



**HAL**  
open science

# **”Did she just tap me?”: Qualifying Multisensory Feedback for Social Touch during Human-Agent Interaction in Virtual Reality**

Grégoire Richard, Fabien Boucaud, Catherine Pelachaud, Indira Thouvenin

## **► To cite this version:**

Grégoire Richard, Fabien Boucaud, Catherine Pelachaud, Indira Thouvenin. ”Did she just tap me?”: Qualifying Multisensory Feedback for Social Touch during Human-Agent Interaction in Virtual Reality. *Frontiers in Virtual Reality*, 2025, 6, <10.3389/frvir.2025.1718198>. <hal-05386969>

**HAL Id: hal-05386969**

**<https://hal.science/hal-05386969v1>**

Submitted on 28 Nov 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

# "Did she just tap me?": Qualifying Multisensory Feedback for Social Touch during Human-Agent Interaction in Virtual Reality

Grégoire Richard<sup>1,\*</sup>, Fabien Boucaud<sup>2</sup>, Catherine Pelachaud<sup>2</sup> and Indira Thouvenin<sup>1</sup>

<sup>1</sup>CNRS, UMR 7253 Heudiasyc, Université de Technologie Compiègne, Compiègne, France

<sup>2</sup>CNRS, UMR 7222 ISIR, Sorbonne Université, Paris, France

Correspondence\*:

Grégoire Richard

gregoire.richard@hds.utc.fr

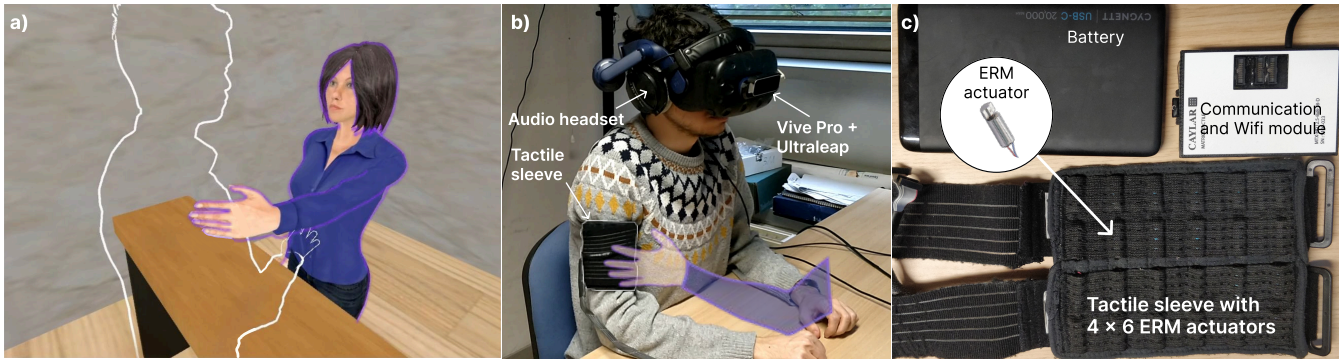
## 2 ABSTRACT

3 Social touch is one of the many modalities of social communication. It can take many forms (from  
4 handshakes to hugs) and can convey a wide variety of emotions and intentions. Social touch is  
5 also multimodal: it is not only comprised of haptic feedback (both tactile and kinesthetic), but  
6 also visual feedback (gestures) and even audio feedback (sound of the hand movement on the  
7 body). Virtual agents (VA) can perceive and interact with users by making use of multimodality  
8 to express attitudes and complex emotions. There are still few studies that have investigated  
9 how to integrate touch into VAs' social abilities, despite a growing interest in haptic interactions  
10 within immersive virtual environments (IVE). While prior work has examined haptic feedback,  
11 auditory substitution, or agent-based touch in isolation, no study has systematically disentangled  
12 the respective and combined contributions of visual, auditory, and tactile cues to the perception of  
13 being socially touched by a virtual agent. To address this gap, we conducted three experiments  
14 that progressively isolate and combine modalities, revealing how each shapes touch recognition,  
15 believability, and agency attribution. Taken together, our results show that multisensory feedback  
16 improves experience of social touch, which is in line with previous research in IVEs. They also  
17 show that visual feedback has the potential to guide the recognition of the stimulus the most but  
18 both audio and tactile feedback further help disambiguate the recognition of a touch gesture in  
19 particular cases. The results also show that both level of animation and interpersonal distance  
20 are essential to how much the VA is felt as a social agent when initiating touch. While visual  
21 feedback and tactile feedback are the main contributors to participants feeling touched by the VA,  
22 audio feedback also has a significant impact.

23 **Keywords:** virtual reality, haptic feedback, social touch, human-agent interaction

## 1 INTRODUCTION

24 Social touch is a central modality in human communication. It plays a key role in regulating emotions,  
25 signaling intentions, and shaping interpersonal bonds from early development to adulthood (Montagu,



**Figure 1.** Experimental setup for the experiments: a) In-game view of the experiment, with the virtual agent (highlighted in purple here) touching the participant on the upper arm of their avatar. (b) Illustration of the experimental setup with the vibrotactile sleeve (as shown here on a member of the research team for illustration purposes). c) The haptic sleeve.

26 1984) (Hertenstein et al., 2009) (Hauser et al., 2019). Touch can fulfill multiple functions: ritualistic (e.g., a  
 27 handshake), functional (e.g., capturing attention), or affective (e.g., comforting through a caress) (Jones  
 28 and Yarbrough, 1985) (Héron et al., 2024). The ability of touch to convey meaning and emotion is thus  
 29 integral to how people establish and maintain social relationships.

30 With the growing sophistication of embodied virtual agents (VAs), integrating social touch into  
 31 human-agent interactions has become an important research frontier. Equipping VAs with the ability to  
 32 perform touch gestures could enhance their social and emotional credibility, increasing trust, engagement,  
 33 and perceived empathy. However, simulating social touch in immersive virtual environments (IVEs)  
 34 remains a challenge: while haptic technologies such as vibrotactile or force-feedback devices can provide  
 35 sensations on the skin, these rarely capture the richness of human touch in terms of timing, texture, or  
 36 intent. Moreover, touch is inherently multimodal, combining physical sensations with visual and auditory  
 37 cues. Leveraging this multimodality through carefully designed combinations of visual animation, tactile  
 38 stimulation, and sound may offer a way to create more convincing and believable social touch interactions.

39 In this work, we investigate how different sensory modalities and their combinations contribute to the  
 40 perception of social touch initiated by a VA in VR. Prior research has explored sensory substitution  
 41 for touch (de Lagarde et al., 2023), social touch in human-robot interaction (Teyssier et al., 2020), or  
 42 human-agent touch in non-immersive settings (Huisman et al., 2016), but the systematic evaluation of  
 43 visual, audio, and tactile cues in immersive VR remains largely unexplored. To our knowledge, this study  
 44 provides the first controlled comparison of these modalities (alone, matched, and mismatched) within an  
 45 immersive setup. Across three experiments, we assess not only whether participants feel touched, but also  
 46 how they recognize the gesture and attribute agency to the VA. This exploratory approach clarifies the  
 47 conditions under which multimodal feedback enhances the subjective experience of being touched in VR,  
 48 and offers a foundation for designing richer and more credible virtual social touch interactions.

## 2 RELATED WORK

### 49 2.1 Social and affective touch

50 Social touch can be understood as signals used to communicate between a sender and a receiver. The sender  
 51 encodes an emotion or an intention within a touch ("*I am going to show that I care about you*"), and the  
 52 receiver performs a sensory judgment based on the information they perceived through their senses ("*The*  
 53 *sender wanted to show their care for me*"). Following this notion, McIntyre *et al.* recorded different touch

54 gestures by varying several parameters: types of motions, points of contact, velocity, duration, and emotion  
55 associated, as perceived by the receiver (McIntyre et al., 2022). This allowed them to build a taxonomy of  
56 touch gestures and their perceived meanings and they confirmed that human participants can recognize  
57 both those touch gestures and the socio-emotional intentions associated, including *attention grabbing*,  
58 *love* and *happiness*. Recent works from the neuro-physiological field on the CT-afferents (a specific class  
59 of unmyelinated cutaneous mechanoreceptors (Schirmer et al., 2023a)) have demonstrated that certain  
60 stroking speeds can maximize the pleasantness of affective touch (Schirmer et al., 2023b), and can induce  
61 psycho-physiological regulation responses (Püschel et al., 2022). However, the acceptability, pleasantness,  
62 and meaning of human-human social touch can be influenced by individual preferences, the relationship  
63 between the toucher and the touchee (Suvilehto et al., 2015), and socio-cultural context (Van Erp and  
64 Toet, 2015). Thus, as with all the other social abilities of humans, social touch on its own is an ambiguous  
65 communication modality. This can lead to misunderstandings or even feelings of invasion, as touch is also  
66 an intimate modality, which involves people’s bodies directly.

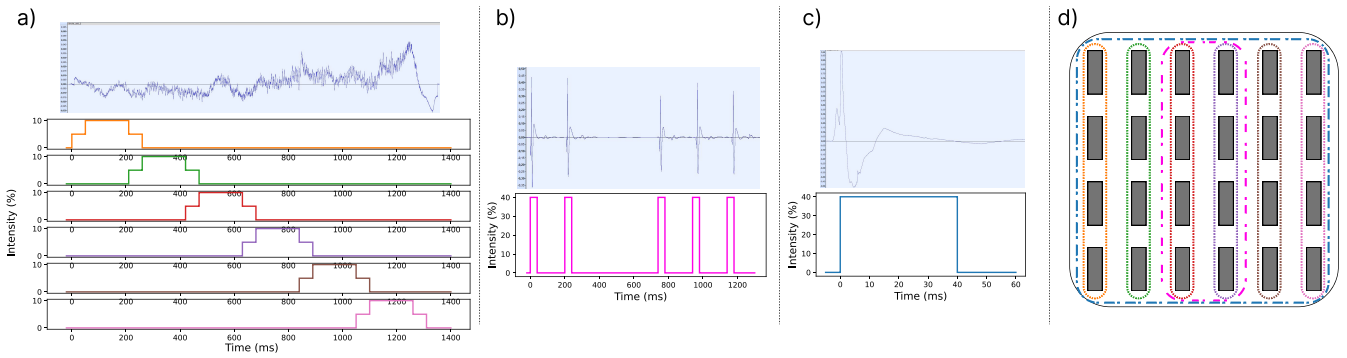
## 67 **2.2 Virtual social touch and virtual agents**

68 Simulating touch *via* technology (also called virtual touch in the context of IVEs) for social interactions  
69 usually relies on some form of haptic feedback, typically generated through haptic devices (Huisman  
70 et al., 2013b) (Huisman et al., 2016) (Ju et al., 2021). Alternatively, several studies have examined  
71 zoomorphic (Geva et al., 2020) and/or social (Song and Yamada, 2017) robots, which are tangible and  
72 physically embodied. Touching social robots can lead to strong physiological responses (Geva et al.,  
73 2020), and while some evidence suggests that people prefer touching a robot rather than being touched by  
74 it (Hirano et al., 2018), other studies show that robot-initiated touch can approximate the pleasantness of  
75 human touch when performed at the right velocity (Willemse et al., 2016).

76 The design space for simulating social touch is immense: parameters such as spatiotemporal profile,  
77 force, texture, or temperature can all be manipulated, yet no single technology currently reproduces the  
78 full richness of human touch (Teyssier et al., 2020). Vibrotactile feedback, despite its “unnatural” feel, has  
79 become the most common technology because it is affordable, customizable, and capable of producing  
80 tactile illusions such as caresses (Huisman et al., 2013b) (Huisman et al., 2016) (Ju et al., 2021). Studies  
81 using vibrotactile stroking have shown that pleasantness peaks at speeds between 1–10 cm/s, in line with  
82 CT-afferent activation in human-human touch (Huisman et al., 2016).

83 Huisman *et al.* (Huisman et al., 2013a) introduced one of the first systems enabling a VA to touch users  
84 via a vibrotactile sleeve composed of eccentric rotating mass motors (ERMs). Subsequent work has refined  
85 this approach: Ju *et al.* (Ju et al., 2021) designed vibrotactile patterns with the TECHTILE toolkit to convey  
86 emotions; Boucaud *et al.* (Boucaud et al., 2021) (Boucaud et al., 2023) integrated vibrotactile cues with  
87 a VA in immersive VR; and other studies have examined human responses to VA touch in controlled  
88 settings (Hoppe et al., 2020). A common thread is that vibrotactile touch can replicate some benefits of  
89 human touch, but its relative importance compared to visual or auditory cues remains unclear.

