Embodiment in Smartphone Augmented Reality: Effects on User Performance

Han Loong Low
University College London
London, UK
hanloong.low.20@alumni.ucl.ac.uk

Daniele Giunchi
University of Birmingham
Birmingham, UK
d.giunchi@bham.ac.uk

Riccardo Bovo
University of Greenwich
London, UK
R.Bovo@greenwich.ac.uk

Pasquale Cascarano
University of Bologna
Bologna, Italy
pasquale.cascarano2@unibo.it

Nick Ritchie
BBC
London, UK
nick.ritchie@bbc.co.uk

Anthony Steed
University College London
London, UK
a.steed@ucl.ac.uk

Enrico Costanza
University College London
London, UK
e.costanza@ucl.ac.uk

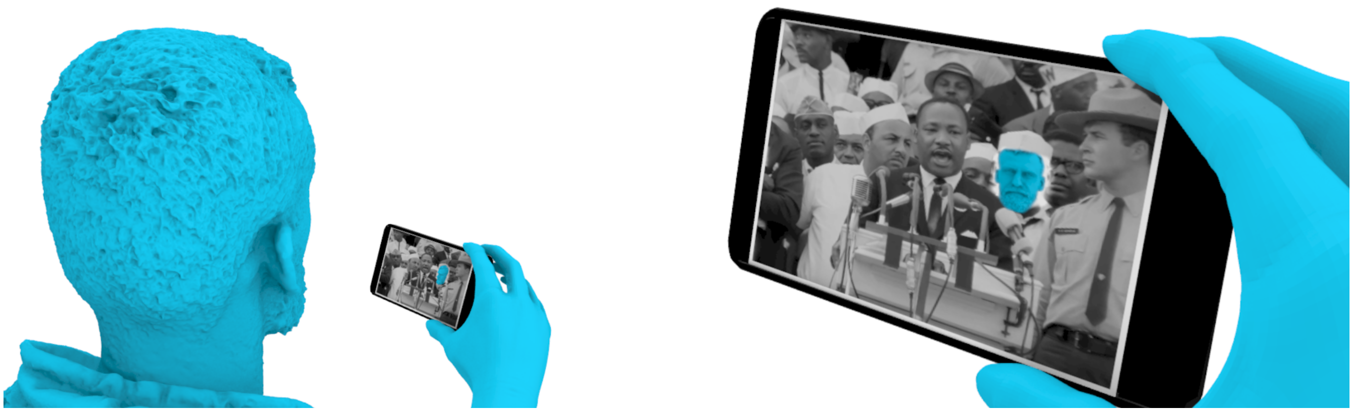

Figure 1: An AR face filter overlaying a participant's realistic face onto a historical photograph of Martin Luther King Jr.'s speech.

Abstract

Embodiment in mixed reality describes the sensation of experiencing a virtual representation as an extension of one's own body. While research has extensively examined embodiment in virtual reality (VR) and head-mounted augmented reality (AR), its impact on smartphones remains underexplored. This study examines how smartphone-based AR embodiment affects user engagement and cognitive performance in a comprehension task. A study involving 24 participants explored whether using a smartphone AR face-filter to embody a virtual audience member influenced the recall of a historical speech. Findings show that participants in the AR condition scored higher on a factual quiz than those in the control group. At the same time, stronger perceived embodiment, especially self-location, was negatively associated with quiz performance, consistent with Cognitive Load Theory. These results should be interpreted cautiously: our comparison contrasted a static image (no AR) with AR that included facial embodiment, so we did not include an "AR without embodiment" condition to fully separate AR novelty from embodiment. Stimuli were also restricted to a single speech and a single historical scene presented as a static image, limiting generalizability to other content and to dynamic or interactive AR. Finally, the sample was modest (N=24), so estimates are preliminary and warrant replication. We discuss implications for designing smartphone AR that balances engagement with cognitive efficiency.

CCS Concepts

• **Human-centered computing** → **Human computer interaction (HCI)**; **Empirical studies in HCI**; **Mixed / augmented reality**.

