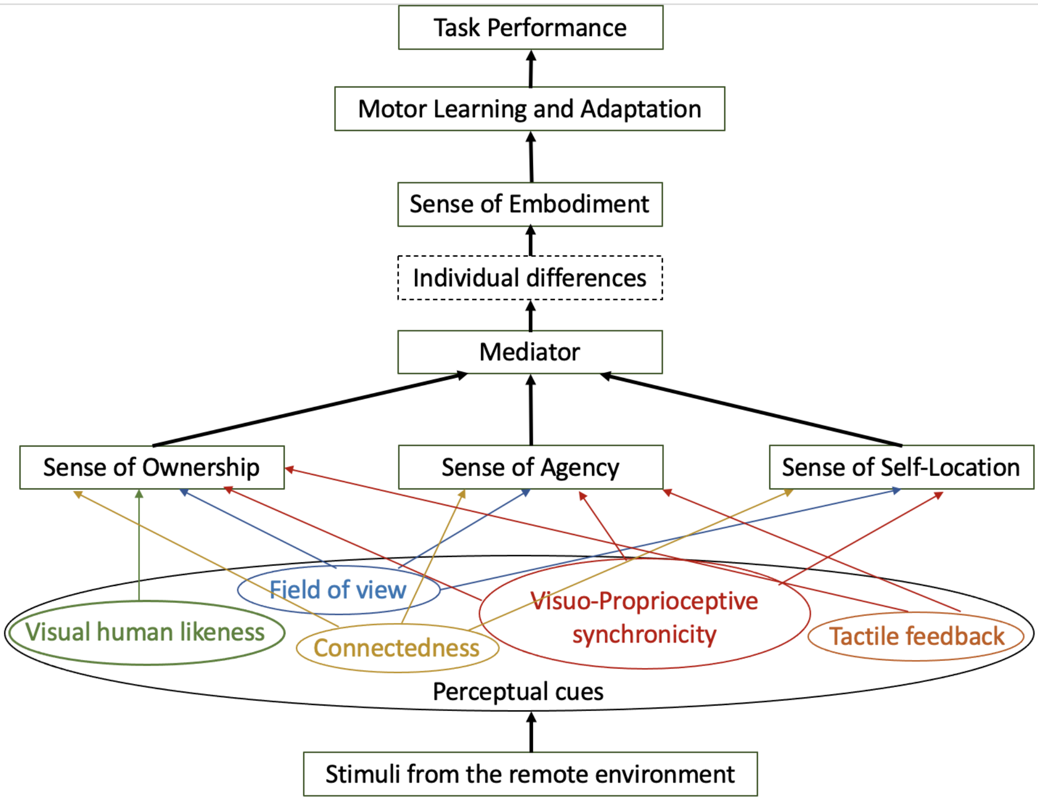


Figure 8.1: The figure represents an architecture of the predicted final model of the Sense of Embodiment in teleoperation, after our investigations.

The descriptive model that we present needs to be ground proofed. We suggested that collecting behavioral data from motor learning and active sensing could result in the development of a predictive model of embodiment. This will be the next step in the work presented in this manuscript.

In the Introduction, we presented two well-established cognitive models as possible explanations of the complex mechanism that describes the relationship between the

SoE and teleoperation: the *Bayesian perceptual learning* and the *predictive encoding theory*. While both theories involve updating internal representations based on sensory information, Bayesian perceptual learning is deeply rooted in updating internal beliefs or priors based on sensory evidence in a probabilistic manner, while predictive encoding theory emphasizes the predictive nature of the brain and the minimization of prediction errors. Although we do not have enough evidence to draw a conclusion, our inclination leans toward supporting predictive encoding theory. It emphasizes the adjustment of internal predictions to minimize prediction errors, involving a constant cycle of prediction, error evaluation, and learning. When there is a mismatch between predictions and sensory input (prediction error), the brain updates its internal models to reduce future prediction errors. This aligns more closely with our empirical findings and the insights derived from the literature we reviewed.

9

Conclusion and Future Directions

> If you don't know where you are going any road can take you there.
>
> *Lewis Carroll, Alice in Wonderland*

We aimed at providing complete pictures, according to the literature and the new insights, on the concept of the Sense of Embodiment. In particular, we related it to teleoperation. To address the presented research questions, we spaced from cognitive science to robotics. We demonstrated the multidisciplinarity of this topic and how it can be of interest for different research fields. Mostly, we showed its complexity and, sometimes, ambiguity which make this concept of embodiment fascinating while also lacking well-grounded theory to be explained.

As a first step towards a standardized framework, we presented a toolbox and guidelines to design, test, and assess the SoE in teleoperation scenarios. The aim of standardization was driven by the need to help the research community compare the different studies and, ideally, create a predictive model of the level of SoE in a task-oriented system before its implementation. This predictive model, part of our future directions, would become a starting point for the improvement and optimization of new teleoperation systems, VR setups and, more generally, embodiment experiences.

Looking closer at the toolbox steps, we focused on the SoE measures. We distinguished explicit and implicit measures, not only in their definition, but also in the SoE aspects that they addressed. We suggested a combination of the two for an optimal assessment of the embodiment components. Moreover, we recognize that further investigations are necessary to arrive at specific and most reliable measures, since the current ones are either too subject to individual differences, noise, or they were not developed to specifically assess the SoE.

To test the SoE and connect it to teleoperation, we then address the role of perceptual cues and how they affect the manipulation of its components and task performance. Our full factorial experiment resulted in a rank order of five different perceptual cues with respect to their effect on the sense of ownership, the sense of agency, and the performance of tasks. Different rankings were found, but visuo-proprioceptive synchronicity affected the three outcome measures the most strongly. We did not observe an interaction effect between the perceptual cues and we found a weak relation between the level of SoE and task performance. These findings helped to decide on the factors to optimize in a system in order to achieve a high sense of ownership, agency, and task performance. Furthermore, that was a necessary baseline to manipulate SoE in the following studies. This was also the starting point of our investigation of the effect of manipulation of the SoE level on task performance in a teleoperation scenario.

Our results indicate that SoE does not have a direct effect on task performance, but it seems to affect motor learning and adaptation, and to reduce the perception of the surrogate as a mediator. A supportive embodiment could facilitate the operator's learning experience of a setup while also increasing its use over time. The quality of movement execution can be improved through practice. Human operators are required to recalibrate their body schema, namely the sensorimotor representations of the body that guide actions, considering the setup that mediates between their own body and the avatar in the remote environment. This implies that human operators need to build a new control to perform actions that they already know how to perform with their own bodies.

Finally, we explored how the mind and body perception of the avatar can affect the recipient's experience of the operator in a teleoperation scenario. To do that, we designed an upgraded version of the EchoBorg method, the Multimodal EchoBorg. In summary, the results of our study do not support our initial assumption that an experience with an MEB would always be rated favorably over the same interaction with a VH based on the belief that the actor was a human. Instead, we see that the limited artificial mind may have a higher impact than expected, limiting such favorable ratings. Therefore, the mind perception plays a fundamental role in the interaction, and the body appearance and rendering of the operator can only affect the co-presence level and engagement, without affecting the perception of the operator.

To conclude, in terms of aiming for beyond state-of-the-art, we investigated the postulates on enhanced embodiment as stated in [246], where they have postulated that the embodiment of the robotic system leads to optimal perceptual transparency and increases task performance. This investigation was a pioneering study in the field of telerobotics, more importantly, these studies investigated the relationship between the SoE and teleoperation in different selected use-cases: industrial robotics, social robotics, and virtual reality simulations. We collected behavioral data that will be used to model active sensing and motor learning, which play an important role in the embodiment experience in teleoperation. This will allow us to predict the embodiment level and to design the most efficient cockpit or setup to achieve the highest level of embodiment according to the teleoperation context or scenario. We will apply the conceptual framework of embodied immersion in embodying any arbitrary object, just like this is usually done in telerobotics. We will show that enhanced embodiment is observable, measurable, reproducible, and most importantly, predictable, thanks to the modeling of active sensing and motor learning.

Since I started this thesis with a story, I would like to conclude it in the same way. I visited my grandmother a few months ago, and unfortunately it does not happen that often, since I live abroad. Therefore, I can pretty well experience the progression of her arthritis. The last time, she explained to me that she had completely lost tactile feedback from her hands. Now, my grandmother used to be a professional seamstress, and she still does it as hobby for family and friends. She still manages to accomplishe most of the tasks she could address before, but she is less precise and it takes more time. As stated in previous chapters, the concept of SoE is a complex abstraction, its effect on task performance does not appear to be straightforward to interpret, and its assessment is strongly task related. I find it fascinating that it applies to the teleoperation context, such as to any other sort of surrogate, or the self-body. I find it fascinating that it can be applied to everyday life examples, that everyone can get it, but none can truly understand and explain it. This is why I can conclude this thesis with my grandmother's story, being certain that she got what I did in the past four years, but she did not truly understand it, as most of the research on this topic.