90 Two recent surveys provide comprehensive overviews of this growing area. Jacucci *et al.* (Jacucci  
91 et al., 2024) reviewed a decade of work on social haptics in VR, identifying three main research  
92 directions: how technologically mediated touch shapes behavior, how it conveys emotion, and how  
93 it can serve as a notification channel. Their synthesis also organized prior work by device type and  
94 study context, underscoring the need for in-the-wild validation and for tools that support richer design  
95 exploration. Olugbade *et al.* (Olugbade et al., 2023) examined affective touch across HCI, human-robot  
96 interaction, and virtual human communication, covering applications, datasets, recognition methods, and



**Figure 2.** Sound spectrums and profiles of tactile pattern as a signal and on the actuators' matrix for a) caress gesture; b) tap gesture; c) hit gesture. d) the actuators' matrix displays the actuators involved for each tactile pattern.

97 ethical considerations. Together, these reviews offer a consolidated view of current technologies and  
 98 methodological approaches, and we refer readers to them for broader coverage of the field.

99 Recent experimental studies reflect this momentum toward increasingly complex forms of mediated  
 100 social touch. Kim *et al.* (Kim and Choi, 2024) used a multimodal haptic glove (combining force, tactile,  
 101 and thermal cues) for user-initiated touch, showing that passive haptics (e.g., touching a silicone limb)  
 102 can outperform active devices in supporting embodiment and copresence. Sun *et al.* (Sun et al., 2024)  
 103 similarly demonstrated that multisensory richness matters: when an experimenter delivered combined  
 104 visual, auditory, and tactile cues, participants reported stronger effects than with unimodal stimulation.

105 As social touch is inherently multimodal, a natural research question emerges: does supplementing  
 106 vibrotactile feedback with additional cues (e.g., visual and auditory) enhance the realism and effectiveness  
 107 of simulated touch in IVEs?

### 108 2.3 Supplementing touch with other sensory cues

109 A signal can substitute another signal (e.g. a gesture can indicate a referenced object without having to  
 110 mention it verbally) or supplement it to disambiguate and/or complement the conveyed meaning (e.g. saying  
 111 'the red box' while pointing at a set of boxes (Lenay and Tixier, 2018)). De Lagarde *et al.* (de Lagarde  
 112 et al., 2023) demonstrated that, given enough context (participants being told that the sounds are produced  
 113 in human-human interaction), audio stimuli can effectively be recognized as social touch gestures and  
 114 associated with socio-emotional intentions, which is consistent with haptics' results in the social touch  
 115 literature. Their study focused on four gesture types: tapping, hitting, caressing (stroking), and rubbing.  
 116 The stimuli were created by recording vibrations produced by a touch performed directly on the skin, using  
 117 a violin microphone attached to a human arm. These findings, along with previous research on touch and  
 118 visual perception (Ravaja et al., 2017), highlight how closely interwoven the modalities of social touch are,  
 119 reinforcing the importance of studying each modality's contribution to eliciting a feeling of being socially  
 120 touched by a VA.

## 3 RESEARCH QUESTIONS

121 The overarching objective of this study is to disentangle the respective contributions of *visual*, *audio*, and  
 122 *tactile* feedback in shaping the perception of social touch gestures performed by a VA in immersive VR.  
 123 Specifically, we address both the *sensory* and the *social* dimensions of touch: not only whether participants  
 124 feel a sensation, but also whether they interpret it as a socially meaningful action initiated by the agent.

125 Our research is guided by the following questions:

126 •**RQ1:** How does each sensory modality (visual, audio, tactile) influence the recognition of specific social  
127 touch gestures?

128 •**RQ2:** To what extent does each modality contribute to the subjective feeling of being socially touched by  
129 a virtual agent?

130 •**RQ3:** Does the attribution of *agency* to the VA, that is, perceiving the VA as the intentional source of the  
131 touch, depend on the available sensory modalities?

132 To answer these questions, we conducted three complementary experiments. The first two focus on  
133 RQ1 and RQ2 by systematically manipulating modality combinations, while the third expands these  
134 investigations to RQ3 by introducing variations in interpersonal distance and non-verbal behavior, thus  
135 probing how sensory and social cues jointly affect the perception of touch.

## 4 METHODOLOGY

136 The following section describes the rationale and methodological choices guiding the design of our  
137 stimuli across experiments. To investigate our research questions, we designed three experiments aimed  
138 at evaluating how different combinations of the three sensory modalities affect the perception of social  
139 touch in interactions with a VA. The experimental protocol was developed with two primary goals: (1)  
140 to isolate the contribution of each sensory modality to the perception of being touched and (2) to assess  
141 the impact of multimodal combinations on the realism and believability of the touch experience. This  
142 study was approved by our local IRB (INSEAD IRB approval n°2024-46). For each experiment, a different  
143 sample of participants was recruited, and there was no overlap of participants between experiments. They  
144 were all recruited in France, but we did not enforce a strict cultural background screening.

145 Our design leverages a controlled virtual environment to simulate touch gestures performed by a VA and  
146 incorporates precise synchronization of sensory feedback across modalities. Participants interact with a  
147 female-looking VA, who performs social touch gestures validated in prior studies (Boucaud et al., 2019),  
148 namely tapping, hitting, and caressing. These gestures were chosen for their relevance to human-human  
149 social touch and adapted to match the capabilities of the haptic and audio systems employed in the study  
150 described further down in this section.

151 The protocol not only evaluates the effectiveness of individual modalities by integrating measures of  
152 gesture recognition and sensory perception, but also explores the influence of their interaction on the  
153 subjective experience of being touched. This methodology ensures a systematic and replicable approach to  
154 exploring the multisensory nature of social touch in virtual environments.

### 155 4.1 Methodology rationale

156 Our three experiments were designed to systematically explore the role of sensory matching through  
157 multimodal combinations in shaping the perception of social touch. Simulating social touch interactions  
158 involves navigating an expansive design space, as noted by Teyssier *et al.* (Teyssier et al., 2020), who  
159 demonstrated that device-initiated touch can convey diverse emotions depending on subtle variations in  
160 sensory parameters. Their work underscores the complexity of simulating touch, where even minor changes  
161 in timing, intensity, or modality alignment can significantly alter user perception. Similarly, Deli *et al.* (Deli  
162 et al., 2016) tackled this complexity in the domain of facial expressions by combining multiple parameters  
163 for emotion categorization. Their findings showed that isolating and comparing parameter effects leads to a  
164 deeper understanding of which elements are most critical in creating believable and emotionally resonant  
165 facial expression for virtual agents.

166 In our study, we extend this principle to social touch by isolating and combining sensory modalities to  
167 determine their individual and synergistic contributions to touch recognition and perception. We thus say  
168 that the sensory modalities are **matched** when each modality’s stimulus presented to the user is intended to  
169 represent the same gesture (e.g. the audio **and** the tactile **and** the visual animation of a tap gesture). On the  
170 other hand, we say that the sensory modalities are **mismatched** when the stimuli presented are not intended  
171 to represent the same gesture (e.g. the audio of a tap gesture combined with the vibrotactile pattern designed  
172 to simulate a hit, with the visual animation of a caress). Finally, we say that the modalities are partially  
173 matched when the three sensory modalities are present but only two of them are matched and the third one  
174 features a mismatched stimulus. Each trial could thus be a complete match, a complete mismatch or a partial  
175 match. Including mismatched trials enables us to investigate how mismatched stimuli disrupt perception,  
176 an approach supported by Teyssier *et al.* ’s observation that multisensory congruence plays a pivotal role in  
177 establishing emotional coherence in touch. By studying both matched and mismatched feedback across  
178 modalities, we aim to identify the conditions necessary for achieving realistic and compelling social touch  
179 interactions with virtual agents.

## 180 4.2 Vibrotactile device

181 The device used in all three experiments, an MTX-Prototype designed by *CAYLAR SAS*, is made up of  
182 a sleeve consisting of  $4 \times 6$  ERMs, a microcontroller, and a Wi-Fi card. Each actuator is cylindrical and  
183 has a size of  $7.7 \text{ mm} \times 22 \text{ mm}$ . The specifications of the actuators are similar to the one found at *Precision*  
184 *MicroDrive*<sup>TM</sup> (Microdrive<sup>TM</sup>, 2023). Actuators’ output can be controlled through voltage, and there exists  
185 a quasi-linear relationship between output and frequency, with 100 Hz at 10% to a nominal frequency of  
186 300 Hz at 100% of the maximal output. The sleeve was always worn directly on the participants’ upper  
187 arm using Velcro straps, with clothing adjusted or rolled up to avoid additional layers between the actuators  
188 and the skin. The teaser image in Figure 1 illustrates the setup, although for illustration purposes the device  
189 is shown placed over clothing. The total size of the matrix is  $110 \text{ mm} \times 170 \text{ mm} \times 15 \text{ mm}$  (see Figure 1).

## 190 4.3 Designing the visual and tactile stimuli from audio recordings

191 Few studies have validated the combination of tactile and auditory stimuli for social touch (Huisman et al.,  
192 2013b) (Teyssier et al., 2020). For our study, we relied on the dataset by De Lagarde *et al.* (de Lagarde  
193 et al., 2023), who showed that recordings of human-human social touch (tap, caress, rub, hit) could be  
194 recognized above chance level when presented unimodally. To maintain experimental feasibility while still  
195 covering a range of social intentions, we restricted our set to three gestures: *Tap*, *Hit*, and *Caress*, selected  
196 for their higher recognition rates in the original validation. This choice reflects the exploratory nature of  
197 the present study, which cannot exhaustively cover the large design space of social touch.

### 198 4.3.1 Tactile stimuli design.

199 The vibrotactile patterns were implemented on our ERM-based sleeve (see Section 4.2). Although ERMs  
200 lack the precision of voice-coils or grounded haptic devices, they remain practical, robust, and easy to  
201 integrate into immersive VR setups, making them suitable for our exploratory study (Remache-Vinueza  
202 et al., 2021). To derive tactile patterns, we analyzed the recordings in terms of *gesture duration*, *intensity*,  
203 *and rhythm*. Rhythm was prioritized as the most critical parameter for cross-modal alignment (Jack et al.,  
204 2015), ensuring that the onset and offset of vibrations corresponded to the temporal structure of the audio.  
205 In addition to directly audio-aligned patterns, we systematically varied stimulus *duration* (five levels) and  
206 *intensity* (three levels), yielding 15 candidate patterns per gesture.

### 207 4.3.2 Pilot study.

208 A pilot study with 12 participants (6 males, 6 females,  $M = 31$ ,  $SD = 9.1$ ) was conducted to identify the  
209 most representative tactile patterns. Participants experienced all 15 variations for each gesture and selected  
210 the one they considered most appropriate. They then compared this chosen pattern with the audio-aligned  
211 pattern, rating both combinations on a 7-point Likert scale regarding:

- 212 •Synchronization between audio and vibrations,
- 213 •Accuracy in reproducing the intended gesture,
- 214 •Perceptual coherence of the combined stimulus.

215 No significant differences emerged between the participant-chosen patterns and the audio-aligned ones.  
216 We therefore retained the *audio-aligned tactile patterns* for the main study, ensuring consistency across  
217 participants by presenting the same tactile patterns irrespective of individual preferences identified in the  
218 pilot study.

### 219 4.3.3 Final tactile stimuli.

220 The final patterns, summarized in Figure 2, were:

- 221 •**Hit**: a single 40 ms square-wave pulse across all actuators at 70% intensity.
- 222 •**Tap**: five 40 ms pulses at 40% intensity on 8 central actuators, arranged as two quick taps, a 500 ms  
223 pause, and three quick taps (160 ms gaps).
- 224 •**Caress**: the actuator layout of the sleeve (four rows of motors, see Fig. 1) did not allow us to reproduce  
225 a continuous linear stroke *along* the length of the arm. During early design exploration, we therefore  
226 compared trajectories that progressed longitudinally (proximal–distal) with trajectories that progressed  
227 *around* the arm, i.e., a smooth, linear sweep of activations along the arm’s circumference from the front  
228 toward the back. Participants did not report systematic differences in how these variants were interpreted  
229 as a caress, while they perceived motion better with the circumferential pattern. We consequently  
230 implemented the around-the-arm sweep, which provided the smoothest and most continuous sensation  
231 given the hardware constraints. The final pattern lasted 1300 ms at 10% intensity (approx. 13 cm/s  
232 stroking speed).