MUM '25, Enna, Italy

ACM ISBN 979-8-4007-2015-4/25/12
https://doi.org/10.1145/3771882.3771908

Keywords

embodiment, mobile augmented reality, face filters, engagement

ACM Reference Format:
Han Loong Low, Daniele Giunchi, Riccardo Bovo, Pasquale Cascarano, Nick Ritchie, Anthony Steed, and Enrico Costanza. 2025. Embodiment in Smartphone Augmented Reality: Effects on User Performance. In *24th International Conference on Mobile and Ubiquitous Multimedia (MUM '25), December 01–04, 2025, Enna, Italy*. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3771882.3771908

1 Introduction

Recent developments have led to various technologies that can simulate, modify, or enhance our physical and virtual environments. Among these, smartphone-based augmented reality (AR) has gained significant attention from industry and academia due to the ubiquity and accessibility of smartphones [7]. Technological advances, such as mobile processing, embedded cameras, sensors, interaction techniques, software and computer vision improvements, have made smartphones the predominant platform for AR systems. The growing adoption of smartphones, with a projected user base of 6.2 billion by 2029 [43], ensures that smartphone AR experiences are highly accessible to a global audience.

Embodiment refers to the psychological experience of identifying with a virtual representation of oneself, influencing user engagement, perception, and cognition. Despite the growing body of research on embodiment in augmented/virtual/mixed reality, studies have focused on fully immersive environments such as virtual reality (VR) [41, 44] or specialized smart glasses. On the other hand, despite the increasing accessibility of smartphone-based AR, research on embodiment within this context remains limited. This may be attributed to technical constraints such as smaller screens and lower tracking fidelity, which reduce immersion and make embodiment less intuitive than fully immersive VR or smart glasses. These constraints may influence how users experience embodiment, potentially affecting cognitive load and task performance. While previous research suggests that embodiment can enhance engagement and memory recall [25, 28], it remains unclear whether smartphone-based AR embodiment yields similar cognitive benefits or introduces trade-offs that impact user performance.

This study investigates the role of embodiment in smartphone-based AR by examining its effects on user engagement and cognitive performance in a comprehension task. Specifically, we explore whether embodying a virtual audience member via a smartphone AR face filter influences the recall of a historical speech. We designed an AR scenario (as shown in Figure 1) using a face filter that superimposes participants' realistic faces onto a static image of an audience during Dr. Martin Luther King Jr.'s iconic speech.

To isolate the effects of embodiment from external social influences, we employed a passive audience setup, where participants observed themselves within an audience rather than actively engaging with virtual agents. Moreover, we selected a static image from the historical video rather than using a dynamic, full-motion sequence as a moving video could introduce additional visual complexity, temporal attention shifts, and cognitive load, potentially influencing memory recall and engagement independently of embodiment effects. This choice reduces motion-related attentional shifts and temporal complexity but narrows ecological validity compared with dynamic video, interactive 3D, or temporally/location-registered AR; moreover, using a single speech and a single image further limits generalizability across content. As such, resulting inferences should not be over-generalized to richer or multi-stimulus AR experiences. This design choice allows for a controlled investigation of embodiment in smartphone-based AR without introducing confounding variables related to real-time interaction, social presence, or response dynamics. Previous research suggests that social interaction in immersive environments can modulate cognitive load and attention allocation, potentially influencing learning outcomes and engagement [17, 39]. Furthermore, passive audience experiences are common in smartphone-based AR applications, such as virtual museum tours, educational content, and historical reconstructions, where users primarily assume an observer role.

Incorporating a user's real face into virtual or augmented environments enhances the sense of embodiment. This phenomenon, often called the enfacement illusion, is characterized by increased self-identification and body ownership when users see their real faces reflected in a virtual body and demonstrated to increase self-identification with avatars in VR [14]. Studies in avatar customization have found that avatars personalized with users' facial features strengthen the user's sense of control and agency over the virtual body, leading to a heightened embodiment experience [50], impacting in creativity [22], user engagement and sense of self within augmented environments [50].