Bibliography

[1] Daniel Aarno, Staffan Ekvall, and Danica Kragic. Adaptive virtual fixtures for machine-assisted teleoperation tasks. In *Proceedings of the 2005 IEEE international conference on robotics and automation*, pages 1139–1144. IEEE, 2005.

[2] Pooya Adami, Patrick B Rodrigues, Peter J Woods, Burcin Becerik-Gerber, Lucio Soibelman, Yasemin Copur-Gencturk, and Gale Lucas. Effectiveness of vr-based training on improving construction workers' knowledge, skills, and safety behavior in robotic teleoperation. *Advanced Engineering Informatics*, 50:101431, 2021.

[3] Manuel Aiple and Andre Schiele. Towards teleoperation with human-like dynamics: Human use of elastic tools. In *2017 IEEE World Haptics Conference (WHC)*, pages 171–176. IEEE, 2017.

[4] M Alimardani, S Nishio, and H Ishiguro. Bci-teleoperated androids; a study of embodiment and its effect on motor imagery learning. In *2015 IEEE 19th International Conference on Intelligent Engineering Systems (INES)*, pages 347–352. IEEE, 2015.

[5] Maryam Alimardani, Shuichi Nishio, and Hiroshi Ishiguro. Effect of biased feedback on motor imagery learning in bci-teleoperation system. *Frontiers in systems neuroscience*, 8:52, 2014.

[6] Nalini Ambady and Robert Rosenthal. Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of personality and social psychology*, 64(3):431, 1993.

[7] Robert J Anderson and Mark W Spong. Bilateral control of teleoperators with time delay. In *Proceedings of the 1988 IEEE International Conference on Systems, Man, and Cybernetics*, volume 1, pages 131–138. IEEE, 1988.

[8] Cédric Anthierens, Christine Libersa, Mohamed Touaibia, Maurice Bétemps, Marc Arsicault, and Nicolas Chaillet. Micro robots dedicated to small diameter canalization exploration. In *Proceedings. 2000 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2000)(Cat. No. 00CH37113)*, volume 1, pages 480–485. IEEE, 2000.

[9] Matthew AJ Apps and Manos Tsakiris. The free-energy self: A predictive coding account of self-recognition. *Neuroscience & Biobehavioral Reviews*, 41:85–97. https://doi.org/10.1016/j.neubiorev.2013.01.029, 2014.

[10] Michael Argyle. *Bodily communication*. Routledge, 2013.

[11] K Carrie Armel and Vilayanur S Ramachandran. Projecting sensations to external objects: evidence from skin conductance response. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(1523):1499–1506, 2003.

[12] Shahar Arzy, Gregor Thut, Christine Mohr, Christoph M Michel, and Olaf Blanke. Neural basis of embodiment: distinct contributions of temporoparietal junction and extrastriate body area. *Journal of Neuroscience*, 26(31):8074–8081, 2006.

[13] Tomohisa Asai. Agency elicits body-ownership: proprioceptive drift toward a synchronously acting external proxy. *Experimental brain research*, 234(5):1163–1174, 2016.

[14] Yochai Ataria. Sense of ownership and sense of agency during trauma. *Phenomenology and the Cognitive Sciences*, 14(1):199–212, 2015.

[15] Laura Aymerich-Franch, Damien Petit, Gowrishankar Ganesh, and Abderrahmane Kheddar. Embodiment of a humanoid robot is preserved during partial and delayed control. In *2015 IEEE International Workshop on Advanced Robotics and its Social Impacts (ARSO)*, pages 1–5. IEEE, 2015.

[16] Laura Aymerich-Franch, Damien Petit, Gowrishankar Ganesh, and Abderrahmane Kheddar. Non-human looking robot arms induce illusion of embodiment. *International Journal of Social Robotics*, 9(4):479–490, 2017.

[17] Jeremy N Bailenson, Nick Yee, Dan Merget, and Ralph Schroeder. The effect of behavioral realism and form realism of real-time avatar faces on verbal disclosure, nonverbal disclosure, emotion recognition, and copresence in dyadic interaction. *Presence: Teleoperators and Virtual Environments*, 15(4):359–372, 2006.

[18] Domna Banakou, Alejandro Beacco, Solène Neyret, Marta Blasco-Oliver, Sofia Seinfeld, and Mel Slater. Virtual body ownership and its consequences for implicit racial bias are dependent on social context. *Royal Society Open Science*, 7(12):201848, 2020.

[19] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics*, 1(1):71–81, 2009.

[20] Ori Ben-Porat, Moshe Shoham, and Joachim Meyer. Control design and task performance in endoscopic teleoperation. *Presence: Teleoperators & Virtual Environments*, 9(3):256–267, 2000.

[21] ES Beursken. Transparancy in bci: the effect of the mapping between an imagined movement and the resulting action on a user's sense of agency. 2012.

[22] Elisabetta Bevacqua, Igor Stanković, Ayoub Maatallaoui, Alexis Nédélec, and Pierre De Loor. Effects of coupling in human-virtual agent body interaction. In *International Conference on Intelligent Virtual Agents*, pages 54–63. Springer, 2014.

[23] Chris Bevan and Danaë Stanton Fraser. Shaking hands and cooperation in tele-present human-robot negotiation. In *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 247–254. IEEE, 2015.

[24] Omar S Bholat, Randy S Haluck, Willie B Murray, Paul J Gorman, and Thomas M Krummel. Tactile feedback is present during minimally invasive surgery. *Journal of the American College of Surgeons*, 189(4):349–355. https://doi.org/10.1016/S1072–7515(99)00184–2, 1999.

[25] Andreas Birk, Sören Schwertfeger, and Kaustubh Pathak. A networking framework for teleoperation in safety, security, and rescue robotics. *IEEE Wireless Communications*, 16(1):6–13, 2009.

[26] Olaf Blanke, Mel Slater, and Andrea Serino. Behavioral, neural, and computational

principles of bodily self-consciousness. *Neuron*, 88(1):145–166, 2015.

[27] Jim Blascovich, Jack Loomis, Andrew C Beall, Kimberly R Swinth, Crystal L Hoyt, and Jeremy N Bailenson. Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry*, 13(2):103–124, 2002.

[28] Jim Blascovich and Cade McCall. Social influence in virtual environments. 2013.

[29] Matthew Botvinick and Jonathan Cohen. Rubber hands 'feel'touch that eyes see. *Nature*, 391(6669):756–756. https://doi.org/10.1038/35784, 1998.

[30] Margaret M Bradley and Peter J Lang. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, 25(1):49–59, 1994.

[31] Cynthia Breazeal, Kerstin Dautenhahn, and Takayuki Kanda. Social robotics. In *Springer handbook of robotics*, pages 1935–1972. Springer, 2016.

[32] Simon Brodbeck. *The bhagavad gita*. Penguin UK, 2003.

[33] Claudio Brozzoli, H Henrik Ehrsson, and Alessandro Farnè. Multisensory representation of the space near the hand: from perception to action and interindividual interactions. *The Neuroscientist*, 20(2):122–135, 2014.

[34] Nicola Bruno and Marco Bertamini. Haptic perception after a change in hand size. *Neuropsychologia*, 48(6):1853–1856, 2010.

[35] Nitasha Buldeo. Interoception: A measure of embodiment or attention? *International Body Psychotherapy Journal*, 14(1), 2015.

[36] Dalila Burin, Francesca Garbarini, Valentina Bruno, Carlotta Fossataro, Cristina Destefanis, Anna Berti, and Lorenzo Pia. Movements and body ownership: evidence from the rubber hand illusion after mechanical limb immobilization. *Neuropsychologia*, 107:41–47, 2017.

[37] Jennifer L Burke, Matthew S Prewett, Ashley A Gray, Liuquin Yang, Frederick RB Stilson, Michael D Coovert, Linda R Elliot, and Elizabeth Redden. Comparing the effects of visual-auditory and visual-tactile feedback on user performance: a meta-analysis. In *Proceedings of the 8th international conference on Multimodal interfaces*, pages 108–117. https://doi.org/10.1145/1180995.1181017, 2006.

[38] Maria Eugenia Cabrera and Juan Pablo Wachs. A human-centered approach to one-shot gesture learning. *Frontiers in Robotics and AI*, 4:8, 2017.

[39] Angelo Cafaro, Hannes Högni Vilhjálmsson, Timothy Bickmore, Dirk Heylen, Kamilla Rún Jóhannsdóttir, and Gunnar Steinn Valgardsson. First impressions: Users' judgments of virtual agents' personality and interpersonal attitude in first encounters. In *International conference on intelligent virtual agents*, pages 67–80. Springer, 2012.