### 233 4.3.4 VA gesture animations.

234 The visual animations were adapted from the GRETA platform (Rosis et al., 2003; Boucaud et al., 2019),  
235 which provides validated social touch gestures performed by an embodied conversational agent. We  
236 preserved the core kinematics of these animations so that the three gestures remained comparable to prior  
237 work, while adjusting their timing and endpoint to align with our tactile and auditory patterns. In particular,  
238 Tap and Caress share a similar reaching phase of the forearm toward the participant’s upper arm, after  
239 which their trajectories diverge: Tap ends with a brief tapping motion, whereas Caress involves a slower,  
240 sliding hand movement. By contrast, Hit includes a more pronounced preparatory swing and a sharper  
241 impact, making it visually distinctive from the outset.

242 Because our goal was to investigate how modalities combine rather than to maximize photorealistic  
243 realism, we prioritized *plausible and internally consistent* movements over ultra–high-fidelity, motion-  
244 captured animation. Using GRETA-based gestures ensured continuity with validated social-touch datasets  
245 and provided tight control over rhythm and amplitude across conditions. As we discuss in Sections 5.7  
246 and 6.5, the main ambiguity between Tap and Caress emerges from the intrinsic subtlety of their endpoint  
247 motions and their shared reaching phase; structural properties that would persist even with motion-captured

248 animations. This ambiguity is further amplified by the limited field of view of the headset, which can  
249 partially occlude the fine hand movements that differentiate these two gestures.

250 It is worth noting that, in immersive VR, animation *fidelity* does not necessarily entail photorealistic  
251 motion capture. Prior research on presence and virtual humans shows that users' responses depend more  
252 on the *plausibility* and *coherence* of an agent's behavior than on graphical realism (Slater, 2009, 2018;  
253 Pan and Hamilton, 2018). When the kinematic structure of an action is preserved, stylized or simplified  
254 characters can convey emotional and social meaning as effectively as highly realistic ones (McDonnell  
255 et al., 2008, 2012). Moreover, extremely high visual realism may expose small animation imperfections  
256 and reduce perceived believability (McDonnell et al., 2012). In this context, our use of GRETA-based  
257 gestures reflects a deliberate methodological choice to prioritize controlled and plausible social motion  
258 which is critical for multisensory experiments, over maximizing visual realism for its own sake.

## 5 EXPERIMENT 1

259 Onward, we call *gestures* the emblematic types of social touches featuring at least one sensory feedback  
260 among the three studied here. For this first experiment, we explore how the coupling of visual animations  
261 of the gesture performed by a VA and either an audio or a tactile stimulus can foster the feeling of being  
262 socially touched by naive participants and how the recognition of the touch gestures is impacted by the  
263 different combinations.

### 264 5.1 Apparatus for experiment 1

265 Using *Unity 3D 2020.3.48f1*, we set up a basic VR environment similar to the literature (Boucaud et al.,  
266 2023), using a *Vive Pro 2* headset (HMD) combined with a *Leap Motion Controller* for tracking, respectively,  
267 the user's head and hands, allowing for basic presence and sense of embodiment towards a given avatar  
268 (see Figure 1.a). This avatar consists of a white outline allowing users to see their virtual hands (detected  
269 by the Leap Motion Controller) and their virtual arms (controlled through inverse kinematics). The Vive  
270 Pro HMD provides a field of view (FOV) of approximately  $116^\circ$  and a refresh rate of 90 Hz. This relatively  
271 wide FOV ensured good visibility of the VA's gestures, though some peripheral movements may still have  
272 extended beyond the participant's immediate visual field. As discussed in Section 4.3, we used the GRETA  
273 platform to control the visual behaviour of our VA (the tap, hit and caress visual animations). The activation  
274 of the audio and of the tactile feedback are synchronized with the moment when the hand of the agent  
275 reaches the human user for each gesture. The virtual agent kept a neutral facial expression throughout the  
276 experiment. Participants were seated in front of a desk in the real world with a matching virtual desk in the  
277 IVE. In front of them, behind the desk, they could see the VA positioned at arm's reach (see Figure 1.a). As  
278 touch from a woman is usually more acceptable for both female and male touchees (Suvilehto et al., 2015),  
279 we used a VA with a female gender. The tactile feedback was rendered through the tactile sleeve described  
280 in Section 4.2. To play the audio, and prevent participants from hearing the vibrations, participants were  
281 equipped with noise cancellation *DT770 Pro* headphones from *Beyerdynamic*. All the questionnaires could  
282 be answered directly in the IVE through direct manual interaction with virtual panels. Examples of trials  
283 and of the virtual environment can be seen both in Figure 1.a and in the video submitted as supplementary  
284 material.

### 285 5.2 Hypotheses

286 We have three hypotheses for the first experiment:

287 •**H1**: adding either audio or tactile feedback will increase the feeling of being touched by the VA compared  
288 to the condition with visual feedback only (Melo et al., 2020) (**RQ2**).

289 •H2: mismatched multisensory feedback will lead to a decreased perception of being socially  
290 touched (Jeunet et al., 2018) (Richard et al., 2021) (RQ2).

291 •H3: visual animations will be the strongest factor in the recognition rate of the social touch gestures  
292 (RQ1).

### 293 5.3 Measures

294 The central dependent measure in Experiment 1 was participants' subjective sense of being touched by the  
295 virtual agent (VA). We constructed a short questionnaire (Table 1) combining four items that targeted both  
296 the perception of a physical stimulus and the attribution of that stimulus to the VA. Items included, for  
297 example, whether the participant felt physically touched and whether the touch was perceived as caused by  
298 the agent.

299 While several validated instruments exist for evaluating user experience with virtual agents (e.g., the  
300 Artificial Social Agent Questionnaire, ASAQ, for believability and social presence - published after we ran  
301 our experiments (Fitriani et al., 2025)) or attitudes toward social touch in human-human interaction (e.g.,  
302 the Social Touch Questionnaire, STQ (Lapp and Croy, 2021)), none to our knowledge assess the combined  
303 experience of *multimodal* (visual-auditory-haptic) social touch or the recognition of touch gestures in  
304 immersive environments. We therefore designed exploratory composite indices tailored to the present  
305 study's aims. The resulting FEELINGTOUCHED index drew on wording and conceptual inspiration from  
306 embodiment and social-presence research (e.g., Jeunet et al., 2018; Huisman et al., 2016; Kilteni et al.,  
307 2012) and achieved excellent internal reliability ( $\alpha = 0.91$ , 95% CI [0.90, 0.92]).

308 Responses were given on 7-point Likert scales and aggregated into a single score representing participants'  
309 overall feeling of being touched by the agent. Although reliability was high, the four items encompass  
310 partially distinct constructs (e.g., tactile sensation, social attribution, spatial presence). The aggregated score  
311 should therefore be interpreted as an *exploratory composite index* rather than a validated unidimensional  
312 scale. This limitation motivated refinements introduced in Experiment 3, where the constructs of perceived  
313 TOUCHAGENCY and TOUCHBELIEVABILITY were measured separately.

314 In addition, gesture recognition was assessed by asking participants to identify which gesture (*hit*, *tap*,  
315 or *caress*) they believed the VA had performed, chosen from a predefined list. This categorical outcome  
316 allowed for subsequent analyses of recognition accuracy across modality conditions.

### 317 5.4 Experimental design and conditions

318 To keep the experiment at a reasonable length for participants, we settled for three experimental conditions  
319 for this first experiment: only visual feedback (agent animations), which serves as a control condition  
320 (VISUONLY); visual and haptic (vibrotactile) feedback (VISUHAPTIC); visual and audio feedback  
321 (VISUAUDIO). We chose to always have visual feedback present as our research is aimed at interactions  
322 with VAs that always feature some virtual body, regardless of other modalities. The experiment followed  
323 a within-subject design. For our three kinds of gestures (tap, hit, and caress) we used the audio and  
324 vibrotactile stimuli and the animations described in Section 4.3 to produce all possible combinations for  
325 each condition, both **matched** and **mismatched**. There are thus 3 trials for the VISUONLY condition and  
326  $3 \times 3$  trials for each of the VISUHAPTIC and VISUAUDIO conditions, for a total of 21 trials.

### 327 5.5 Procedure

328 After being seated in front of a plain office desk, participants read information about the experiment through  
329 an information notice and signed an informed consent form. Participants were instructed that the virtual

**Table 1.** Questionnaire used for each trial.

ID	Questions	Format
<b>Recognition of Gesture</b>		
Rec-1	What kind of touch did you receive?	Single-choice (Hit/Tap/Caress)
<b>Perception of the agent’s touch</b>		
Per-1	I felt the agent touch me.	7-pt Likert
Per-2	It felt like I was in the same place as the agent.	7-pt Likert
Per-3	The agent could reach me.	7-pt Likert
Per-4	I did not feel a touch.	7-pt Likert

330 agent would perform a series of touch gestures directed toward their avatar. They were told that the gestures  
331 might vary in their sensory characteristics (visual, auditory, tactile) and that their task was to (i) report how  
332 touched they felt by the agent and (ii) identify which gesture the agent performed when applicable. No  
333 information was given about congruence or mismatch between modalities. The experimenter then helped  
334 them put the vibrotactile device on their upper arm and adjust the head-mounted display (HMD) and the  
335 headphones. Participants were immersed in the IVE described in Section 5.1 and familiarized themselves  
336 with how to answer the questionnaires with their virtual hands using the Leap Motion Controller through a  
337 short tutorial. Once ready, they pressed a button to start the trials, in which the agent would touch them on  
338 the arm visually and with either vibrotactile feedback or audio feedback depending on the trial condition.  
339 They answered the questionnaire shown in Table 1) for each trial right after experiencing a stimulus. This  
340 was repeated for each of the 21 trials. The presentation order of the trials was fully randomized.

## 341 5.6 Results

342 Twenty-five participants took part in Experiment 1 (16 male, 9 female;  $M = 25.5$ ,  $SD = 6.3$ ), recruited by  
343 word of mouth. All analyses described below (and additional checks) are available in the supplementary  
344 material. Shapiro–Wilk tests indicated violations of normality; therefore, only non-parametric analyses  
345 were conducted.

346 The experiment comprised 21 trials: 3 VISUONLY trials and 18 multisensory trials that combined vision  
347 with either haptics or audio in matched or mismatched form. For analysis, we grouped trials into five  
348 *matching groups*: VISUONLY, VISUAUDIOMATCH, VISUAUDIOMISMATCH, VISUHAPTICMATCH, and  
349 VISUHAPTICMISMATCH. Distributions per group are shown in Figure 3.

### 350 5.6.1 Perception of the agent’s touch.

351 We computed the aggregated FEELINGTOUCHED score by averaging items *Per-1*, *Per-2*, *Per-3*, and *Per-4*  
352 (reverse-scored). The scale showed excellent internal reliability (Cronbach’s  $\alpha = 0.91$ ). A Friedman test  
353 across matching groups was significant, and post-hoc Wilcoxon signed-rank tests with Holm correction  
354 revealed significant pairwise differences among groups. Detailed pairwise outcomes are annotated in  
355 Figure 3.

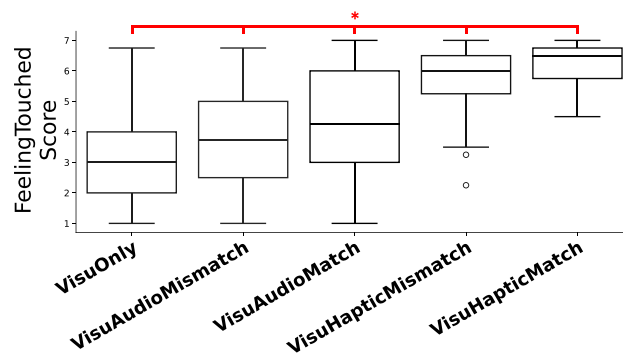
### 356 5.6.2 Recognition of the gesture.