Our findings reveal that participants in the experimental group outperformed those in the control group on a factual comprehension quiz, indicating that smartphone-based AR embodiment can positively influence memory retention. However, a negative correlation between embodiment levels and quiz performance was observed, suggesting that while embodiment may enhance engagement, it can also introduce cognitive trade-offs by increasing cognitive load. These results align with Cognitive Load Theory (CLT), which posits that excessive presence or agency in an immersive environment can divert attentional resources away from content processing, leading to reduced comprehension despite heightened engagement. This study contributes to the growing body of work on the cognitive and perceptual effects of AR embodiment, providing empirical insights into how immersive experiences should be designed to balance engagement with cognitive effectiveness.

2 Related Work

The following section presents a broad overview of literature related to the rationale and research questions addressed in the present study with a focus on embodiment.

2.1 Embodiment

It is important to clarify the term 'embodiment' as it has been used to refer to vastly different concepts depending on the context and discipline [8]. It generally refers to theories about how our physical bodies and environment shape cognition and behavior [26]. Within HCI, several usages exist [1]; for instance, to refer to spatial [21], on-body (e.g., skin [16]) or kinaesthetic interactions, for example gestures [18], as well as tangible interfaces and wearable technology [9, 10].

Research in embodiment is relatively young, but there is active interest in the field to grow our understanding of the subject. Early research by Longo et al. [24] adopted a psychometric approach to examine embodiment further and found three subcomponents through principal components analysis: (i) body ownership, the extent the virtual body is experienced as user's own body; (ii) agency, the extent the actions of the virtual body are perceived as from oneself; and (iii) location, the extent users feel they are co-located in the same place as the virtual body. Subsequently, Slater [40] proposed a fourth component, plausibility, which is the extent that events depicted in XR are realistically occurring. However, this has yet to be accepted by all researchers.

Different effects of embodiment on perception, attitudes and behavior have been examined. Early studies tended to focus on embodying a specific other. For instance, adults embodying a smaller child body led to perspective distortions and overestimating object size [47]; White people embodying a dark-skinned body led to lowered implicit racial bias [3, 31]; and people embodying a figure that signified intelligence (Albert Einstein) exhibited enhanced cognitive performance [4]. These effects have promising applications and have already been applied in domains such as pain management [27], therapy [15, 30, 33, 46], and rehabilitation [6].

More recently, a few studies have approached embodiment through the lens of immersive journalism, embodying users in specific scenarios about historical and current affairs. In these studies, the emphasis is less on the characteristics of the virtual body itself and more on reliving the scenario through the body of another [41, 44]. This is also the approach of the current study.

Enfacement Illusion. Enfacement illusion (EI) is a perceptual phenomenon when an individual perceives his own reflection while watching the face of another individual [32]. Recent studies have explored different approaches to induce and enhance the enfacement illusion, or measuring the impact in VR on other quantities such as embodiment or agency. Gonzalez et al. [14] investigated avatar embodiment in VR and its ability to generate the enfacement illusion. They looked at how varied amounts of facial animation affected self-identification with avatars. They attempted to improve the impression of enfacement by incorporating synchronous lip motion and additional facial animations. Similarly, La Rocca et al. [20] used computer-generated faces to trigger the enfacement illusion to evaluate the impact of diverse situations on embodiment and enfacement experiences using synchronous multisensory stimulation (visuotactile and visuomotor).

2.2 AR Face filers inducing Embodiment

Several studies [29, 36, 38] have tried to induce embodiment using AR instead of VR. However, these studies adopted full-body tracking using Microsoft Kinect, and with a head-mounted display (HoloLens, Oculus Rift). There are no existing studies examining embodiment in MAR. Of particular interest to the present study are opportunities for AR face filters to enable users to embody a virtual avatar. This may be achieved through various levels of sophistication. The user's face is superimposed or 'swapped' onto the virtual body. This relatively simple implementation is already widely accessible and familiar to the masses through current apps such as Snapchat and Instagram. Hence, this is the implementation adopted in the present study. In contrast, stylised cartoon figure can be programmed to respond to a user's movements in real-time. This can currently be achieved through commercial applications such as loom.ai [34] and FaceRig [45], however, they may be relatively costly to develop or purchase. This implementation is similar to DeepFakes [49], and requires advanced artificial intelligence and processing power [52]. These requirements mean that it is currently not feasible for such an implementation to be conducted in real-time, especially in a mobile context.