[40] Anne Campbell and J Philippe Rushton. Bodily communication and personality. *British Journal of Social and Clinical Psychology*, 17(1):31–36, 1978.

[41] Mark Carey, Laura Crucianelli, Catherine Preston, and Aikaterini Fotopoulou. The effect of visual capture towards subjective embodiment within the full body illusion. *Scientific reports*, 9(1):1–12, 2019.

[42] Emilie A Caspar, Axel Cleeremans, and Patrick Haggard. The relationship between human agency and embodiment. *Consciousness and cognition*, 33:226–236. https://doi.org/10.1016/j.concog.2015.01.007, 2015.

[43] Michael J Christoe, Jialuo Han, and Kourosh Kalantar-Zadeh. Telecommunications and data processing in flexible electronic systems. *Advanced Materials Technologies*, 5(1):1900733, 2020.

[44] Alain Codourey, Miguel Rodriguez, and Ion Pappas. A task-oriented teleoperation system for assembly in the microworld. In *1997 8th International Conference on Advanced Robotics. Proceedings. ICAR'97*, pages 235–240. IEEE, 1997.

[45] Silvia Coradeschi, Amy Loutfi, Annica Kristoffersson, Gabriella Cortellessa, and Kerstin Severinson Eklundh. Social robotic telepresence. In *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 5–6. IEEE, 2011.

[46] Kevin Corti and Alex Gillespie. Offscreen and in the chair next to your: conversational agents speaking through actual human bodies. In *Intelligent Virtual Agents*, pages 405–417. Springer, 2015.

[47] Kevin Corti and Alex Gillespie. A truly human interface: interacting face-to-face with someone whose words are determined by a computer program. *Frontiers in Psychology*, 6:634, 2015.

[48] Kevin Corti and Alex Gillespie. Co-constructing intersubjectivity with artificial conversational agents: people are more likely to initiate repairs of misunderstandings with agents represented as human. *Computers in Human Behavior*, 58:431–442, 2016.

[49] Kirsten Cowan and Seth Ketron. A dual model of product involvement for effective virtual reality: The roles of imagination, co-creation, telepresence, and interactivity. *Journal of Business Research*, 100:483–492, 2019.

[50] Damiano Crivelli, Elisa Polimeni, Daniele Crotti, Gabriella Bottini, and Gerardo Salvato. Bilateral skin temperature drop and warm sensibility decrease following modulation of body part ownership through mirror-box illusion. *Cortex*, 135:49–60, 2021.

[51] Laura Crucianelli, Charlotte Krahé, Paul M Jenkinson, and Aikaterini Katerina Fotopoulou. Interoceptive ingredients of body ownership: Affective touch and cardiac awareness in the rubber hand illusion. *Cortex*, 104:180–192, 2018.

[52] Jianhong Cui, Sabri Tosunoglu, Rodney Roberts, Carl Moore, and Daniel W Repperger. A review of teleoperation system control. In *Proceedings of the Florida Conference on Recent Advances in Robotics*, pages 1–12. Florida Atlantic University Boca Raton, FL, 2003.

[53] Nils Dahlbäck, Arne Jönsson, and Lars Ahrenberg. Wizard of Oz studies — why and how. *Knowledge-Based Systems*, 6(4):258–266, December 1993.

[54] J Dargahi and S Najarian. Human tactile perception as a standard for artificial tactile sensing—a review. *The international journal of medical robotics and computer assisted surgery*, 1(1):23–35. https://doi.org/10.1016/S1072–7515(99)00184–2, 2004.

[55] Nicole David, Bettina H Bewernick, Michael X Cohen, Albert Newen, Silke Lux, Gereon R Fink, N Jon Shah, and Kai Vogeley. Neural representations of self versus other: visual-spatial perspective taking and agency in a virtual ball-tossing game. *Journal of cognitive neuroscience*, 18(6):898–910, 2006.

[56] Mark H Davis et al. A multidimensional approach to individual differences in empathy. 1980.

[57] Alyanne M de Haan, Haike E Van Stralen, Miranda Smit, Anouk Keizer, Stefan Van der Stigchel, and H Chris Dijkerman. No consistent cooling of the real hand in the rubber

hand illusion. *Acta psychologica*, 179:68–77, 2017.

[58] Celso M. de Melo and Jonathan Gratch. Beyond believability: Quantifying the differences between real and virtual humans. In Willem-Paul Brinkman, Joost Broekens, and Dirk Heylen, editors, *Intelligent Virtual Agents*, pages 109–118, Cham, 2015. Springer International Publishing.

[59] Helena De Preester. A radical phenomenology of the body: subjectivity and sensations in body image and body schema. *Body Schema and Body Image: New Directions*, page 52, 2021.

[60] Frédérique De Vignemont. Body schema and body image—pros and cons. *Neuropsychologia*, 48(3):669–680, 2010.

[61] Frédérique De Vignemont. Embodiment, ownership and disownership. *Consciousness and cognition*, 20(1):82–93. https://doi.org/10.1016/j.concog.2010.09.004, 2011.

[62] Henrique G Debarba, Eray Molla, Bruno Herbelin, and Ronan Boulic. Characterizing embodied interaction in first and third person perspective viewpoints. In *2015 IEEE Symposium on 3D User Interfaces (3DUI)*, pages 67–72. IEEE, 2015.

[63] Henrique Galvan Debarba, Sidney Bovet, Roy Salomon, Olaf Blanke, Bruno Herbelin, and Ronan Boulic. Characterizing first and third person viewpoints and their alternation for embodied interaction in virtual reality. *PloS one*, 12(12), 2017.

[64] Maria Di Maro, Sara Falcone, and Francesco Cutugno. Prosodic analysis in human-machine interaction. *Studi AISV*, 1, 2018.

[65] Nicola Diolaiti and Claudio Melchiorri. Teleoperation of a mobile robot through haptic feedback. In *IEEE International Workshop HAVE Haptic Virtual Environments and Their*, pages 67–72. https://doi.org/10.1109/HAVE.2002.1106916. IEEE, 2002.

[66] Svetlana Drobyazko, Iryna Hryhoruk, Halyna Pavlova, Liudmyla Volchanska, and Sergiy Sergiychuk. Entrepreneurship innovation model for telecommunications enterprises. *Journal of Entrepreneurship Education*, 22(2):1–6, 2019.

[67] Timothy Dummer, Alexandra Picot-Annand, Tristan Neal, and Chris Moore. Movement and the rubber hand illusion. *Perception*, 38(2):271–280, 2009.

[68] Mariano D'Angelo, Giuseppe Di Pellegrino, and Francesca Frassinetti. Invisible body illusion modulates interpersonal space. *Scientific reports*, 7(1):1–9, 2017.

[69] H Henrik Ehrsson. The experimental induction of out-of-body experiences. *Science*, 317(5841):1048–1048, 2007.

[70] H Henrik Ehrsson, Nicholas P Holmes, and Richard E Passingham. Touching a rubber hand: Feeling of body ownership is associated with activity in multisensory brain areas. *Journal of neuroscience*, 25(45):10564–10573. https://doi.org/10.1523/JNEUROSCI.0800–05.2005, 2005.

[71] H Henrik Ehrsson, Charles Spence, and Richard E Passingham. That's my hand! activity in premotor cortex reflects feeling of ownership of a limb. *Science*, 305(5685):875–877, 2004.

[72] H Henrik Ehrsson, Katja Wiech, Nikolaus Weiskopf, Raymond J Dolan, and Richard E Passingham. Threatening a rubber hand that you feel is yours elicits a cortical anxiety response. *Proceedings of the National Academy of Sciences*, 104(23):9828–9833. https://doi.org/10.1073/pnas.0610011104, 2007.

[73] Albert Einstein. Zur Elektrodynamik bewegter Körper. (German) [On the electrodynamics of moving bodies]. *Annalen der Physik*, 322(10):891–921, 1905.

[74] Sara Falcone, Anne-Marie Brouwer, Ioana Cocu, Kaj Gijsbertse, Dirk Heylen, and Jan van Erp. The relative contribution of five key perceptual cues and their interaction to the sense of embodiment. *Technology, Mind, and Behavior*, 2022.

[75] Sara Falcone, Anne-Marie Brouwer, Dirk Heylen, Jan Van Erp, Liang Zhang, Saket Sachin Pradhan, Ivo V Stuldreher, Ioana Cocu, Martijn Heuvel, Pieter Simke de Vries, et al. Pupil diameter as implicit measure to estimate sense of embodiment. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 44, 2022.

[76] Sara Falcone, Gwenn Englebienne, Jan Van Erp, and Dirk Heylen. Toward standard guidelines to design the sense of embodiment in teleoperation applications: A review and toolbox. *Human–Computer Interaction*, pages 1–30, 2022.