357 Figure 4 reports recognition rates by matching group. Because the visual animation was present in all  
358 conditions, we recoded responses into a binary outcome indicating whether the selected gesture matched  
359 the intention of the *visual* animation (*yes/no*). We then fit a mixed-effects logistic regression with participant  
360 ID as a random effect (525 observations; 25 participants; variance = 0.026,  $s.d. = 0.161$ ) and matching  
361 group as a fixed effect (summary in Table 2). For example, relative to VISUONLY, VISUAUDIOMATCH  
362 significantly reduced the odds of an incorrect choice ( $\beta = -0.976$ ,  $SE = 0.462$ ,  $z = -2.11$ ,  $p = .035$ ).

**Table 2.** Selected results from fixed effects (condition groups) of mixed-effects logistic regression (full results available in supplementary materials). The outcome variable (whether the gesture chosen by the participants corresponds to the intended visual animation) is coded as Yes = 0 and No = 1.

	Estimates	Std. Err.	Z value	Pr(> z )
<b>Reference level: VisuOnly</b>				
(Intercept)	-1.1595	0.2736	-4.237	2.26e-05 ***
VisuAudioMatch	-0.9760	0.4623	-2.111	0.0347 *
VisuHapticMatch	-0.4112	0.4084	-1.007	0.3140
VisuAudioMismatch	1.1058	0.3173	3.485	0.0004 ***
VisuHapticMismatch	0.6382	0.3197	1.996	0.0458 *
<b>Reference level: VisuHapticMatch</b>				
(Intercept)	-1.5707	0.3083	-5.095	3.50e-07 ***
VisuOnly	0.4113	0.4085	1.007	0.3140
VisuAudioMismatch	1.5170	0.3476	4.364	1.27e-05 ***
VisuAudioMatch	-0.5648	0.4834	-1.168	0.2426
VisuHapticMismatch	1.0495	0.3497	3.001	0.0026 **
<b>Reference level: VisuAudioMismatch</b>				
(Intercept)	-0.0537	0.1671	-0.321	0.7478
VisuHapticMismatch	-0.4676	0.2358	-1.983	0.047387 *

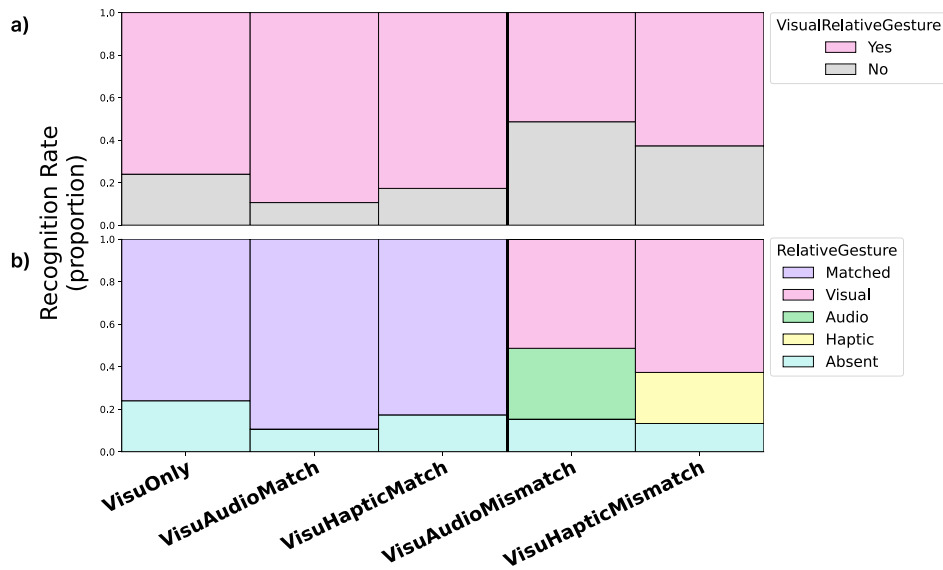
363 To examine which cue participants followed in *mismatched* trials, we ran chi-squared tests on the  
 364 distribution of chosen gestures within each mismatched group. In VISUAUDIOMISMATCH, choices differed  
 365 significantly across the VISUAL vs. AUDIO intended gestures ( $\chi^2(2, N = 150) = 5.74, p = 0.0166$ ). In  
 366 VISUHAPTICMISMATCH, choices differed significantly across the VISUAL vs. TACTILE intended gestures  
 367 ( $\chi^2(2, N = 150) = 25.9, p < 0.0001$ ). The breakdown of responses for mismatched trials is shown in  
 368 Figure 4.b.



**Figure 3.** Box plot of participants' average FEELINGTOUCHED scores (Likert scale, 1–7) across multisensory matching groups, as described in Section 6.4. The red asterisk indicates that all pairwise comparisons involving this group reached statistical significance. Within each modality combination, matched conditions elicited higher FEELINGTOUCHED ratings than mismatched ones.

## 369 5.7 Discussion

370 The results of Experiment 1 show that participants reported a stronger feeling of being touched by the VA  
 371 whenever additional sensory feedback was present, compared to visual feedback alone. This effect held even  
 372 for incongruent trials, confirming that multisensory stimulation enhances the experience of social touch.  
 373 Tactile feedback led to the highest ratings overall, and congruent conditions consistently outperformed  
 374 incongruent ones. These findings provide clear support for **H1** and **H2**, and are consistent with previous



**Figure 4.** Gesture recognition rates across multisensory matching conditions. (a) Proportion of trials in which participants correctly recognized the gesture relative to the visual animation (“yes” if their choice matched the intended animation, “no” otherwise). (b) Distribution of chosen gestures across matching conditions. For matched groups, responses are classified as either the *matched gesture* or *absent gesture* (i.e., a gesture not present in the stimuli, such as recognizing a tap when the actual feedback - audio and visual - was a caress). For mismatched groups, responses are attributed to the modality that drove the choice (visual, tactile, or audio) or to *absent*. Overall, recognition was higher in matched conditions than in mismatched ones.

375 work highlighting the importance of congruence for user experience (Jeunet et al., 2018) (Richard et al.,  
 376 2021).

377 Gesture recognition revealed a more nuanced pattern. The VISUAUDIOMATCH condition significantly  
 378 improved recognition compared to VISUONLY, while tactile feedback did not significantly alter recognition  
 379 rates. In incongruent trials, audio cues influenced participants’ answers more strongly than tactile ones.  
 380 Taken together, these findings suggest that audio feedback can be more effective than haptic feedback for  
 381 disambiguating gestures, even though haptics contributes more strongly to the subjective sensation of being  
 382 touched. This provides only partial support for **H3**.

383 Qualitative feedback provided additional context for these findings. Several participants described the  
 384 Hit gesture as particularly easy to recognize, largely due to its distinctive preparatory motion and sharp  
 385 impact. By contrast, Tap and Caress were sometimes reported as visually similar. Both gestures involved  
 386 the agent leaning forward and extending the forearm toward the participant’s upper arm, with differences  
 387 mainly in the final hand movement at the point of contact (a brief tap versus a slower caress). In the HMD,  
 388 participants often focused on the VA’s face, and the limited field of view meant that this endpoint motion  
 389 was not always fully visible in peripheral vision. As a result, the shared reaching phase may have made  
 390 Tap and Caress appear more alike, prompting participants to rely more heavily on additional sensory cues  
 391 for recognition. Importantly, this reduced recognizability of the Caress gesture was already present in  
 392 the visual-only trials, indicating that the ambiguity stems from the intrinsic subtlety of the gesture and  
 393 viewpoint constraints rather than from limitations of the tactile rendering.

394 In summary, Experiment 1 demonstrates that multimodal feedback significantly increases the subjective  
 395 feeling of being touched, with tactile feedback playing the strongest role. Audio feedback, however,

396 appeared more influential than haptic cues in guiding recognition. These findings suggest a functional  
397 dissociation between modalities: tactile feedback enhances the affective dimension of touch, while audio  
398 feedback supports gesture recognition. Importantly, the study did not examine how audio and haptic  
399 feedback might interact when combined. Addressing this gap motivated the design of Experiment 2, which  
400 directly investigates tri-modal (visual–audio–haptic) stimulation and explores whether adding the third  
401 modality produces additive or interaction effects beyond those observed for bimodal feedback.

## 6 EXPERIMENT 2

402 Experiment 2 was originally conceived as part of the same experimental study as Experiment 1. The initial  
403 design included every possible combination of visual, auditory, and haptic feedback (both matched and  
404 mismatched) within a single comprehensive framework. However, piloting revealed that completing all  
405 conditions in one session would lead to excessive duration and participant fatigue, potentially compromising  
406 data quality.

407 To preserve methodological rigor and participant engagement, we therefore divided the original design  
408 into two complementary experiments. **Experiment 1** focused on unimodal and bimodal combinations to  
409 isolate pairwise contributions, whereas **Experiment 2** extended this framework to full tri-modal integration.  
410 This division maintained the integrity of the overall research design while allowing independent analyses  
411 and feasible session lengths. The present experiment thus specifically investigates whether adding a third  
412 sensory channel yields additive or interactive effects on perceived realism and gesture recognition.

### 413 6.1 Measures

414 Measures were identical to those of Experiment 1 to allow direct comparison. The FEELINGTOUCHED  
415 questionnaire (Table 1) was again aggregated into a single exploratory index, and gesture recognition was  
416 assessed using the same forced-choice task.

### 417 6.2 Experimental design and conditions

418 The experimental setup is identical to the one from Section 5, in terms of the animations, sounds and tactile  
419 patterns used, as well as apparatus, measures and procedure. Only conditions differ.

420 Experiment 2 introduced a *tri-modal* configuration combining all three sensory feedback modalities  
421 (visual, auditory, and tactile) alongside a visual-only control condition. This design enabled us to examine  
422 how congruence across all sensory channels influences both the subjective feeling of being touched and  
423 gesture recognition.

424 Two high-level experimental conditions were tested:

425 •**Visual-only (VISUONLY)**: participants saw the gesture animation without any accompanying audio or  
426 tactile feedback (3 trials);

427 •**Visuo–Audio–Haptic (VISUAUDIOHAPTIC)**: participants received simultaneous visual, auditory, and  
428 tactile feedback, which could vary in congruence across modalities.

429 Within the VISUAUDIOHAPTIC condition, each modality could represent one of the three intended  
430 gestures (*Hit*, *Tap*, or *Caress*). All possible combinations were presented, resulting in matched, partially  
431 matched, and completely mismatched trials:

432 •**CompleteMatch** – all three modalities correspond to the same gesture (3 trials);

- 433 •**AudioHapticMatch**, **VisuHapticMatch**, and **VisuAudioMatch** – two modalities match and one differs  
434 (6 trials each);  
435 •**CompleteMismatch** – each modality represents a different gesture (6 trials).

436 In total, participants experienced 3 VISUONLY trials and  $3 \times 3 \times 3 = 27$  VISUAUDIOHAPTIC trials,  
437 yielding 30 trials per participant. The experiment followed a within-subjects design, so every participant  
438 experienced all conditions in randomized order.

### 439 6.3 Hypotheses

440 In this second experiment, a new level of mismatch is introduced as trials are made of three different  
441 sensory feedback. In consequence, our hypotheses are:

- 442 •**H3**: Completely matching trials will elicit a higher sense of being socially touched compared to partially  
443 matching trials, which will be better than completely mismatched trials (**RQ2**).  
444 •**H4**: For partially or completely mismatched trials, visual feedback will be predominant for gesture  
445 recognition (**RQ1**).  
446 •**H5**: When visual feedback is more ambiguous, audio feedback will be predominant over tactile feedback  
447 for gesture recognition (**RQ1**).

448 Hypothesis **H5** was formulated as a result of the first experiment, where audio feedback seemed to drive  
449 recognition better than tactile feedback.

### 450 6.4 Results

451 Twenty-four participants took part in Experiment 2 (10 male, 14 female;  $M = 26$ ,  $SD = 7.2$ ). As in  
452 Experiment 1, analyses were conducted based on sensory matching conditions (see Section 6.2 for grouping  
453 details). The aggregated FEELINGTOUCHED scores were computed in the same way as in Experiment 1  
454 (see Section 5.6), and group distributions are shown in Figure 5. Shapiro–Wilk tests indicated violations of  
455 normality; therefore, only non-parametric tests were applied.