2.3 Embodiment in Smartphone AR

In smartphone AR, front-facing cameras on smartphones can perform face/head tracking, blending this real-world information with virtual content. Popular usage of face filters includes modifying facial features or trying on different virtual objects (e.g., eye frames). There is a growing interest in how face filters can affect user perception, attitudes and behavior [5, 11, 35], but studies remain limited. Makled et al. (2022) [25] investigate user embodiment by integrating avatars using smartphone sensors. The study introduces ARIKA, an inverse-kinematic avatar solution that leverages the tracking capabilities of smartphones that allows users to control their avatar's movements creating a sense of agency and self-location. Similarly, Rosa et al.[37] examine virtual body ownership and self-presence in augmented environments by manipulating movement congruency and anthropomorphism. Their findings indicate that movement congruency enhances body ownership while visual similarity plays a lesser role. Our study extends this work by exploring facial movement congruency, where the user's face is dynamically embedded into an AR audience scene. Nimcharoen et al. [28] explore AR reflections and their impact on self-perception and body ownership, demonstrating that virtual mirrors can reinforce embodiment by allowing users to see augmented versions of their real selves. While these works highlight how AR mirrors and avatars influence body awareness and self-location, they often assume active participation. In contrast, our study focuses on passive interaction, examining whether embodied presence in a static AR setting impacts cognitive skills such as attention and comprehension.

Li and Duh (2012) [23] discuss cognitive issues in mobile AR from an embodied perspective, identifying three key factors: information representation, physical interaction, and shared experience. Their framework emphasizes that mobile AR experiences differ from VR or smart glasses due to limited immersion and tracking fidelity, which could impact user cognition and embodiment effects. Our study aligns with these concerns, particularly regarding information representation—by embedding users into a historical context through a static AR frame, we examine whether cognitive processing and recall benefit from embodied engagement despite reduced interactivity.

Additionally, Stalheim and Somby (2024) [42] highlight the role of embodied learning in mobile AR, showing that bodily engagement and social collaboration enhance motivation, engagement, and learning outcomes. Their findings suggest that AR can encourage cognitive engagement even in passive or observational experiences, supporting that embodied presence in AR may contribute to knowledge retention and comprehension. While their work focuses on active movement-based AR, our study extends this inquiry to passive AR settings, investigating whether passively experiencing embodiment (e.g., as part of an audience) can still influence cognitive performance.

Genay et al.[12] explore how full-body avatars improve agency and self-location in AR, reinforcing the importance of realistic representations for embodiment. While full-body tracking enhances presence, our study investigates a partial-body approach—focusing solely on facial embodiment and its cognitive implications in AR-mediated historical narratives. Our study therefore targets a controlled, passive, single-scene framing to probe embodiment's cognitive trade-offs on smartphones; broader claims about interactive/mobile AR require dynamic, multi-scene stimuli. In addition, it contributes to this ongoing discussion by investigating whether smartphone-based AR embodiment can enhance cognitive processing despite its passive nature and reduced sensory immersion.

3 Experiment

This study aimed to investigate the effects of embodiment on user performance in smartphone AR. A one-factorial, between-subjects design was adopted. The independent variable (IV) was the embodiment, while the main dependent variable (DV) was user performance, measured through a factual quiz about the speech participants listened to. The participants' quiz performance was the primary measure of how well they comprehended and retained the information. A quantitative approach was employed to compare two groups: (i) a control group that listened to Martin Luther King's speech while viewing only a static image (no embodiment) and (ii) an experimental group using a smartphone AR face filter that introduced elements of embodiment. The goal was to assess whether embodiment in smartphone AR improved participants' ability to comprehend and recall the speech, as indicated by their quiz performance. In addition, this study also examined user engagement to explore potential relationships with embodiment in smartphone AR. User engagement was assessed using behavioural outcomes, specifically whether participants clicked on a link to learn more about the speech at the end of the study.