[77] Sara Falcone, Jan Kolkmeier, Merijn Bruijnes, and Dirk Heylen. The multimodal echoborg: not as smart as it looks. *Journal on Multimodal User Interfaces*, 16(3):293–302, 2022.

[78] Sara Falcone, Saket Sachin Pradhan, Jan Van Erp, and Dirk Heylen. Individuals with high kinesthetic intelligence experience an active embodiment illusion assessed with pupil dilation. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 43, 2021.

[79] Chlöé Farrer and Chris D Frith. Experiencing oneself vs another person as being the cause of an action: the neural correlates of the experience of agency. *Neuroimage*, 15(3):596–603, 2002.

[80] Naomi T Fitter, Megan Strait, Eloise Bisbee, Maja J Mataric, and Leila Takayama. You're wigging me out! is personalization of telepresence robots strictly positive? In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, pages 168–176, 2021.

[81] Alessia Folegatti, Frederique De Vignemont, Francesco Pavani, Yves Rossetti, and Alessandro Farnè. Losing one's hand: visual-proprioceptive conflict affects touch perception. *PLoS One*, 4(9):e6920, 2009.

[82] Pierre-Pascal Forster, Harun Karimpur, and Katja Fiehler. Why we should rethink our approach to embodiment and presence. 2022.

[83] Carlotta Fossataro, Valentina Bruno, Patrizia Gindri, Lorenzo Pia, Anna Berti, and Francesca Garbarini. Feeling touch on the own hand restores the capacity to visually discriminate it from someone else'hand: Pathological embodiment receding in brain-damaged patients. *Cortex*, 104:207–219, 2018.

[84] Rebecca Fribourg, Ferran Argelaguet, Anatole Lécuyer, and Ludovic Hoyet. Avatar and sense of embodiment: Studying the relative preference between appearance, control and point of view. *IEEE transactions on visualization and computer graphics*, 26(5):2062–2072, 2020.

[85] Karl Friston. Prediction, perception and agency. *International Journal of Psychophysiology*, 83(2):248–252. https://doi.org/10.1016/j.ijpsycho.2011.11.014, 2012.

[86] Karl Friston and Stefan Kiebel. Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1521):1211–1221. https://doi.org/10.1098/rstb.2008.0300, 2009.

[87] Karl J Friston. Functional and effective connectivity: A review. *Brain connectivity*, 1(1):13–36. https://doi.org/10.1089/brain.2011.0008, 2011.

[88] Jakob Fröhner, Gionata Salvietti, Philipp Beckerle, and Domenico Prattichizzo. Can wearable haptic devices foster the embodiment of virtual limbs? *IEEE transactions on haptics*, 12(3):339–349, 2018.

[89] Christina T Fuentes, Mariella Pazzaglia, Matthew R Longo, Giorgio Scivoletto, and Patrick Haggard. Body image distortions following spinal cord injury. *Journal of Neurology, Neurosurgery & Psychiatry*, 84(2):201–207, 2013.

[90] Janez Funda, Russell H Taylor, Ben Eldridge, Stephen Gomory, and Kreg G Gruben. Constrained cartesian motion control for teleoperated surgical robots. *IEEE Transactions on Robotics and Automation*, 12(3):453–465, 1996.

[91] M Fusaro, G Tieri, and SM Aglioti. Influence of cognitive stance and physical perspective on subjective and autonomic reactivity to observed pain and pleasure: An immersive virtual reality study. *Consciousness and cognition*, 67:86–97, 2019.

[92] Martina Fusaro, Gaetano Tieri, and Salvatore Maria Aglioti. Seeing pain and pleasure on self and others: behavioral and psychophysiological reactivity in immersive virtual reality. *Journal of neurophysiology*, 116(6):2656–2662, 2016.

[93] Shaun Gallagher. Philosophical conceptions of the self: implications for cognitive science. *Trends in cognitive sciences*, 4(1):14–21. https://doi.org/10.1016/S1364–6613(99)01417–5, 2000.

[94] Shaun Gallagher. *Ambiguity in the sense of agency*. Oxford: Oxford University Press, 2013.

[95] Philippe Giguere and Gregory Dudek. A simple tactile probe for surface identification by mobile robots. *IEEE Transactions on Robotics*, 27(3):534–544. https://doi.org/10.1109/TRO.2011.2119910, 2011.

[96] Alex Gillespie and Kevin Corti. The body that speaks: recombining bodies and speech sources in unscripted face-to-face communication. *Frontiers in psychology*, 7:1300, 2016.

[97] Melita Joy Giummarra, BM Fitzgibbon, Nellie Georgiou-Karistianis, Megan Beukelman, A Verdejo-Garcia, Zachary Blumberg, M Chou, and Stephen J Gibson. Affective, sensory and empathic sharing of another's pain: The e mpathy for p ain s cale. *European Journal of Pain*, 19(6):807–816, 2015.

[98] Dean Giustini and Maged N Kamel Boulos. Google scholar is not enough to be used alone for systematic reviews. *Online journal of public health informatics*, 5(2):214, 2013.

[99] Claudia González, J Ernesto Solanes, Adolfo Munoz, Luis Gracia, Vicent Girbés-Juan, and Josep Tornero. Advanced teleoperation and control system for industrial robots based on augmented virtuality and haptic feedback. *Journal of Manufacturing Systems*, 59:283–298, 2021.

[100] Mar Gonzalez-Franco and Tabitha C Peck. Avatar embodiment. towards a standardized questionnaire. *Frontiers in Robotics and AI*, 5:74, 2018.

[101] Geoffrey Gorisse, Olivier Christmann, Etienne Armand Amato, and Simon Richir. First-and third-person perspectives in immersive virtual environments: Presence and performance analysis of embodied users. *Frontiers in Robotics and AI*, 4:33, 2017.

[102] Geoffrey Gorisse, Olivier Christmann, Samory Houzangbe, and Simon Richir. From robot to virtual doppelganger: Impact of visual fidelity of avatars controlled in third-person perspective on embodiment and behavior in immersive virtual environments. *Frontiers in Robotics and AI*, 6:8, 2019.

[103] Klaudia Grechuta, Jelena Guga, Giovanni Maffei, Belen Rubio Ballester, and Paul FMJ Verschure. Visuotactile integration modulates motor performance in a perceptual decision-making task. *Scientific reports*, 7(1):1–13, 2017.

[104] Weston B Griffin, William R Provancher, and Mark R Cutkosky. Feedback strategies for telemanipulation with shared control of object handling forces. *Presence: Teleoperators & Virtual Environments*, 14(6):720–731, 2005.

[105] Arvid Guterstam, Malin Björnsdotter, Giovanni Gentile, and H Henrik Ehrsson. Posterior cingulate cortex integrates the senses of self-location and body ownership. *Current Biology*, 25(11):1416–1425. https://doi.org/10.1016/j.cub.2015.03.059, 2015.

[106] Michael O Gutjahr, Wolfgang Ellermeier, Sandro Hardy, Stefan Göbel, and Josef Wiemeyer. The pupil response as an indicator of user experience in a digital exercise game. *Psychophysiology*, 56(10):e13418, 2019.

[107] Antal Haans, Wijnand A IJsselsteijn, and Yvonne AW de Kort. The effect of similarities in skin texture and hand shape on perceived ownership of a fake limb. *Body Image*, 5(4):389–394. https://doi.org/10.1016/j.bodyim.2008.04.003, 2008.

[108] Aarne Halme, Jussi Suomela, and Marko Savela. Applying telepresence and augmented reality to teleoperate field robots. *Robotics and autonomous systems*, 26(2-3):117–125, 1999.

[109] RE Hampson, Ioan Opris, and SA Deadwyler. Neural correlates of fast pupil dilation in nonhuman primates: relation to behavioral performance and cognitive workload. *Behavioural brain research*, 212(1):1–11, 2010.

[110] Blake Hannaford, Laurie Wood, Douglas A McAffee, and Haya Zak. Performance evaluation of a six-axis generalized force-reflecting teleoperator. *IEEE transactions on systems, man, and cybernetics*, 21(3):620–633, 1991.

[111] Yuval Noah Harari. Why fiction trumps truth. *New York Times*, 24, 2019.

[112] Sandra G Hart. Nasa-task load index (nasa-tlx); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 50, pages 904–908. Sage publications Sage CA: Los Angeles, CA, 2006.

[113] Christian Hatzfeld and Thorsten A Kern. *Engineering haptic devices*. Springer, 2016.

[114] Clint Heyer. Human-robot interaction and future industrial robotics applications. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4749–4754. IEEE, 2010.

[115] Hinze Hogendoorn, Marjolein PM Kammers, Thomas A Carlson, and Frans AJ Verstraten. Being in the dark about your hand: Resolution of visuo-proprioceptive conflict by disowning visible limbs. *Neuropsychologia*, 47(13):2698–2703, 2009.