#### 456 6.4.1 Perception of the agent’s touch.

457 A Friedman test on the FEELINGTOUCHED scores revealed a significant effect of sensory congruence  
458 ( $F = 33.3$ ,  $p < 0.001$ ). Post-hoc Wilcoxon signed-rank tests with Holm correction identified 11 significant  
459 pairwise differences. Specifically:

- 460 •The *CompleteMatch* condition differed significantly from all other groups.  
461 •The *VisuOnly* condition also differed significantly from all other groups.  
462 •The *CompleteMismatch* condition differed significantly from *AudioHapticMatch* ( $W = 48.0$ ,  $p = 0.007$ )  
463 and from *VisuHapticMatch* ( $W = 41.0$ ,  $p = 0.008$ ).

464 The full set of comparisons is reported in the supplementary material.

#### 465 6.4.2 Recognition of the gesture.

466 As in Experiment 1, responses were recoded as *yes/no* depending on whether participants’ choices matched  
467 the intended *visual* animation. This binary transformation enabled the use of a mixed-effects logistic  
468 regression (Table 3), with participant ID as a random effect (720 observations; 24 participants; variance  
469 = 0.1475,  $s.d. = 0.3841$ ) and matching group as the fixed effect. Results confirmed a significant influence  
470 of condition on gesture recognition rates.

**Table 3.** Mixed-effects logistic regression’s fixed effects (condition groups) selected results (full results available in supplementary materials). Outcome variable (whether the gesture chosen by the participants is the one intended to correspond to the visual animation) is coded as Yes = 0 and No = 1.

	Estimates	Std. Err.	Z value	Pr(> z )
<b>Reference level: VisuOnly</b>				
(Intercept)	-0.4080	0.2567	-1.589	0.1120
CompleteMatch	-1.3554	0.4114	-3.295	0.0009 ***
CompleteMismatch	0.7270	0.2989	2.432	0.0150 *
AudioHapticMatch	0.9064	0.3007	3.014	0.0025 **
VisuHapticMatch	-0.2772	0.3027	-0.916	0.3598
VisuAudioMatch	-0.3418	0.3038	-1.125	0.2606
<b>Reference level: CompleteMatch</b>				
(Intercept)	-1.7634	0.3407	-5.176	2.26e-07 ***
CompleteMismatch	2.0824	0.3738	5.571	2.53e-08 ***
AudioHapticMatch	2.2618	0.3754	6.026	1.69e-09 ***
VisuHapticMatch	1.0782	0.3759	2.868	0.0041 **
VisuAudioMatch	1.0137	0.3768	2.691	0.0071 **
<b>Reference level: VisuAudioMatch</b>				
(Intercept)	-0.74974	0.19721	-3.802	0.0001 ***
CompleteMismatch	1.06874	0.24976	4.279	1.88e-05 ***
AudioHapticMatch	1.24810	0.25198	4.953	7.30e-07 ***
VisuHapticMatch	0.06451	0.25394	0.254	0.7994

#### 471 6.4.3 Influence of individual modalities.

472 To further analyze mismatched trials, a chi-squared test was conducted on gesture distributions within the  
 473 COMPLETEMISMATCH condition. No significant differences were found across modalities ( $\chi^2(2, N =$   
 474 144) = 5.79,  $p = 0.055$ ). In contrast, when comparing mismatched response rates across the partially  
 475 matched groups (Figure 6.b), a significant effect was observed for AUDIOHAPTICMATCH ( $\chi^2(2, N =$   
 476 144) = 10.36,  $p = 0.0056$ ), suggesting a stronger role of audio–haptic congruence in guiding recognition.

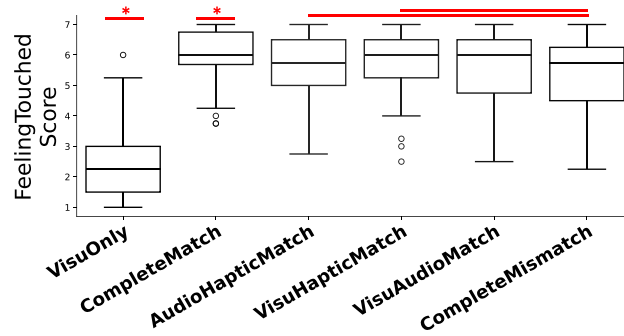
#### 477 6.4.4 Role of the hit animation.

478 Recognition rates across all 27 multisensory trials (excluding VISUONLY) are presented in Figure 7. The  
 479 *hit* animation consistently dominated recognition, overriding incongruent sensory cues in most cases. The  
 480 only exception occurred when a *tap* tactile pattern was combined with a non-hit audio cue, leading to  
 481 split recognition between tap and hit. After removing all *hit* trials, a chi-squared test showed that in the  
 482 COMPLETEMISMATCH condition, modalities were equally likely to guide responses ( $\chi^2(2, N = 93) =$   
 483 0.1875,  $p = 0.91$ ). Similarly, no significant differences were found across the three partial match groups  
 484 (AUDIOHAPTICMATCH, VISUHAPTICMATCH, VISUAUDIOMATCH;  $\chi^2(2, N = 76) = 0.421$ ,  $p = 0.81$ ).

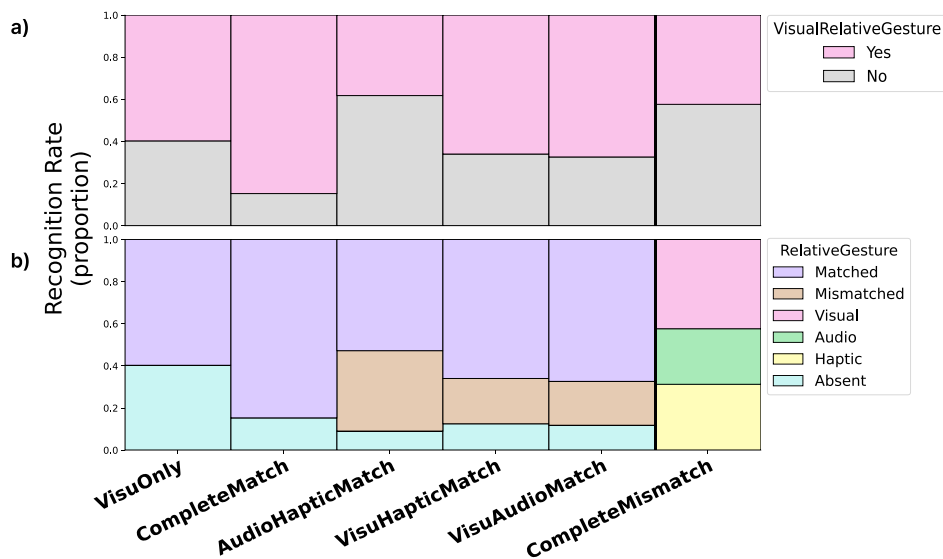
### 485 6.5 Discussion

486 Reflecting on the FEELINGTOUCHED scores, *CompleteMatch* trials elicited the strongest sense of  
 487 being socially touched, followed by *partially matching* conditions, and then *CompleteMismatch*. Even  
 488 mismatched trials, however, produced higher ratings than *VisuOnly*. These results support **H3**, align with the  
 489 findings of Experiment 1, and reinforce prior evidence that multisensory integration enhances immersion in  
 490 IVEs (Sallnäs, 2004) (Kilteni et al., 2012) (Melo et al., 2020), with congruence playing a key role (Richard  
 491 et al., 2021) (Jeunet et al., 2018).

492 Turning to **H4**, recognition results were more nuanced. In *CompleteMismatch* conditions, participants  
 493 distributed their responses equally across visual, audio, and haptic cues, showing no dominant modality.



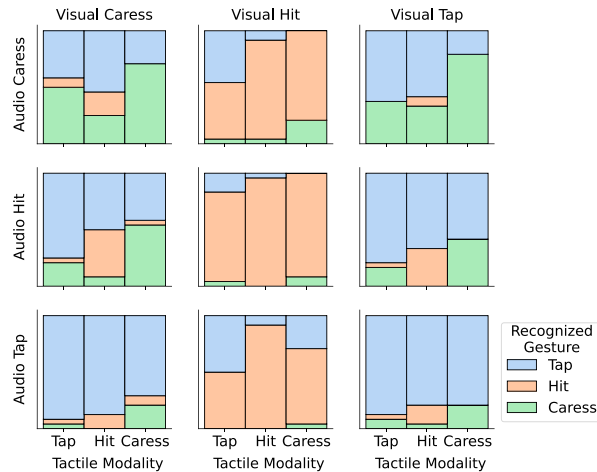
**Figure 5.** Box plot of participants’ average FEELINGTOUCHED scores (Likert scale, 1–7) across multisensory matching groups, as described in Section 6.4. The red asterisk indicates that all pairwise comparisons involving this group reached statistical significance. Overall, matched groups elicited higher FEELINGTOUCHED ratings than mismatched ones.



**Figure 6.** Recognition rates across multisensory matching groups. (a) Proportion of correct recognitions relative to the visual animation (y-axis = proportion of trials; “yes” = chosen gesture matched the animation, “no” = otherwise). (b) Distribution of recognized gestures across matching groups. For matched groups, responses are classified as either the intended gesture or a *absent gesture* (i.e. a gesture not present in the stimuli). For mismatched groups, responses are classified by the modality selected (visual, tactile, audio) or as absent. Visual feedback generally dominated recognition in matched conditions, whereas responses in mismatched conditions varied by modality.

494 However, in *AudioHapticMatch* trials, participants more often relied on the mismatched visual cue than  
 495 in *VisuAudioMatch* or *VisuHapticMatch*, suggesting that visual feedback can still dominate even when  
 496 incongruent with other modalities.

497 Qualitative feedback offers a partial explanation: the Hit animation was consistently recognized due  
 498 to its distinctive preparatory motion and clearly defined impact, while Tap and Caress animations were  
 499 harder to differentiate. As in Experiment 1, participants noted that for Tap and Caress the agent first  
 500 performed a similar reaching movement toward the upper arm, with the gestures diverging only in the more  
 501 subtle hand motion at contact. Given the headset’s field of view, this endpoint difference was not always  
 502 prominent, which likely contributed to the visual ambiguity between these two gestures. This pattern  
 503 matches the recognition data (Figure 7), where Hit reached over 80% accuracy regardless of sensory cues.



**Figure 7.** Recognition rates of the three gestures across the 27 multisensory trials of Experiment 2 (excluding VISUONLY trials). Columns correspond to the visual animation, rows to the audio feedback condition, and stacked bars within each cell to the tactile feedback condition. Colors indicate the gesture recognized. The hit animation achieved consistently higher recognition than tap or caress, regardless of the presence of additional sensory cues.

504 For Tap and Caress, participants reported depending more on the additional modalities, highlighting the  
 505 role of multisensory congruence when visual cues are ambiguous. As in Experiment 1, this ambiguity also  
 506 appeared in the visual-only trials, showing that the limited recognizability of the Caress gesture originates  
 507 from the gesture’s subtle visual form and shared reaching phase rather than from constraints of the tactile  
 508 feedback design.

509 Regarding **H5**, results did not confirm that audio feedback drives recognition more strongly than  
 510 haptic feedback. VISUAUDIOMATCH and VISUHAPTICMATCH showed comparable performance, and  
 511 mismatched conditions revealed no advantage for either modality. One possible explanation is the  
 512 ecological validity of the audio stimuli. Several participants noted that some sounds, especially Caress, felt  
 513 unconvincing and even resembled environmental noises, raising the possibility that realistic recordings  
 514 may appear less believable in immersive contexts (Marini et al., 2012).

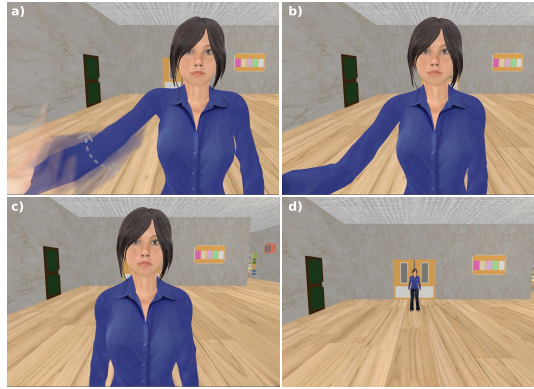
515 A few limitations should be noted. The questionnaire continued to rely on the term “touch,” which may  
 516 have biased participants toward tactile sensations. In addition, the audio feedback, despite being validated  
 517 in prior work (de Lagarde et al., 2023), may have been less effective in VR due to context-specific factors.  
 518 These limitations motivated refinements in Experiment 3.