3.1 Procedure

We designed an experiment where participants had to virtually attend Martin Luther King Jr.'s 'I Have a Dream' public speech held in 1963. This event was picked due to journalistic and public interests in civil rights and the Black Lives Matter (BLM) movement. Audio and image of the speech were taken from an online article by National Public Radio (NPR) [19], a non-profit American news organisation. The original full speech lasts over 17 minutes; this was shortened to 5 minutes and 18 seconds based on previous studies, which typically used an embodiment process lasting 2 – 5 minutes (e.g., [2, 4, 41, 51]). This duration was also considered reasonable for sustaining participants' attention without distraction or fatigue.

Relevant snippets of the speech were extracted based on the clarity of the audio (some segments of the speech were muffled or distorted by applause, etc.), as well as their relevance to the BLM movement (e.g., police brutality).

This image of Martin Luther King Jr. speaking was sized at 1024*768 pixels. The photograph was chosen because it depicts audience members facing the camera squarely, providing opportunities for participants to embody through a face filter. To maximize experimental control, we used a single speech excerpt and a single still image as the stimulus. This avoids content-level variance (topic difficulty, writing style, visual composition, motion) that could confound embodiment effects, but it also restricts generalizability to other historical speeches and visual scenes. Future work should employ multiple speeches and images (or dynamic video/3D scenes) to test whether the effects replicate across content.

System. A web application for the AR face filter was developed. This allowed for remote testing of participants. This browser-based implementation was chosen as it allows participants to access the study without additional downloads or installations, which may impose a higher burden on participants and security concerns. The application was built using jeelizFaceFilter [48], a publicly available, open-source library based on JavaScript and WebGL. The code for the final application used in the study is publicly available through the following GitHub repository: removed for anonymous review.

Participants' faces were detected from their front-facing webcams using a neural network and tracked in real-time. This was then superimposed or 'swapped' onto an audience member in the image to create the sense that they are experiencing the scenario as a crowd member (see Figure 2 3b). The audience member in the photograph who underwent the face swap was chosen based on two criteria. Firstly, he immediately faced the camera, ensuring his face was captured optimally. Second, his body below the neck was hidden, making the face swap appear more natural and seamless. As the application only tracks the face and not the rest of the body, this helps to maintain congruence and visuomotor synchrony between participants' real and virtual bodies. Various factors were considered when building the face filter. This includes the position, size, dimensions, rotation, and colour correction of the superimposed face onto the image, that is gray scaled. A pilot of the face filter was conducted with two colleagues. Qualitative feedback was collected to design the face filter to look natural and convincing iteratively.

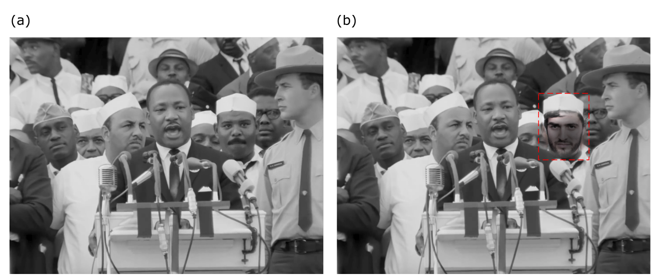

Figure 2: (a) Control condition with a static image; (b) Participants' faces were detected in real-time and superimposed onto an audience member, as shown inside the red contour. Image from National Public Radio [19].

Demographic Measures and Checks. Basic demographic information about age and gender was collected for this study. Participants were also asked if they had listened to the 'I Have a Dream' speech within (i) the past 6 months; (ii) the past year; (iii) the past 3 years;

(iv) greater than 3 years ago; or (v) never heard the speech. This was done to check their familiarity and prior knowledge.

The duration spent on each study page was also collected to ensure that participants did not speed through the study, especially when listening to the speech.

Embodiment Measure. In this study, we adapted the embodiment questionnaire from the framework proposed by Gonzalez-Franco and Peck [13], which calculates embodiment as a composite of body ownership, body localization, agency, tactile sensation, external appearance and response to external stimuli. This was not administered in the control condition due to the complete absence of a virtual body. We propose a tailored version of the embodiment questionnaire designed explicitly for such a mobile experience, as suggested by the authors. Thus, we measured embodiment through three key dimensions: body ownership, agency and motor control, and localization of the body. Participants responded to statements assessing how much they felt the virtual body was their own, their control over it, and its alignment with their real body. The "External Appearance" metrics were excluded from the embodiment questionnaire because the visual modifications in this study were limited to the participant's face, without changes to broader aspects such as body shape, gender, race, or clothing, which are typically assessed using these metrics. The final instrument administered consisted of 10 items. Participants rated on a 7-point Likert scale ranging from -3 (strongly disagree) to +3 (strongly agree) their agreement on various statements reflecting the dimensions. Component scores, as well as a single combined embodiment score, were derived by summing and averaging the corresponding items based on Gonzalez-Franco & Peck's recommendations.