[116] Maarten A Hogervorst, Anne-Marie Brouwer, and Jan BF Van Erp. Combining and comparing eeg, peripheral physiology and eye-related measures for the assessment of mental workload. *Frontiers in neuroscience*, 8:322, 2014.

[117] Jakob Hohwy. *The predictive mind*. Oxford University Press, 2013.

[118] Jakob Hohwy and Bryan Paton. Explaining away the body: Experiences of supernaturally caused touch and touch on non-hand objects within the rubber hand illusion. *PloS one*, 5(2), 2010.

[119] Adria EN Hoover and Laurence R Harris. Inducing ownership over an 'other'perspective with a visuo-tactile manipulation. *Experimental brain research*, 234(12):3633–3639, 2016.

[120] Robert D Howe and Dimitrios A Kontarinis. Task performance with a dexterous teleoperated hand system. In *Telemanipulator Technology*, volume 1833, pages 199–207. International Society for Optics and Photonics, 1993.

[121] WA IJsselsteijn, YAW De Kort, and Karolien Poels. The game experience questionnaire. 2013.

[122] Shafiqul Islam, PX Liu, Abdulmotaleb El Saddik, and Yubin B Yang. Bilateral control of teleoperation systems with time delay. *IEEE/ASME Transactions on Mechatronics*, 20(1):1–12, 2014.

[123] Siddhartha Joshi, Yin Li, Rishi M Kalwani, and Joshua I Gold. Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. *Neuron*, 89(1):221–234, 2016.

[124] Andreas Kalckert and H Henrik Ehrsson. Moving a rubber hand that feels like your own: a dissociation of ownership and agency. *Frontiers in human neuroscience*, 6:40, 2012.

[125] Andreas Kalckert and H Henrik Ehrsson. The moving rubber hand illusion revisited: Comparing movements and visuotactile stimulation to induce illusory ownership. *Consciousness and cognition*, 26:117–132, 2014.

[126] Andreas Kalckert and HH Ehrsson. The onset time of the ownership sensation in the moving rubber hand illusion. *Frontiers in psychology*, 8:344, 2017.

[127] Marjolein PM Kammers, Frederique de Vignemont, Lennart Verhagen, and H Chris Dijkerman. The rubber hand illusion in action. *Neuropsychologia*, 47(1):204–211, 2009.

[128] Daisuke Kaneko, IV Stuldreher, Anne Reuten, Alexander Toet, Jan BF Van Erp, and Anne-Marie Brouwer. Comparing explicit and implicit measures for assessing cross-cultural food experience. *Frontiers in Neuroergonomics*, 2(2), 2021.

[129] Dominic Kao. The effects of anthropomorphic avatars vs. non-anthropomorphic avatars in a jumping game. In *Proceedings of the 14th International Conference on the Foundations of Digital Games*, pages 1–5, 2019.

[130] Mohamed Zouheir Kastouni and Ayoub Ait Lahcen. Big data analytics in telecommunications: Governance, architecture and use cases. *Journal of King Saud University-Computer and Information Sciences*, 2020.

[131] Jari Kätsyri, Riitta Hari, Niklas Ravaja, and Lauri Nummenmaa. The opponent matters: elevated fmri reward responses to winning against a human versus a computer opponent during interactive video game playing. *Cerebral Cortex*, 23(12):2829–2839, 2013.

[132] Konstantina Kilteni, Raphaela Groten, and Mel Slater. The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments*, 21(4):373–387, 2012.

[133] Konstantina Kilteni, Antonella Maselli, Konrad P Kording, and Mel Slater. Over my fake body: Body ownership illusions for studying the multisensory basis of own-body perception. *Frontiers in human neuroscience*, 9:141. https://doi.org/10.3389/fnhum.2015.00141, 2015.

[134] Günther Knoblich and Tilo TJ Kircher. Deceiving oneself about being in control: conscious detection of changes in visuomotor coupling. *Journal of Experimental Psychology: Human Perception and Performance*, 30(4):657, 2004.

[135] Elena Kokkinara, Konstantina Kilteni, Kristopher J Blom, and Mel Slater. First person perspective of seated participants over a walking virtual body leads to illusory agency over the walking. *Scientific reports*, 6(1):1–11, 2016.

[136] Elena Kokkinara and Mel Slater. Measuring the effects through time of the influence of visuomotor and visuotactile synchronous stimulation on a virtual body ownership illusion. *Perception*, 43(1):43–58. https://doi.org/10.1068/p7545, 2014.

[137] Elena Kokkinara, Mel Slater, and Joan López-Moliner. The effects of visuomotor calibration to the perceived space and body, through embodiment in immersive virtual reality. *ACM Transactions on Applied Perception (TAP)*, 13(1):1–22, 2015.

[138] Jan Kolkmeier, Merijn Bruijnes, Dennis Reidsma, and Dirk Heylen. An asap realizer-unity3d bridge for virtual and mixed reality applications. In *International Conference on Intelligent Virtual Agents*, pages 227–230. Springer, 2017.

[139] Jan Kolkmeier, Jered Vroon, and Dirk Heylen. Interacting with virtual agents in shared space: Single and joint effects of gaze and proxemics. In *International Conference on Intelligent Virtual Agents*, pages 1–14. Springer, 2016.

[140] Ryota Kondo, Maki Sugimoto, Kouta Minamizawa, Takayuki Hoshi, Masahiko Inami, and Michiteru Kitazaki. Illusory body ownership of an invisible body interpolated between virtual hands and feet via visual-motor synchronicity. *Scientific reports*, 8(1):1–8. https://doi.org/10.1038/s41598–018–25951–2, 2018.

[141] Kean Kouakoua, Cyril Duclos, Rachid Aissaoui, Sylvie Nadeau, and David R Labbe. Rhythmic proprioceptive stimulation improves embodiment in a walking avatar when added to visual stimulation. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 573–574. IEEE, 2020.

[142] John W Krakauer, Alkis M Hadjiosif, Jing Xu, Aaron L Wong, and Adrian M Haith. Motor learning. *Compr Physiol*, 9(2):613–663, 2019.

[143] Claudia Krogmeier, Christos Mousas, and David Whittinghill. Human–virtual character interaction: Toward understanding the influence of haptic feedback. *Computer Animation and Virtual Worlds*, 30(3-4):e1883. https://doi.org/10.1002/cav.1883, 2019.

[144] Bouke N Krom, Milène Catoire, Alexander Toet, Roelof JE van Dijk, and Jan BF van Erp. Effects of likeness and synchronicity on the ownership illusion over a moving virtual robotic arm and hand. In *2019 IEEE World Haptics Conference (WHC)*, pages 49–54. IEEE, 2019.

[145] Philipp Kulms and Stefan Kopp. More human-likeness, more trust? the effect of anthropomorphism on self-reported and behavioral trust in continued and interdependent human-agent cooperation. In *Proceedings of mensch und computer 2019*, pages 31–42. 2019.

[146] Timothy Lane, Su-Ling Yeh, Philip Tseng, and An-Yi Chang. Timing disownership ex-

periences in the rubber hand illusion. *Cognitive research: principles and implications*, 2(1):4, 2017.

[147] John Lawrence, Mark Snaith, Barbara Konat, Katarzyna Budzynska, and Chris Reed. Debating technology for dialogical argument: Sensemaking, engagement, and analytics. *ACM Transactions on Internet Technology (TOIT)*, 17(3):1–23, 2017.

[148] Nam-In Lee. *Edmund Husserls Phänomenologie der Instinkte*, volume 128. Springer-Verlag, 2013.

[149] Bigna Lenggenhager, Michael Mouthon, and Olaf Blanke. Spatial aspects of bodily self-consciousness. *Consciousness and cognition*, 18(1):110–117, 2009.

[150] Bigna Lenggenhager, Tej Tadi, Thomas Metzinger, and Olaf Blanke. Video ergo sum: manipulating bodily self-consciousness. *Science*, 317(5841):1096–1099, 2007.

[151] Maurice J Levesque and David A Kenny. Accuracy of behavioral predictions at zero acquaintance: A social relations analysis. *Journal of Personality and Social Psychology*, 65(6):1178, 1993.

[152] Elizabeth Lewis and Donna M Lloyd. Embodied experience: A first-person investigation of the rubber hand illusion. *Phenomenology and the Cognitive Sciences*, 9(3):317–339, 2010.

[153] Jamy Li. The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents. *International Journal of Human-Computer Studies*, 77:23–37, 2015.

[154] Minas V Liarokapis, Panagiotis K Artemiadis, and Kostas J Kyriakopoulos. Mapping human to robot motion with functional anthropomorphism for teleoperation and telemanipulation with robot arm hand systems. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2075–2075. IEEE, 2013.