519 Taken together, Experiments 1 and 2 show that congruent multimodal feedback strengthens both the  
 520 feeling of being touched and gesture recognition, with tactile input playing a central role. Yet, the relative  
 521 contributions of audio and tactile cues remain inconclusive, as both improved recognition only when visual  
 522 input was ambiguous. To further address the role of agency and the social interpretation of touch, we  
 523 revised the protocol in Experiment 3 to manipulate the VA’s non-verbal behavior and interpersonal distance.

## 7 THIRD EXPERIMENT

### 524 7.1 Experimental and stimuli design

525 For this third experiment, we refined our approach to better isolate the respective contributions of each  
 526 sensory modality to the feeling of being socially touched. While Experiments 1 and 2 provided initial  
 527 insights, they also highlighted limitations in stimulus realism and interpretability, particularly for the



**Figure 8.** Visual feedback conditions used in Experiment 3. (a) FULLANIMATION: the VA within peripersonal space, performing the tap gesture (blur indicates motion). (b) FIXED: the VA within reach, with the arm extended but no animation. (c) CLOSE: the VA standing motionless within reach. (d) FAR: the VA standing motionless outside peripersonal space.

528 auditory and visual channels. In Experiment 3, we therefore introduced a new set of stimuli with more  
 529 controlled visual conditions and revised audio recordings, focusing exclusively on the *tap gesture*.

530 This choice was guided by both practical and conceptual considerations. First, including only one  
 531 gesture kept the number of conditions manageable ( $4 \times 4 = 16$ ), avoiding excessive participant fatigue  
 532 and overly complex analyses. Second, compared to the other gestures, *Tap* struck a balance of neutrality  
 533 and recognizability: it was consistently well-recognized in the earlier studies, but without the strong  
 534 visual salience of *Hit* or the interpretive ambiguity of *Caress*. In addition, its short and repeatable rhythm  
 535 facilitated precise alignment of tactile and auditory feedback. We therefore considered *Tap* the most suitable  
 536 candidate for systematically probing modality contributions under refined conditions.

### 537 7.1.1 Visual feedback conditions

538 To better control the VA's presence and movement, we defined four distinct visual feedback levels forming  
 539 one independent variable:

- 540 •FULLANIMATION: the VA is positioned within peripersonal space (70 cm away from the participant) and  
 541 performs a tap animation.
- 542 •FIXED: the VA is within reach, with its arm extended and hand close to the participant's arm, but without  
 543 animation.
- 544 •CLOSE: the VA stands motionless within reach, facing the participant.
- 545 •FAR: the VA stands motionless outside of peripersonal space (public proxemic distance), 7 meters away.

546 This structure allowed us to examine the extent to which movement and proximity contributed to perceptions  
 547 of agency and believability, independently of other modalities.

### 548 7.1.2 Audio stimuli

549 Feedback from earlier studies suggested that direct skin-to-skin recordings (de Lagarde et al., 2023) lacked  
 550 believability, as they captured only contact vibrations without the airborne qualities by which humans  
 551 normally perceive touch sounds. In response, we recorded a new tap sound using a microphone positioned  
 552 to capture air-propagated transmission. The stimulus was generated by gently tapping a clothed arm to  
 553 mirror the physical context of the experiment. Rather than aiming for perfect ecological fidelity, this design

554 prioritized believability: a sound that listeners would plausibly associate with a tap on their own arm in VR,  
 555 even if simplified compared to real-world acoustics.

### 556 7.1.3 Tactile stimuli

557 For consistency, we reused the tactile pattern for tap defined in Experiments 1 and 2: five 40 ms square-wave  
 558 pulses at 40% intensity, applied to the eight central actuators, with a temporal structure matching the audio  
 559 rhythm (two pulses separated by 160 ms, a 500 ms pause, followed by three pulses with 160 ms gaps). This  
 560 ensured comparability across experiments.

### 561 7.1.4 Animation design

562 A new tap animation was created using motion capture to improve movement naturalism and to reduce  
 563 the mechanical appearance reported in earlier experiments. The rhythm of the tap motion was aligned  
 564 with the revised audio and tactile feedback. To simplify interaction and free the participants' right arm  
 565 for responding, the tap animation was applied to the participants' left arm. Despite these improvements,  
 566 we acknowledge that the animation remained limited in expressiveness, as it did not include facial or  
 567 non-verbal cues that are typically integral to social interactions.

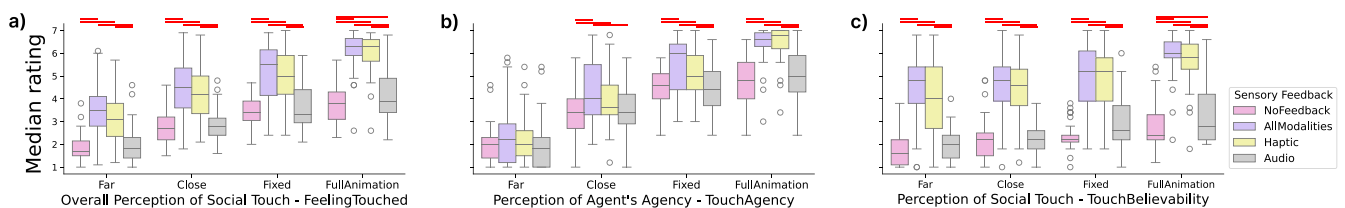
### 568 7.1.5 Multisensory feedback combinations

569 In addition to visual feedback, we defined a second independent variable with four levels of multisensory  
 570 feedback:

- 571 •NOFEEDBACK: no supplementary audio or tactile stimuli.
- 572 •HAPTIC: tactile feedback only.
- 573 •AUDIO: audio feedback only.
- 574 •ALLMODALITIES: combined tactile and audio feedback.

575 Participants experienced all 16 combinations of visual and multisensory feedback ( $4 \times 4$ ). This design  
 576 enabled us to investigate both the isolated and combined influence of sensory modalities under controlled  
 577 visual conditions.

## 578 7.2 Procedure



**Figure 9.** Box plots of all 16 trials in Experiment 3, grouped by visual feedback conditions. (a) FEELINGTOUCHED. (b) TOUCHAGENCY. (c) TOUCHBELIEVABILITY. Red bars indicate significant differences in pairwise comparisons.

579 The experiment comprised all 16 combinations of the two independent variables. Each trial followed  
 580 the same sequence: the scene first faded to black, during which the VA was positioned according to the  
 581 condition (e.g., close, far, arm fixed near the participant, etc.). The scene then faded back in, and the  
 582 stimulus was delivered with the corresponding sensory feedback (audio, tactile, both, or none). After  
 583 stimulus presentation, the screen faded to black again, and participants completed the questionnaire

**Table 4.** Questionnaire used in Experiment 3.

ID	Questions	Format
<b>The agent being the cause of the touch (<i>TouchAgency</i>)</b>		
A1	I felt the agent touch me.	7-pt Likert
A2	It felt like I was not in the same place as the agent.	7-pt Likert
A3	The agent could reach me.	7-pt Likert
A4	The agent wanted to touch me.	7-pt Likert
A5	The stimulus was caused by the agent.	7-pt Likert
<b>Believability of the stimuli as touch (<i>TouchBelievability</i>)</b>		
T1	I did not feel touched.	7-pt Likert
T2	I expected to feel something on my body.	7-pt Likert
T3	I was touched on the arm.	7-pt Likert
T4	The stimulus conveyed social touch believably.	7-pt Likert
T5	The stimulus was not sufficient to elicit the impression of being touched.	7-pt Likert

584 described in Section 7.3. They could then initiate the next trial by pressing a designated button. Trial order  
585 was fully randomized.

### 586 **7.3 Measures**

587 In Experiments 1 and 2, our questionnaire on the feeling of being socially touched relied heavily on the  
588 word *touch*. This phrasing may have unintentionally biased participants toward interpreting their experience  
589 primarily in terms of vibrotactile sensations, rather than as a broader social phenomenon involving multiple  
590 modalities. For Experiment 3, we therefore revised the questionnaire to reduce this bias and to disentangle  
591 different dimensions of the experience more clearly.

592 Specifically, we introduced two subscales:

593 •**Touch Believability:** the extent to which participants felt that the stimulus was convincing as a touch-like  
594 event, regardless of modality. This captures the sensory plausibility of the experience.

595 •**Touch Agency:** the extent to which participants attributed the stimulus to the VA as an intentional, social  
596 actor. This reflects whether participants identified the VA as the cause of the stimulus and perceived it as  
597 a social act rather than a purely external event.

598 To our knowledge, no validated questionnaire currently exists for assessing perceived interpersonal  
599 touch or social attribution in immersive human-agent interaction. Consequently, both subscales were  
600 custom-designed for the present study, with item phrasing and conceptual inspiration drawn from prior  
601 work on embodiment, social presence, and human-agent behavior. Items were written to avoid direct use  
602 of the word *touch*, emphasizing instead believability, causality, and agency. Each item used a 7-point  
603 Likert scale, and three items (A2, T1, and T5) were reverse-scored. Internal consistency was satisfactory  
604 for both subscales (TOUCHAGENCY:  $\alpha = 0.81$ ; TOUCHBELIEVABILITY:  $\alpha = 0.80$ ), indicating good  
605 internal reliability. Item-level diagnostics suggested that all Agency items contributed positively to the  
606 scale, whereas within Believability, the item "I expected to feel touch" showed weaker coherence with the  
607 others, which could be refined in future iterations.

608 The resulting questionnaire is presented in Table 4, with items grouped under the two subscales. Although  
609 this questionnaire has not yet been validated as a psychometric instrument, it represents a methodological  
610 step toward mitigating lexical bias and explicitly targeting the constructs of sensory believability and social  
611 attribution.

612 Unlike Experiments 1 and 2, gesture recognition was not assessed in Experiment 3, as the focus shifted  
613 to the role of interpersonal distance and non-verbal behavior in shaping participants' attributions of  
614 believability and agency. By refining the questionnaire in Experiment 3 to reduce lexical bias and separate  
615 the constructs of believability and agency, we obtained a more precise account of how participants attributed  
616 both the sensory plausibility of the stimulus and the intentionality of the VA, strengthening the interpretation  
617 of findings across experiments.

## 618 7.4 Hypotheses

619 The main motivation of this third experiment is to distinguish and quantify the contribution of each sensory  
620 modality to the feeling of being socially touched by a VA. Based on related works (Section 2) and our  
621 Experiments 1 and 2 results, we hypothesize that:

- 622 •**H6**: Visual, tactile, and audio feedback will contribute differently to the sense of being touched (**RQ2**).
- 623 •**H6.1**: Multisensory feedback will increase the perception of being socially touched, and tactile feedback  
624 will increase it more than audio feedback (**RQ2**).
- 625 •**H7**: Visual feedback will enable feeling **socially** touched when the VA is inside the participants'  
626 peripersonal space (**RQ3**).

## 627 7.5 Results

628 Thirty-one participants took part in Experiment 3 (16 female, 13 male, 2 preferred not to disclose;  $M = 25$ ,  
629  $SD = 5.0$ ).

630 From the post-trial questionnaire (see Section 7.3 and Table 4), we computed three outcome metrics:  
631 TOUCHAGENCY, TOUCHBELIEVABILITY, and FEELINGTOUCHED. The first two were obtained by  
632 averaging the relevant items (with reverse-scored items handled appropriately), while FEELINGTOUCHED  
633 was calculated as the mean of both scales combined. Distributions of these scores across conditions are  
634 shown in Figure 9.

### 635 7.5.1 Main effects of visual and multisensory feedback.

636 Levene's tests indicated significant violations of homogeneity of variances for both visual ( $p = 0.013$ )  
637 and multisensory feedback ( $p < 0.0001$ ). Since the data were also non-normal, we applied rank-based  
638 repeated-measures ANOVAs for each outcome. Results revealed significant main effects of visual feedback  
639 on TOUCHAGENCY ( $F(3, 480) = 218.49$ ,  $p < 0.0001$ ), TOUCHBELIEVABILITY ( $F(3, 480) = 31.42$ ,  
640  $p < 0.0001$ ), and FEELINGTOUCHED ( $F(3, 480) = 123.09$ ,  $p < 0.0001$ ). Similarly, multisensory feedback  
641 showed significant effects on TOUCHBELIEVABILITY ( $F(3, 480) = 156.88$ ,  $p < 0.0001$ ), TOUCHAGENCY  
642 ( $F(3, 480) = 24.64$ ,  $p < 0.0001$ ), and FEELINGTOUCHED ( $F(3, 480) = 100.57$ ,  $p < 0.0001$ ). Finally, there  
643 was a significant interaction between visual and multisensory feedback for TOUCHAGENCY ( $F(9, 480) =$   
644  $1.96$ ,  $p = 0.043$ ).