User Engagement Measures. Engagement was assessed using a series of measures. First, engagement was measured using an adapted 3-item questionnaire previously adopted in the Lenin VR study by Slater et al. [41]. Participants indicated on a 7-point Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree) their agreement on statements reflecting curiosity, interest and follow-up intent. This questionnaire was used as a subjective self-reported measure of engagement.

Quiz. Participants were also quizzed on the contents of the speech through four multiple-choice questions. Each question comprised a single correct answer out of four different choices. Questions and answers were derived from the author's personal reading of the speech. The questions were designed to be relatively straightforward, testing facts about what was being said throughout the speech without the need for inference. Martin Luther King Jr.'s 'I Have a Dream' speech is commonly found in educational curricula[1]. The quiz was scored by summing up the number of questions participants answered correctly. Finally, participants were presented with the opportunity to visit a website[2] at the end of the study to find out more about the scenario they had witnessed. Whether or not participants clicked on the link to the website was tracked.

[1] e.g., https://www.pbs.org/newshour/extra/lessons-plans/i-have-a-dream-as-a-work-of-literature-martin-luther-king-jr/

[2] https://www.pbs.org/newshour/extra/lessons-plans/i-have-a-dream-as-a-work-of-literature-martin-luther-king-jr/

3.2 Participants

Twenty-four participants were recruited to participate in the study. Participants (11 female, 13 male, mean age 40.4 with standard deviation 16.3) participated in the study. Given the modest sample size (12 per condition), the study is powered for detecting large effects only; estimates should be treated as preliminary and replicated with larger samples.

To account for any potential bias related to ethnic group identification with the content of the study, we carefully recorded the demographic composition of our sample; all participants in both the experimental and control groups were of either white or Asian ethnic backgrounds. While this was not done by choice, it ensured that any differences observed in embodiment, comprehension, or engagement could be attributed to the experimental conditions rather than to participants' personal or historical connections to the content of the study. Participants were recruited on a voluntary basis with no compensation through university and snowball sampling (participants were asked to forward the study to their families with a good knowledge of spoken English). All participants were blind to the hypotheses of the study. The inclusion criteria for the study required participants to: (i) be at least 18 years old, (ii) own a mobile that supports a display resolution of at least 1024x768 pixels and with a functional front-facing webcam. These requirements ensured consistency and compatibility when running the AR face filter app.

3.3 Procedure

The study was conducted remotely and asynchronously through a web platform hosted on an ANON server. Participants were sent a URL to the study and told them they could complete it on their own time and pace. Upon clicking the link, participants were presented with information about the study. At this stage, checks were conducted, and access was denied to participants whose web browsers were smaller than 1024*768 pixels. After providing informed consent, participants were presented with a set of instructions. They were informed that they would be listening to a 5-minute speech and instructed to "try to imagine that you are at the scene of the speech and listen carefully to what is being said". They were also asked to retrieve and put on a pair of earphones/headphones and ensure they were in a well-lit environment with mobile front-camera access before proceeding with the study.

On the next page, players were either presented with a static image (control condition) or the image with their face superimposed onto an audience member (experimental condition). The speech audio was triggered to play automatically once the AR application detected a face. Participants were required to manually click to the next page once the speech had ended. Subsequently, participants were presented with the questionnaires in the following order: (i) demographics and control, (ii) embodiment questionnaire, (iii) engagement quiz about the speech, and (iv) engagement 3-item self-report.

On the final page, participants were told that they had reached the end of the study and that their results had been recorded. They were also presented with a link to an external website hosted by HistoryExtra, the official website for BBC History Magazine and BBC History Revealed. Participants were told that the site provided additional information and commentary about Martin Luther King Jr. and the civil rights movement that might interest them.