[155] Sohye Lim and Byron Reeves. Computer agents versus avatars: Responses to interactive game characters controlled by a computer or other player. *International Journal of Human-Computer Studies*, 68(1-2):57–68, 2010.

[156] Jakub Limanowski and Felix Blankenburg. Integration of visual and proprioceptive limb position information in human posterior parietal, premotor, and extrastriate cortex. *Journal of Neuroscience*, 36(9):2582–2589, 2016.

[157] Jakub Limanowski and Felix Blankenburg. That's not quite me: limb ownership encoding in the brain. *Social cognitive and affective neuroscience*, 11(7):1130–1140, 2016.

[158] John M Linebarger and G Drew Kessler. The effect of avatar connectedness on task performance. *Lehigh Univ TR*, 2002.

[159] Yizhi Liu, Mahmoud Habibnezhad, and Houtan Jebelli. Brain-computer interface for hands-free teleoperation of construction robots. *Automation in Construction*, 123:103523, 2021.

[160] Matthew R Longo, Friederike Schüür, Marjolein PM Kammers, Manos Tsakiris, and Patrick Haggard. What is embodiment? a psychometric approach. *Cognition*, 107(3):978–998, 2008.

[161] Tomás Lozano-Pérez, Joseph L. Jones, Emmanuel Mazer, and Patrick A. O'Donnell. Task-level planning of pick-and-place robot motions. *Computer*, 22(3):21–29, 1989.

[162] Penny A MacDonald and Tomáš Paus. The role of parietal cortex in awareness of

self-generated movements: a transcranial magnetic stimulation study. *Cerebral Cortex*, 13(9):962–967, 2003.

[163] Kazuo Machida, Kenzo Akita, Keitaro Ohno, Masayoshi Moriya, Hirotaka Nishida, and Tsutomu Ohsawa. Space experiment evaluation of advanced robotic hand system boarded on ets-vii. *The Journal of Space Technology and Science*, 14(2):2_21–2_30, 1998.

[164] Akhil J Madhani, Günter Niemeyer, and J Kenneth Salisbury. The black falcon: a teleoperated surgical instrument for minimally invasive surgery. In *Proceedings. 1998 IEEE/RSJ International Conference on Intelligent Robots and Systems. Innovations in Theory, Practice and Applications (Cat. No. 98CH36190)*, volume 2, pages 936–944. IEEE, 1998.

[165] Gary Brent Madison. The phenomenology of merleau-ponty: A search for the limits of consciousness. 1981.

[166] Tamar R Makin, Nicholas P Holmes, and H Henrik Ehrsson. On the other hand: Dummy hands and peripersonal space. *Behavioural brain research*, 191(1):1–10. https://doi.org/10.1016/j.bbr.2008.02.041, 2008.

[167] Paul D Marasco, Jacqueline S Hebert, Jon W Sensinger, Courtney E Shell, Jonathon S Schofield, Zachary C Thumser, Raviraj Nataraj, Dylan T Beckler, Michael R Dawson, Dan H Blustein, et al. Illusory movement perception improves motor control for prosthetic hands. *Science translational medicine. https://doi.org/10.1126/scitranslmed.aao6990*, 10(432), 2018.

[168] Antonella Maselli and Mel Slater. Sliding perspectives: dissociating ownership from self-location during full body illusions in virtual reality. *Frontiers in human neuroscience*, 8:693, 2014.

[169] Michael J Massimino and Thomas B Sheridan. Sensory substitution for force feedback in teleoperation. *Presence: Teleoperators & Virtual Environments*, 2(4):344–352, 1993.

[170] Michael J Massimino and Thomas B Sheridan. Teleoperator performance with varying force and visual feedback. *Human factors*, 36(1):145–157, 1994.

[171] Samuel D McDougle, Krista M Bond, and Jordan A Taylor. Explicit and implicit processes constitute the fast and slow processes of sensorimotor learning. *Journal of Neuroscience*, 35(26):9568–9579, 2015.

[172] Samuel D McDougle and Jordan A Taylor. Dissociable cognitive strategies for sensorimotor learning. *Nature communications*, 10(1):40, 2019.

[173] Mitchell McEwan, Daniel Johnson, Peta Wyeth, and Alethea Blackler. Videogame control device impact on the play experience. In *Proceedings of The 8th Australasian Conference on Interactive Entertainment: Playing the System*, pages 1–3, 2012.

[174] Maurice Merleau-Ponty. *Phenomenology of perception*. Routledge, 2013.

[175] Thomas Metzinger. Why are out-of-body experiences interesting for philosophers?: The theoretical relevance of obe research. *cortex*, 45(2):256–258, 2009.

[176] Sébastien Mick, Arnaud Badets, Pierre-Yves Oudeyer, Daniel Cattaert, and Aymar De Rugy. Biological plausibility of arm postures influences the controllability of robotic arm teleoperation. *Human Factors*, page 0018720820941619, 2020.

[177] Stanley Milgram and L van Gasteren. *Das Milgram-Experiment*.) Rowohlt, 1974.

[178] Sarah Miller. Workload measures. *National Advanced Driving Simulator. Iowa City, United States*, 2001.

[179] Nooralisa Mohd Tuah, Gary Wills, and Ashok Ranchhod. The characteristics and application of anthropomorphic interface: A design spectrum. 2016.

[180] G Lorimer Moseley, Nick Olthof, Annemeike Venema, Sanneke Don, Marijke Wijers, Alberto Gallace, and Charles Spence. Psychologically induced cooling of a specific body part caused by the illusory ownership of an artificial counterpart. *Proceedings of the National Academy of Sciences*, 105(35):13169–13173, 2008.

[181] Michael Neff, Nicholas Toothman, Robeson Bowmani, Jean E Fox Tree, and Marilyn A Walker. Don't scratch! self-adaptors reflect emotional stability. In *International Workshop on Intelligent Virtual Agents*, pages 398–411. Springer, 2011.

[182] Michael Neff, Yingying Wang, Rob Abbott, and Marilyn Walker. Evaluating the effect of gesture and language on personality perception in conversational agents. In *International Conference on Intelligent Virtual Agents*, pages 222–235. Springer, 2010.

[183] Roger Newport, Rachel Pearce, and Catherine Preston. 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, 2010.

[184] Roger Newport and Catherine Preston. Pulling the finger off disrupts agency, embodiment and peripersonal space. *Perception*, 39(9):1296–1298, 2010.

[185] Günter Niemeyer, Carsten Preusche, Stefano Stramigioli, and Dongjun Lee. Telerobotics. In *Springer handbook of robotics*, pages 1085–1108. Springer, 2016.

[186] Jean-Marie Normand, Elias Giannopoulos, Bernhard Spanlang, and Mel Slater. Multisensory stimulation can induce an illusion of larger belly size in immersive virtual reality. *PloS one*, 6(1), 2011.

[187] Aline Normoyle, Fannie Liu, Mubbasir Kapadia, Norman I Badler, and Sophie Jörg. The effect of posture and dynamics on the perception of emotion. In *Proceedings of the ACM Symposium on Applied Perception*, pages 91–98. ACM, 2013.

[188] Kristine L Nowak and Frank Biocca. The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 12(5):481–494, 2003.

[189] Kristine L Nowak and Jesse Fox. Avatars and computer-mediated communication: A review of the definitions, uses, and effects of digital representations. *Review of Communication Research*, 6:30–53, 2018.

[190] Kenji Ogawa and Toshio Inui. Lateralization of the posterior parietal cortex for internal monitoring of self-versus externally generated movements. *Journal of Cognitive Neuroscience*, 19(11):1827–1835, 2007.

[191] Allison M Okamura. Methods for haptic feedback in teleoperated robot-assisted surgery. *Industrial Robot: An International Journal*, 2004.

[192] Sandra Y Okita, Jeremy Bailenson, and Daniel L Schwartz. The mere belief of social interaction improves learning. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 29, 2007.

[193] Manuel Oliva. Pupil size and search performance in low and high perceptual load. *Cognitive, Affective, & Behavioral Neuroscience*, 19(2):366–376, 2019.

[194] Manuel Oliva and Andrey Anikin. Pupil dilation reflects the time course of emotion recognition in human vocalizations. *Scientific reports*, 8(1):4871, 2018.

[195] Jaeheung Park and Oussama Khatib. A haptic teleoperation approach based on contact force control. *The International Journal of Robotics Research*, 25(5-6):575–591, 2006.

[196] Francesco Pavani, Charles Spence, and Jon Driver. Visual capture of touch: Out-of-the-body experiences with rubber gloves. *Psychological science*, 11(5):353–359, 2000.

[197] Enea Francesco Pavone, Gaetano Tieri, Giulia Rizza, Emmanuele Tidoni, Luigi Grisoni, and Salvatore Maria Aglioti. Embodying others in immersive virtual reality: electro-cortical signatures of monitoring the errors in the actions of an avatar seen from a first-person perspective. *Journal of Neuroscience*, 36(2):268–279, 2016.