### 645 7.5.2 Post-hoc comparisons.

646 Wilcoxon signed-rank tests with Holm correction further specified these effects. All pairwise Wilcoxon tests  
647 were corrected for multiple comparisons using the Holm–Bonferroni method. For both FEELINGTOUCHED  
648 and TOUCHBELIEVABILITY, all pairwise comparisons between conditions were significant, except between

649 ALLMODALITIES and HAPTIC (for all visual levels) and between NOFEEDBACK and AUDIO (for FAR,  
650 CLOSE, and FIXED visual conditions). For TOUCHAGENCY, 13 out of 24 possible pairwise comparisons  
651 reached significance. Full results are reported in Figure 9 and in the supplementary analysis repository.

### 652 7.5.3 Item-level analyses.

653 To further examine how visual feedback shaped the social attribution of the agent, we conducted  
654 Wilcoxon pairwise comparisons across the four visual conditions for each of the five items comprising  
655 TOUCHAGENCY. This involved 30 comparisons (6 per item). Of these, 27 were significant. The three  
656 non-significant results were: (i) item A2, comparing CLOSE vs. FIXED ( $W = 115.5$ ,  $p = 0.500$ ); (ii)  
657 item A3, comparing CLOSE vs. FIXED ( $W = 86.5$ ,  $p = 0.084$ ); and (iii) item A3, comparing FIXED vs.  
658 FULLANIMATION ( $W = 71.5$ ,  $p = 0.129$ ).

## 659 7.6 Discussion

660 Experiment 3 extended the first two studies by examining how the *context of interaction*, rather than only  
661 sensory congruence, influences the perception of being touched by a VA. Specifically, we manipulated the  
662 interpersonal distance between participants and the VA to test whether proximity conditions the attribution  
663 of agency and the believability of touch. In doing so, this experiment directly addressed **RQ3**, which asked  
664 how visual cues and multisensory feedback shape participants' perception of the VA's social ability.

665 The results show that distance strongly modulated participants' responses. When the VA was positioned  
666 within reach, tactile events were consistently attributed to her, and the feedback was experienced as  
667 intentional touch. By contrast, when the VA stood further away, participants were less likely to link the  
668 stimuli to her actions and instead described the sensations as external or system-driven. This difference  
669 highlights that the spatial configuration of the interaction is critical for grounding touch as a social act,  
670 even when the same tactile patterns are used.

671 Gesture animation further reinforced agency attribution: static poses such as FIXED or CLOSE elicited  
672 lower scores than the FULLANIMATION condition, where movement clearly conveyed intentionality.  
673 Multisensory feedback amplified this distinction, with haptics again being the dominant contributor and  
674 audio adding smaller but consistent support. These results indicate that participants do not automatically  
675 attribute agency to the VA whenever tactile feedback is present; instead, agency emerges when proximity,  
676 movement, and sensory cues jointly signal the VA as the plausible source of the touch.

677 Experiment 3 therefore adds an important dimension to our investigation. While Experiments 1 & 2  
678 emphasized how visual, auditory, and tactile modalities combine to create convincing sensations, the  
679 present results show that contextual factors, especially distance and animation, determine whether these  
680 sensations are integrated as part of an intentional social gesture. Together, the three experiments indicate  
681 that successful simulation of social touch requires not only multimodal congruence but also a plausible  
682 context, especially regarding interpersonal distance, that affords agency attribution.

## 8 CONCLUSION

683 This work provides the **first systematic exploration** of how visual, auditory, and tactile feedback (alone  
684 and in combination) contribute to the perception of social touch in IVEs. Building on earlier studies of  
685 sensory substitution for touch (de Lagarde et al., 2023), social touch in robotics (Teyssier et al., 2020),  
686 and human-agent touch in non-immersive contexts (Huisman et al., 2016), we present one of the first  
687 **controlled investigations** of how these modalities interact when embodied in a virtual agent. While  
688 multimodal interaction has been widely studied in other domains, the specific question of how visual,  
689 audio, and tactile channels jointly shape *believability*, *gesture recognition*, and *agency attribution* in virtual

690 social touch has not previously been addressed. By structuring three experiments around matched and  
691 mismatched feedback, we provide **novel insights into the design space of virtual social touch**, offering a  
692 foundation for more ecologically rich and socially contextualized investigations.

693 Across three experiments, our findings demonstrate that **multimodal feedback enhances the perception**  
694 **of social touch**, but with distinct contributions from each modality. **Haptic feedback consistently played**  
695 **a central role** in eliciting the sensation of touch. Visual feedback strongly influenced gesture recognition,  
696 although its effect depended on the prominent features of the animation and the immersive setup. **Auditory**  
697 **feedback had a more limited role**: while it sometimes aided recognition, its contribution to believability  
698 and agency was smaller than tactile or visual cues. Experiment 3 further revealed that proxemics and  
699 animation detail modulate whether participants attribute social agency to the VA, underscoring the  
700 importance of contextual factors in touch interactions. This directly addressed **RQ3**, demonstrating  
701 that agency attribution does not arise automatically from multimodal feedback, but depends on whether the  
702 interaction context makes the VA a plausible and intentional source of the touch.

703 While our work deliberately explored only a subset of the vast design space of social touch gestures and  
704 stimuli, it provides a foundation for understanding how modality combinations affect social touch in VR.  
705 By isolating matched and mismatched cues, we highlight both the robustness of haptic feedback and the  
706 fragility of believability when stimuli lack congruence. Taken together, these insights inform the design of  
707 future VR systems aiming for richer, more socially meaningful touch interactions.

## 9 LIMITATIONS & PERSPECTIVES

708 Our findings underscore the promise of multimodal integration for simulating social touch, but they also  
709 highlight important limitations and opportunities for future research. These can be grouped into three broad  
710 categories: *technological*, *methodological*, and *socio-cultural* considerations.

### 711 9.1 Technological limitations

712 The first set of limitations arises from the technical means by which we simulated social touch. Visual  
713 feedback in our experiments was primarily varied through gesture animations and interpersonal distance.  
714 Participants, however, frequently remarked on the absence of accompanying nonverbal cues such as facial  
715 expressions, which are central to conveying intention and affect.

716 Similarly, while vibrotactile ERMs provided controlled and repeatable stimuli, they represent only a  
717 narrow part of the rich spectrum of haptic experiences that social touch entails. For the caress gesture,  
718 the actuator layout of the sleeve (four rows of motors arranged around the arm) constrained our ability to  
719 render a continuous stroke *along* the arm's length. In early design exploration, we therefore compared  
720 different trajectories for the vibration sweep: strokes that progressed longitudinally (proximal–distal) and  
721 strokes that progressed *around* the arm's circumference, with actuators firing sequentially from the front  
722 toward the back of the arm. Participants did not report meaningful differences in how these variants were  
723 interpreted as caresses, and the around-the-arm sweep produced the smoothest and most continuous pattern  
724 given the hardware limitations. Although this simplification reduces the naturalism of the caress stimulus,  
725 recognition difficulties were already present in the visual-only trials of Experiments 1 and 2, indicating that  
726 the ambiguity stems from the inherent subtlety of the gesture rather than from the tactile rendering. As such,  
727 the device constraint affects realism but does not undermine the interpretation of the tactile modality's role  
728 in our findings.

729 Audio feedback proved particularly challenging: although Experiment 3 introduced improved air-  
730 propagated recordings, participants still found them less convincing than visual or tactile cues. This

731 suggests that realism alone may not suffice; instead, sound design techniques from film or games, which  
732 emphasize *believability* over literal reproduction, may be more effective.

733 Finally, technical constraints also affected immersion. The limited field of view of the HMD used in  
734 Experiments 1 and 2 may have restricted participants' ability to perceive wide or peripheral animations,  
735 potentially altering how visual cues were integrated with other modalities. While animation fidelity may  
736 always impose some constraints, the subtle motion differences between Tap and Caress would likely remain  
737 challenging to distinguish even with higher-end motion capture or more expressive facial animation, as  
738 these gestures are intrinsically less visually distinct. Future studies using wider-FOV headsets or CAVE  
739 systems could provide clearer insights into these interactions. For these reasons, we do not interpret the  
740 limited photorealism of our agent's animations as a primary confound for gesture recognition. The Hit  
741 gesture, which included a distinct preparatory swing, was recognized reliably even in visual-only conditions,  
742 whereas Tap and Caress shared a similar reaching phase and differed mainly in a subtle endpoint motion.  
743 This structural similarity, combined with the headset's FOV constraints and participants' natural tendency  
744 to focus on the VA's face, is sufficient to explain the observed ambiguity, and would persist even with more  
745 recent motion-captured animations. Consistent with prior work emphasizing plausibility over graphical  
746 realism in VR (Slater, 2009; Pan and Hamilton, 2018; McDonnell et al., 2008), our results indicate that the  
747 critical factor for recognition is the distinctiveness of the underlying kinematics rather than the absolute  
748 realism of the animation.

## 749 **9.2 Methodological limitations**

750 Beyond technology, methodological choices also shaped our findings. In Experiment 3, we deliberately  
751 restricted the design space to tap gestures in order to improve experimental control and reduce participant  
752 burden. This decision necessarily limited ecological variety, leaving unexplored the distinctive social  
753 meanings of other gestures such as hits or caresses. Similarly, while our questionnaires captured dimensions  
754 such as agency and believability, they remain exploratory and require further validation as robust measures  
755 of social touch.

756 Another omission concerns presence and co-presence. Although these constructs are closely tied to  
757 immersion and social realism in VR, we did not include standardized scales to measure them. As a result,  
758 we cannot determine the extent to which participants' perception of touch was shaped by their sense of  
759 "being there" in the virtual environment or of "being with" the agent. Incorporating validated presence and  
760 co-presence measures would strengthen future studies by clarifying these relationships.

## 761 **9.3 Socio-cultural perspectives**

762 Finally, social touch cannot be reduced to technical or methodological parameters alone: it is inherently  
763 embedded in cultural and relational contexts. In our study, gestures were treated mainly as controlled  
764 sensory stimuli to isolate modality effects. Yet touches such as hits, taps, or caresses carry distinct affective  
765 and relational meanings, such as aggression, attention, or intimacy, that participants may have interpreted  
766 differently depending on their background or expectations. By omitting these meanings to focus on sensory  
767 integration, we gained control but sacrificed some ecological validity.

768 Cultural variation adds another layer: our participant pool was drawn from a single cultural setting,  
769 while norms of touch vary widely across societies. Future work should explore how cultural background,  
770 demographics, and relational context (e.g., perceived gender or role of the agent) shape the interpretation  
771 and acceptability of virtual social touch.

772 These considerations naturally raise ethical questions. Because social touch in VR involves intimacy and  
773 potential power asymmetries, respecting personal boundaries is crucial. Designing systems that allow users  
774 to calibrate their comfort levels (for example, by specifying the type, location, or frequency of touch) will  
775 be essential for the acceptability of such interactions.

776 Overall, our study provides an initial but systematic step toward disentangling how different modalities  
777 contribute to social touch in IVEs. By combining tactile, visual, and auditory cues with controlled variations  
778 in proximity and animation, we highlight both the potential and the challenges of designing believable,  
779 respectful, and meaningful virtual social touch interactions.

## CONFLICT OF INTEREST STATEMENT

780 The authors declare that the research was conducted in the absence of any commercial or financial  
781 relationships that could be construed as a potential conflict of interest.

## AUTHOR CONTRIBUTIONS

782 F.Boucaud and G.Richard contributed equally to the design of the experiments and the analysis of the  
783 results. F.Boucaud, G.Richard, C.Pelachaud and I.Thouvenin contributed equally to the redaction of the  
784 manuscript.