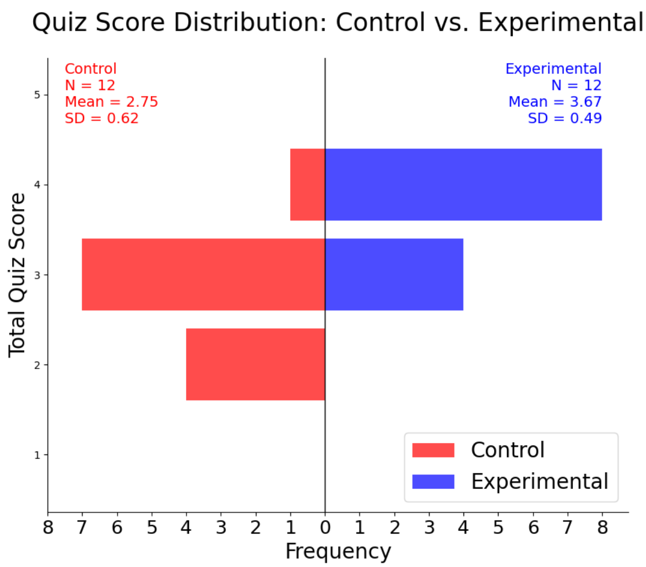

Figure 3: Quiz Scores for experimental condition and control condition participants.

4 Results

The statistical package provided by scipy.stats version 1.14.1 for Python 3.10.9 was employed for all statistical analyses in this study.

4.1 Quiz

We conducted a t-test to assess the difference in quiz scores between the experimental and control groups. The experimental group achieved significantly higher scores (mean = 3.67, SD = 0.49) compared to the control group (mean = 2.75, SD = 0.62), with a t-statistic of 4.00 (p = 0.0006), indicating a statistically significant difference between the two groups, as shown by Figure 3. The effect size, measured by Cohen's d, was significant (d = 1.64), suggesting a substantial impact of the experimental intervention on participants' quiz performance.

4.2 Embodiment

Since participants retained their real facial features through the AR face filter, metrics related to overall external appearance were deemed unnecessary for this context. Descriptive statistics of embodiment measures are reported in Table 1 and Figure 4.

4.3 Engagement

4.3.1 Self-Reported Measures. Descriptive statistics are visualized in Figure 5. For 'curiosity', a Shapiro-Wilk test found that scores in the control group, D(12) = 0.89, p = .13 and in the experimental group, D(12) = 0.87, p = .074 follow a normal distribution. We ran the Kolmogorov-Smirnov test, and the distribution has the same shape (p=0.99). As our sample is sufficiently large, we ran an independent samples t-test. We found that participants in the experimental condition (M = 4.83 SD = 1.62) reported a not significantly different curiosity than control participants (M = 5.25 SD = 0.92), t(24) = 0.73, p = 0.46.

For 'interest', a Shapiro-Wilk test found that scores in both the control group, D(12) = 0.89, p = 0.13 and experimental group, D(12) = 0.91, p = .13, follow a normal distribution. We ran Kolmogorov-Smirnov test, and the distribution have the same shape (p=0.87). An independent samples t-test found no significant difference in interest scores between experimental (M = 4.5, SD = 1.66) and control conditions (M = 5.25, SD = 0.92), t(24) = 1.31, p = 0.20.

For 'follow-up intent', a Shapiro-Wilk test found that scores in both the control group, D(12) = 0.80, p = .011, not following a normal distribution, and the experimental group, D(12) = 0.89, p = 0.13 follows a normal distribution. The distribution of scores for the two groups also share the same shape (p=0.87) using the Kolmogorow-Smirnov test. An independent samples t-test found no significant difference in intent to follow up between experimental (M = 4.83, SD = 1.82) and control conditions (M = 5.25, SD = 0.92), t(24) = 0.68, p = 0.51.