[198] Tabitha C Peck and Mar Gonzalez-Franco. Avatar embodiment. a standardized questionnaire. *Frontiers in Virtual Reality*, 1:44, 2021.

[199] Daniel Perez-Marcos, Matteo Martini, Christina T Fuentes, Anna I Bellido Rivas, Patrick Haggard, and Maria V Sanchez-Vives. Selective distortion of body image by asynchronous visuotactile stimulation. *Body Image*, 24:55–61, 2018.

[200] Daniel Perez-Marcos, Maria V Sanchez-Vives, and Mel Slater. Is my hand connected to my body? the impact of body continuity and arm alignment on the virtual hand illusion. *Cognitive neurodynamics*, 6(4):295–305, 2012.

[201] Valeria I Petkova and H Henrik Ehrsson. If i were you: perceptual illusion of body swapping. *PloS one*, 3(12), 2008.

[202] Nicolas Pouliot and Serge Montambault. Geometric design of the linescout, a teleoperated robot for power line inspection and maintenance. In *2008 IEEE International Conference on Robotics and Automation*, pages 3970–3977. IEEE, 2008.

[203] Polona Pozeg, Giulia Galli, and Olaf Blanke. Those are your legs: the effect of visuo-spatial viewpoint on visuo-tactile integration and body ownership. *Frontiers in psychology*, 6:1749, 2015.

[204] Catherine Preston and H Henrik Ehrsson. Illusory obesity triggers body dissatisfaction responses in the insula and anterior cingulate cortex. *Cerebral Cortex*, 26(12):4450–4460, 2016.

[205] Maria Pyasik, Dalila Burin, and Lorenzo Pia. On the relation between body ownership and sense of agency: A link at the level of sensory-related signals. *Acta Psychologica*, 185:219–228. https://doi.org/10.1016/j.actpsy.2018.03.001, 2018.

[206] Whitney Quesenbery. What does usability mean: Looking beyondease of use'. In *Annual conference-society for technical communication*, volume 48, pages 432–436. Citeseer, 2001.

[207] Carolyn Ranti, Warren Jones, Ami Klin, and Sarah Shultz. Blink rate patterns provide a reliable measure of individual engagement with scene content. *Scientific reports*, 10(1):8267, 2020.

[208] Arran T Reader and H Henrik Ehrsson. Weakening the subjective sensation of own hand ownership does not interfere with rapid finger movements. *PloS one*, 14(10):e0223580, 2019.

[209] Joel Michael Reynolds. Merleau-ponty, world-creating blindness, and the phenomenology of non-normate bodies. *Merleau-Ponty, world-creating blindness, and the phe-*

nomenology of non-normate bodies, pages 419–436, 2018.

[210] Grégoire Richard, Thomas Pietrzak, Ferran Argelaguet, Anatole Lécuyer, and Géry Casiez. Studying the role of haptic feedback on virtual embodiment in a drawing task. *Frontiers in Virtual Reality*, 1:28, 2021.

[211] Martin Riemer, Florian Bublatzky, Jörg Trojan, and Georg W Alpers. Defensive activation during the rubber hand illusion: Ownership versus proprioceptive drift. *Biological psychology*, 109:86–92, 2015.

[212] Martin Riemer, Jörg Trojan, Marta Beauchamp, and Xaver Fuchs. The rubber hand universe: On the impact of methodological differences in the rubber hand illusion. *Neuroscience & Biobehavioral Reviews*, 104:268–280, 2019.

[213] Marieke Rohde, Andrew Wold, Hans-Otto Karnath, and Marc O Ernst. The human touch: skin temperature during the rubber hand illusion in manual and automated stroking procedures. *PloS one*, 8(11):e80688, 2013.

[214] Daniele Romano, Elisa Caffa, Alejandro Hernandez-Arieta, Peter Brugger, and Angelo Maravita. The robot hand illusion: Inducing proprioceptive drift through visuo-motor congruency. *Neuropsychologia*, 70:414–420, 2015.

[215] Septimiu E Salcudean. Control for teleoperation and haptic interfaces. In *Control problems in robotics and automation*, pages 51–66. Springer, 1998.

[216] Septimiu E Salcudean, S Ku, and G Bell. Performance measurement in scaled teleoperation for microsurgery. In *CVRMed-MRCAS'97*, pages 789–798. Springer, 1997.

[217] Maria V Sanchez-Vives and Mel Slater. From presence to consciousness through virtual reality. *Nature Reviews Neuroscience*, 6(4):332–339, 2005.

[218] Maria V Sanchez-Vives, Bernhard Spanlang, Antonio Frisoli, Massimo Bergamasco, and Mel Slater. Virtual hand illusion induced by visuomotor correlations. *PloS one*, 5(4):e10381, 2010.

[219] Atsushi Sato and Asako Yasuda. Illusion of sense of self-agency: discrepancy between the predicted and actual sensory consequences of actions modulates the sense of self-agency, but not the sense of self-ownership. *Cognition*, 94(3):241–255, 2005.

[220] Matthew Schiefer, Daniel Tan, Steven M Sidek, and Dustin J Tyler. Sensory feedback by peripheral nerve stimulation improves task performance in individuals with upper limb loss using a myoelectric prosthesis. *Journal of neural engineering*, 13(1):016001, 2015.

[221] Sebastian Schostek, Marc O Schurr, and Gerhard F Buess. Review on aspects of artificial tactile feedback in laparoscopic surgery. *Medical engineering & physics*, 31(8):887–898. https://doi.org/10.1016/j.medengphy.2009.06.003, 2009.

[222] Lars Schwabe and Olaf Blanke. Cognitive neuroscience of ownership and agency. *Consciousness and cognition*, 16(ARTICLE):661–6, 2007.

[223] Valentin Schwind, Pascal Knierim, Cagri Tasci, Patrick Franczak, Nico Haas, and Niels Henze. " these are not my hands!" effect of gender on the perception of avatar hands in virtual reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 1577–1582, 2017.

[224] S Seghezzi. Modulating the sense of agency in the human brain: an fmri-guided tms study. In *European Workshop on Cognitive Neuropsychology.*, 2019.

[225] Ali Sengül, Giulio Rognini, Michiel van Elk, Jane Elizabeth Aspell, Hannes Bleuler, and Olaf Blanke. Force feedback facilitates multisensory integration during robotic tool use. *Experimental brain research*, 227(4):497–507, 2013.

[226] Silvia Serino, Elisa Pedroli, Anouk Keizer, Stefano Triberti, Antonios Dakanalis, Federica Pallavicini, Alice Chirico, and Giuseppe Riva. Virtual reality body swapping: a tool for modifying the allocentric memory of the body. *Cyberpsychology, Behavior, and Social Networking*, 19(2):127–133, 2016.

[227] Thomas B Sheridan. Space teleoperation through time delay: Review and prognosis. *IEEE Transactions on robotics and Automation*, 9(5):592–606, 1993.

[228] Thomas B Sheridan. Teleoperation, telerobotics and telepresence: A progress report. *Control Engineering Practice*, 3(2):205–214. https://doi.org/10.1016/0967–0661(94)00078–U, 1995.

[229] Mincheol Shin, Sanguk Lee, Stephen W Song, and Donghun Chung. Enhancement of perceived body ownership in virtual reality-based teleoperation may backfire in the execution of high-risk tasks. *Computers in Human Behavior*, 115:106605, 2021.

[230] Metin Sitti. Survey of nanomanipulation systems. In *Proceedings of the 2001 1st IEEE Conference on Nanotechnology. IEEE-NANO 2001 (Cat. No. 01EX516)*, pages 75–80. https://doi.org/10.1109/NANO.2001.966397. IEEE, 2001.

[231] Metin Sitti, Hideki Hashimoto, and O Kaynak. Teleoperated nano scale object manipulation. *Recent Advances on Mechatronics*, pages 322–335, 1999.

[232] Filip Škola and Fotis Liarokapis. 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, 2016.

[233] Mel Slater, Daniel Pérez Marcos, Henrik Ehrsson, and Maria V Sanchez-Vives. Towards a digital body: the virtual arm illusion. *Frontiers in human neuroscience*, 2:6. https://doi.org/10.3389/neuro.09.006.2008, 2008.

[234] Mel Slater, Daniel Pérez Marcos, Henrik Ehrsson, and Maria V Sanchez-Vives. Inducing illusory ownership of a virtual body. *Frontiers in neuroscience*, 3:29, 2009.

[235] Mel Slater, Bernhard Spanlang, Maria V Sanchez-Vives, and Olaf Blanke. First person experience of body transfer in virtual reality. *PloS one*, 5(5), 2010.

[236] Harrison Jesse Smith and Michael Neff. Understanding the impact of animated gesture performance on personality perceptions. *ACM Transactions on Graphics (TOG)*, 36(4):49, 2017.