## FUNDING

785 This work was funded by the ANR-22-CE33-0010 (ANR MATCH) project grant. This work was also  
786 supported by French government funding managed by the National Research Agency under the Investments  
787 for the Future program (PIA) with the grant ANR-21-ESRE-0030 (CONTINUUM project) and the IDEX  
788 Sorbonne Université as part of the French government's support for the Investissements d'Avenir program.

## ACKNOWLEDGMENTS

789 We also thank INSEAD and its members for supporting the organization of the studies. Yohan Bouvet has  
790 all our thanks for his help with recording the motion capture animations.

## SUPPLEMENTAL DATA

791 The complete statistical analyses for this manuscript are available on the Open Science Framework (OSF)  
792 repository: [https://osf.io/buf4d/overview?view\\_only=56899eb8b0714ed6bad18d09938c3ae4](https://osf.io/buf4d/overview?view_only=56899eb8b0714ed6bad18d09938c3ae4). Due to IRB restrictions, raw participant-  
793 level data cannot be shared. Instead, we provide a fully knitted analysis file that reproduces all statistical  
794 tests and figures reported in the manuscript, ensuring transparency and reproducibility.

## REFERENCES

- 795 Boucaud, F., Pelachaud, C., and Thouvenin, I. (2021). Decision model for a virtual agent that can touch and  
796 be touched. In *20th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS*  
797 *2021)*. 232–241
- 798 Boucaud, F., Pelachaud, C., and Thouvenin, I. (2023). "it patted my arm": Investigating social touch from  
799 a virtual agent. In *Proceedings of HAI '23 (ACM)*, 72–80
- 800 Boucaud, F., Tafiani, Q., Pelachaud, C., and Thouvenin, I. (2019). Social touch in human-agent interactions  
801 in an immersive virtual environment. In *HUCAPP*. 129–136
- 802 de Lagarde, A., Pelachaud, C., Kirsch, L. P., and Auvray, M. (2023). Audio-touch: tactile gestures and  
803 emotions can be recognised through their auditory counterparts

804 Delis, I., Chen, C., Jack, R. E., Garrod, O. G., Panzeri, S., and Schyns, P. G. (2016). Space-by-time  
805 manifold representation of dynamic facial expressions for emotion categorization. *Journal of Vision* 16,  
806 14–14

807 Fitrianie, S., Bruijnes, M., Abdulrahman, A., and Brinkman, W.-P. (2025). The artificial social agent  
808 questionnaire (asaq)—development and evaluation of a validated instrument for capturing human  
809 interaction experiences with artificial social agents. *International Journal of Human-Computer Studies*  
810 199, 103482

811 Geva, N., Uzefovsky, F., and Levy-Tzedek, S. (2020). Touching the social robot paro reduces pain  
812 perception and salivary oxytocin levels. *Scientific reports* 10, 9814

813 Hauser, S. C., McIntyre, S., Israr, A., Olausson, H., and Gerling, G. J. (2019). Uncovering human-to-human  
814 physical interactions that underlie emotional and affective touch communication. In *2019 IEEE world*  
815 *haptics conference (WHC)* (IEEE), 407–412

816 Héron, R., Safin, S., Baker, M., Zhang, Z., Lecolinet, E., and Détienne, F. (2024). Touching at a distance:  
817 the elaboration of communicative functions from the perspective of the interactants. *Frontiers in*  
818 *Psychology* 15, 1497289

819 Hertenstein, M. J., Holmes, R., McCullough, M., and Keltner, D. (2009). The communication of emotion  
820 via touch. *Emotion* 9, 566

821 Hirano, T., Shiomi, M., Iio, T., Kimoto, M., Tanev, I., Shimohara, K., et al. (2018). How do communication  
822 cues change impressions of human–robot touch interaction? *International Journal of Social Robotics* 10,  
823 21–31

824 Hoppe, M., Rossmly, B., Neumann, D. P., Streuber, S., Schmidt, A., and Machulla, T.-K. (2020). A  
825 Human Touch: Social Touch Increases the Perceived Human-likeness of Agents in Virtual Reality. In  
826 *Proceedings of CHI '20* (ACM), 1–11

827 Huisman, G., Bruijnes, M., Kolkmeier, J., Jung, M., Frederiks, A. D., and Rybarczyk, Y. (2013a). Touching  
828 Virtual Agents: Embodiment and Mind. In *9th International Summer Workshop on Multimodal Interfaces*  
829 *(eINTERFACE)* (Springer), 114–138

830 Huisman, G., Frederiks, A. D., Van Dijk, B., Hevlen, D., and Kröse, B. (2013b). TaSSt: Tactile sleeve for  
831 social touch. In *World Haptics Conference* (IEEE), 211–216

832 Huisman, G., Frederiks, A. D., Van Erp, J. B., and Heylen, D. K. (2016). Simulating affective touch: Using  
833 a vibrotactile array to generate pleasant stroking sensations. In *EuroHaptics '16* (Springer), 240–250

834 Jack, R., McPherson, A., and Stockman, T. (2015). Designing tactile musical devices with and for deaf  
835 users: a case study. In *Proceedings of ICMEM, Sheffield, UK*. 23–25

836 Jacucci, G., Bellucci, A., Ahmed, I., Harjunen, V., Spape, M., and Ravaja, N. (2024). Haptics in social  
837 interaction with agents and avatars in virtual reality: a systematic review. *Virtual Reality* 28, 170

838 Jeunet, C., Albert, L., Argelaguet, F., and Lécuyer, A. (2018). “do you feel in control?”: towards novel  
839 approaches to characterise, manipulate and measure the sense of agency in virtual environments. *IEEE*  
840 *transactions on visualization and computer graphics* 24, 1486–1495

841 Jones, S. E. and Yarbrough, A. E. (1985). A naturalistic study of the meanings of touch. *Communications*  
842 *Monographs* 52, 19–56

843 Ju, Y., Zheng, D., Hynds, D., Chernyshov, G., Kunze, K., and Minamizawa, K. (2021). Haptic empathy:  
844 Conveying emotional meaning through vibrotactile feedback. In *Extended Abstracts of CHI '21*. 1–7

845 Kilteni, K., Groten, R., and Slater, M. (2012). The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments* 21, 373–387

847 Kim, H. and Choi, S. (2024). Expressing the social intent of touch initiator in virtual reality using  
848 multimodal haptics. In *2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*

849 (IEEE), 416–425

850 Lapp, H. S. and Croy, I. (2021). Insights from the german version of the social touch questionnaire: How  
851 attitude towards social touch relates to symptoms of social anxiety. *Neuroscience* 464, 133–142

852 Lenay, C. and Tixier, M. (2018). From sensory substitution to perceptual supplementation. *Living*  
853 *Machines: A Handbook of Research in Biomimetic and Biohybrid Systems* , 552–559

854 Marini, D., Folgieri, R., Gadia, D., and Rizzi, A. (2012). Virtual reality as a communication process.  
855 *Virtual Reality* 16, 233–241

856 McDonnell, R., Breidt, M., and Bülthoff, H. H. (2012). Render me real? investigating the effect of render  
857 style on the perception of animated virtual humans. *ACM Transactions on Graphics (TOG)* 31, 1–11

858 McDonnell, R., Jörg, S., McHugh, J., Newell, F., and O’Sullivan, C. (2008). Evaluating the emotional  
859 content of human motions on real and virtual characters. In *Proceedings of the 5th symposium on Applied*  
860 *perception in graphics and visualization*. 67–74

861 McIntyre, S., Hauser, S. C., Kusztor, A., Boehme, R., Mounou, A., Isager, P. M., et al. (2022). The  
862 language of social touch is intuitive and quantifiable. *Psychological Science* 33, 1477–1494

863 Melo, M., Gonçalves, G., Monteiro, P., Coelho, H., Vasconcelos-Raposo, J., and Bessa, M. (2020). Do  
864 multisensory stimuli benefit the virtual reality experience? a systematic review. *IEEE transactions on*  
865 *visualization and computer graphics* 28, 1428–1442

866 [Dataset] Microdrive™, P. (2023). Pico vibe 9mm vibration motor

867 Montagu, A. (1984). The skin, touch, and human development. *Clinics in Dermatology* 2, 17–26

868 Olugbade, T., He, L., Maiolino, P., Heylen, D., and Bianchi-Berthouze, N. (2023). Touch technology  
869 in affective human–, robot–, and virtual–human interactions: A survey. *Proceedings of the IEEE* 111,  
870 1333–1354

871 Pan, X. and Hamilton, A. F. d. C. (2018). Why and how to use virtual reality to study human social  
872 interaction: The challenges of exploring a new research landscape. *British Journal of Psychology* 109,  
873 395–417

874 Püschel, I., Reichert, J., Friedrich, Y., Bergander, J., Weidner, K., and Croy, I. (2022). Gentle as a mother’s  
875 touch: C-tactile touch promotes autonomic regulation in preterm infants. *Physiology & Behavior* 257,  
876 113991

877 Ravaja, N., Harjunen, V., Ahmed, I., Jacucci, G., and Spapé, M. M. (2017). Feeling touched: Emotional  
878 modulation of somatosensory potentials to interpersonal touch. *Scientific reports* 7, 40504

879 Remache-Vinueza, B., Trujillo-León, A., Zapata, M., Sarmiento-Ortiz, F., and Vidal-Verdú, F. (2021).  
880 Audio-tactile rendering: a review on technology and methods to convey musical information through the  
881 sense of touch. *Sensors* 21, 6575

882 Richard, G., Pietrzak, T., Argelaguet, F., Lécuyer, A., and Casiez, G. (2021). Studying the role of haptic  
883 feedback on virtual embodiment in a drawing task. *Frontiers in Virtual Reality* 1, 573167

884 Rosis, F. d., Pelachaud, C., Poggi, I., Carofiglio, V., and Carolis, B. D. (2003). From Greta’s mind to  
885 her face: modelling the dynamics of affective states in a conversational embodied agent. *International*  
886 *Journal of Human-Computer Studies* 59, 81–118. doi:10.1016/S1071-5819(03)00020-X

887 Sallnäs, E.-L. (2004). *The effect of modality on social presence, presence and performance in collaborative*  
888 *virtual environments*. Ph.D. thesis, KTH

889 Schirmer, A., Croy, I., and Ackerley, R. (2023a). What are c-tactile afferents and how do they relate to  
890 “affective touch”? *Neuroscience & Biobehavioral Reviews* , 105236

891 Schirmer, A., Lai, O., Cham, C., and Lo, C. (2023b). Velocity-tuning of somatosensory eeg predicts the  
892 pleasantness of gentle caress. *NeuroImage* 265, 119811

- 893 Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual  
894 environments. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364, 3549–3557
- 895 Slater, M. (2018). Immersion and the illusion of presence in virtual reality. *British journal of psychology*  
896 109, 431–433
- 897 Song, S. and Yamada, S. (2017). Expressing emotions through color, sound, and vibration with an  
898 appearance-constrained social robot. In *Proceedings of the 2017 ACM/IEEE International Conference*  
899 *on Human-Robot Interaction*. 2–11
- 900 Sun, W., Banakou, D., Świdrak, J., Valori, I., Slater, M., and Fairhurst, M. T. (2024). Multisensory  
901 experiences of affective touch in virtual reality enhance engagement, body ownership, pleasantness, and  
902 arousal modulation. *Virtual Reality* 28, 1–16
- 903 Suvilehto, J. T., Glerean, E., Dunbar, R. I. M., Hari, R., and Nummenmaa, L. (2015). Topography of  
904 social touching depends on emotional bonds between humans. *Proceedings of the National Academy of*  
905 *Sciences* 112, 13811–13816
- 906 Teyssier, M., Bailly, G., Pelachaud, C., and Lecolinet, E. (2020). Conveying emotions through device-  
907 initiated touch. *IEEE Trans. on Affective Computing* 13, 1477–1488
- 908 Van Erp, J. B. and Toet, A. (2015). Social touch in human–computer interaction. *Frontiers in digital*  
909 *humanities* 2, 2
- 910 Willemse, C. J., Huisman, G., Jung, M. M., van Erp, J. B., and Heylen, D. K. (2016). Observing touch  
911 from video: The influence of social cues on pleasantness perceptions. In *Haptics: Perception, Devices,*  
912 *Control, and Applications: 10th International Conference, EuroHaptics 2016, London, UK, July 4-7,*  
913 *2016, Proceedings, Part II 10* (Springer), 196–205