[237] Jonathan A Smith. Beyond the divide between cognition and discourse: Using interpretative phenomenological analysis in health psychology. *Psychology and health*, 11(2):261–271, 1996.

[238] Aiguo Song, Dan Morris, and J Edward Colgate. Haptic telemanipulation of soft environment without direct force feedback. In *2005 IEEE International Conference on Information Acquisition*, pages 5–pp. IEEE, 2005.

[239] Maximilian Speicher, Brian D Hall, and Michael Nebeling. What is mixed reality? In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pages 1–15, 2019.

[240] Matthis Synofzik, Gottfried Vosgerau, and Albert Newen. I move, therefore i am: A

new theoretical framework to investigate agency and ownership. *Consciousness and cognition*, 17(2):411–424. https://doi.org/10.1016/j.concog.2008.03.008, 2008.

[241] Zoltán Szántó, Lőrinc Márton, Piroska Haller, and Sebestyén György. Performance analysis of wlan based mobile robot teleoperation. In *2013 IEEE 9th International Conference on Intelligent Computer Communication and Processing (ICCP)*, pages 299–305. IEEE, 2013.

[242] Jordan A Taylor, John W Krakauer, and Richard B Ivry. Explicit and implicit contributions to learning in a sensorimotor adaptation task. *Journal of Neuroscience*, 34(8):3023–3032, 2014.

[243] Judith Jarvis Thomson. Killing, letting die, and the trolley problem. *The Monist*, 59(2):204–217, 1976.

[244] Gaetano Tieri, Annamaria Gioia, Michele Scandola, Enea F Pavone, and Salvatore M Aglioti. Visual appearance of a virtual upper limb modulates the temperature of the real hand: a thermal imaging study in immersive virtual reality. *European Journal of Neuroscience*, 45(9):1141–1151, 2017.

[245] Gaetano Tieri, Emmanuele Tidoni, Enea Francesco Pavone, and Salvatore Maria Aglioti. Body visual discontinuity affects feeling of ownership and skin conductance responses. *Scientific reports*, 5:17139, 2015.

[246] A Toet, IA Kuling, BN Krom, and JBF van Erp. Toward enhanced teleoperation through embodiment. *Front. Robot. AI 7: 14. doi: 10.3389/frobt*, 2020.

[247] William T Townsend and Jeffrey A Guertin. Teleoperator slave-wam design methodology. *Industrial Robot: An International Journal*, 1999.

[248] Manos Tsakiris. My body in the brain: a neurocognitive model of body-ownership. *Neuropsychologia*, 48(3):703–712, 2010.

[249] Manos Tsakiris, Marcello Costantini, and Patrick Haggard. The role of the right temporo-parietal junction in maintaining a coherent sense of one's body. *Neuropsychologia*, 46(12):3014–3018, 2008.

[250] Manos Tsakiris and Patrick Haggard. The rubber hand illusion revisited: Visuotactile integration and self-attribution. *Journal of Experimental Psychology: Human Perception and Performance*, 31(1):80. https://psycnet.apa.org/doi/10.1037/0096–1523.31.1.80, 2005.

[251] Manos Tsakiris, Matthew R Longo, and Patrick Haggard. Having a body versus moving your body: neural signatures of agency and body-ownership. *Neuropsychologia*, 48(9):2740–2749, 2010.

[252] Manos Tsakiris, Gita Prabhu, and Patrick Haggard. Having a body versus moving your body: How agency structures body-ownership. *Consciousness and cognition*, 15(2):423–432, 2006.

[253] Esther Van Den Bos and Marc Jeannerod. Sense of body and sense of action both contribute to self-recognition. *Cognition*, 85(2):177–187, 2002.

[254] Björn Van der Hoort and H Henrik Ehrsson. Body ownership affects visual perception of object size by rescaling the visual representation of external space. *Attention, Perception, & Psychophysics*, 76(5):1414–1428, 2014.

[255] Björn Van Der Hoort and H Henrik Ehrsson. Illusions of having small or large invisible

bodies influence visual perception of object size. *Scientific reports*, 6(1):1–9, 2016.

[256] Björn Van Der Hoort, Arvid Guterstam, and H Henrik Ehrsson. Being barbie: the size of one's own body determines the perceived size of the world. *PloS one*, 6(5):e20195, 2011.

[257] Jan BF Van Erp, Camille Sallaberry, Christiaan Brekelmans, Douwe Dresscher, Frank Ter Haar, Gwenn Englebienne, Jeanine Van Bruggen, Joachim De Greeff, Leonor Fermoselle Silva Pereira, Alexander Toet, et al. What comes after telepresence? embodiment, social presence and transporting one's functional and social self. In *2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 2067–2072. IEEE, 2022.

[258] Jelte van Waterschoot, Merijn Bruijnes, Jan Flokstra, Dennis Reidsma, Daniel Davison, Mariët Theune, and Dirk Heylen. Flipper 2.0: A Pragmatic Dialogue Engine for Embodied Conversational Agents. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents*, IVA '18, pages 43–50, Sydney, NSW, Australia, 2018. ACM.

[259] Herwin Van Welbergen, Dennis Reidsma, and Stefan Kopp. An incremental multimodal realizer for behavior co-articulation and coordination. In *International Conference on Intelligent Virtual Agents*, pages 175–188. Springer, 2012.

[260] Terez A Varkonyi, Imre J Rudas, Peter Pausits, and Tamas Haidegger. Survey on the control of time delay teleoperation systems. In *IEEE 18th International Conference on Intelligent Engineering Systems INES 2014*, pages 89–94. IEEE, 2014.

[261] Pim Verhagen, Irene Kuling, Kaj Gijsbertse, Ivo V Stuldreher, Krista Overvliet, Sara Falcone, Jan Van Erp, and Anne-Marie Brouwer. The cross-modal congruency effect as an objective measure of embodiment. In *Companion Publication of the 2020 International Conference on Multimodal Interaction*, pages 107–111, 2020.

[262] Joshua Wainer, David J Feil-Seifer, Dylan A Shell, and Maja J Mataric. Embodiment and human-robot interaction: A task-based perspective. In *RO-MAN 2007-The 16th IEEE International Symposium on Robot and Human Interactive Communication*, pages 872–877. IEEE, 2007.

[263] Chin-An Wang, Talia Baird, Jeff Huang, Jonathan D Coutinho, Donald C Brien, and Douglas P Munoz. Arousal effects on pupil size, heart rate, and skin conductance in an emotional face task. *Frontiers in neurology*, 9:1029, 2018.

[264] Kunlin Wei and Konrad Kording. Relevance of error: what drives motor adaptation? *Journal of neurophysiology*, 101(2):655–664, 2009.

[265] David Weibel, Bartholomäus Wissmath, Stephan Habegger, Yves Steiner, and Rudolf Groner. Playing online games against computer-vs. human-controlled opponents: Effects on presence, flow, and enjoyment. *Computers in Human Behavior*, 24(5):2274–2291, 2008.

[266] Nicolas Wenk, J Penalver-Andres, KA Buetler, Tobias Nef, René Martin Müri, and Laura Marchal-Crespo. Effect of immersive visualization technologies on cognitive load, motivation, usability, and embodiment. *Virtual Reality*, pages 1–25, 2021.

[267] Nick Yee, Jeremy N Bailenson, Mark Urbanek, Francis Chang, and Dan Merget. The unbearable likeness of being digital: The persistence of nonverbal social norms in online virtual environments. *CyberPsychology & Behavior*, 10(1):115–121, 2007.

[268] Su-Ling Yeh, Timothy Joseph Lane, An-Yi Chang, and Sung-En Chien. Switching to the

rubber hand. *Frontiers in psychology*, 8:2172, 2017.

[269] Jabar Yousif. Social and telepresence robots a future of teaching. *Artificial Intelligence & Robotics Development Journal*, 1(1):58–65, 2021.

[270] Ye Yuan and Anthony Steed. Is the rubber hand illusion induced by immersive virtual reality? In *2010 IEEE Virtual Reality Conference (VR)*, pages 95–102. https://doi.org/10.1109/VR.2010.5444807. IEEE, 2010.

[271] Daniel Zeller, Karl J Friston, and Joseph Classen. Dynamic causal modeling of touch-evoked potentials in the rubber hand illusion. *Neuroimage*, 138:266–273, 2016.

[272] Jing Zhang and Bernhard Hommel. Body ownership and response to threat. *Psychological research*, 80(6):1020–1029. https://doi.org/10.1007/s00426–015–0698–1, 2016.

[273] Wen-Hong Zhu and Septimiu E Salcudean. Stability guaranteed teleoperation: an adaptive motion/force control approach. *IEEE transactions on automatic control*, 45(11):1951–1969, 2000.