	N	Min	Max	Mean	SD
Body Ownership	12	-0.67	2.33	0.56	0.83
Agency	12	-1.50	1.00	-0.06	0.87
Location	12	-1.00	2.00	0.06	0.94
Combined Score	12	-0.50	1.78	0.18	0.68

Table 1: Descriptive statistics of Embodiment Questionnaire

	Clicked on link?	
	NO	YES
Control	11	1
Experimental	10	2

Table 2: Two-way contingency table showing the number of participants in different experimental conditions who clicked on follow-up link.

4.3.2 Click-through Web Visit. As seen in Table 2 below, two out of twelve participants (16%) in the experimental condition clicked on the follow-up link at the end of the study, compared to one out of twelve participants (8%) in the control condition.

Due to the number of total samples (24 split in 12 experimental and 12 control) and the categorical data we performed a χ^2-test. The test found no statistically significant difference in follow-up web visits between control and experimental conditions (p = 1.00).

4.4 Embodiment on Quiz Results

We conducted a Spearman rank correlation analysis to investigate the relationship between various aspects of embodiment and participants' quiz performance. Our analysis revealed significant negative correlations between both overall embodiment and the location aspect of embodiment with quiz scores. Specifically, we found that the Spearman rank correlation between total embodiment and quiz

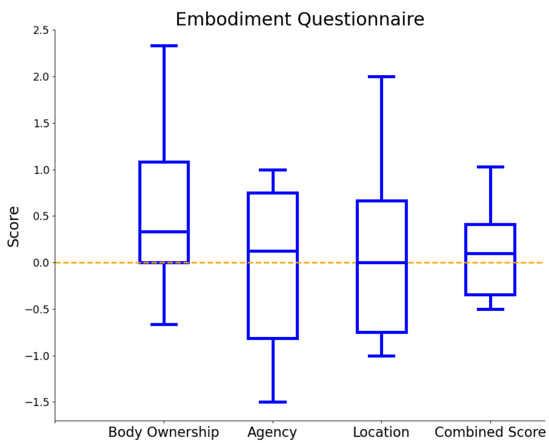

Figure 4: Box-plots showing median and distribution of Embodiment Questionnaire data.

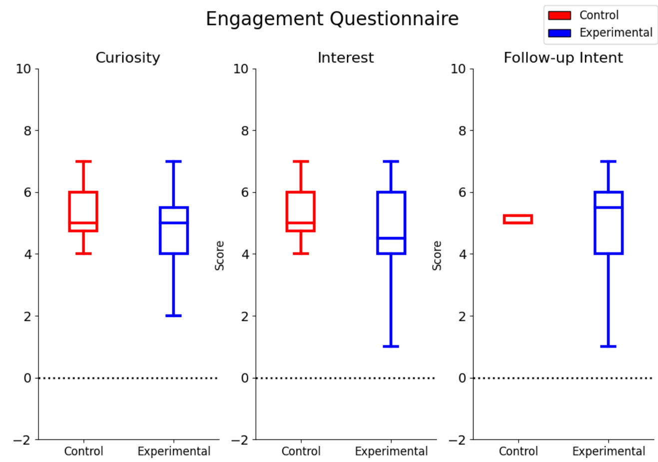

Figure 5: Box-plots showing median and distribution of self-reported engagement measures, grouped by experimental condition.

performance was $r_s = -0.615$, $p = 0.033$, indicating a negative relationship. Additionally, the location aspect of embodiment demonstrated a significant negative correlation with quiz performance, $r_s = -0.648$, $p = 0.023$, suggesting that participants who reported a stronger sense of being collocated with the virtual body tended to perform worse on the quiz. The other components of embodiment did not exhibit significant correlations with quiz scores. Body ownership showed a weak positive but non-significant correlation with quiz performance ($r_s = 0.052$, $p = 0.873$). Meanwhile, agency was moderately negatively correlated with quiz performance, though not significantly ($r_s = -0.492$, $p = 0.104$).

As seen in Figure 6, simple linear regression showed a significant relationship between embodiment and location with quiz scores (for embodiment F = 16.9, p = 0.0014, R2 = 0.29). The regression coefficient B = -0.40 indicated that an increase in one point of embodiment corresponded, on average, to a decrease in quiz score of 0.40. Data met the assumptions of normal distribution for residuals and homoscedasticity.

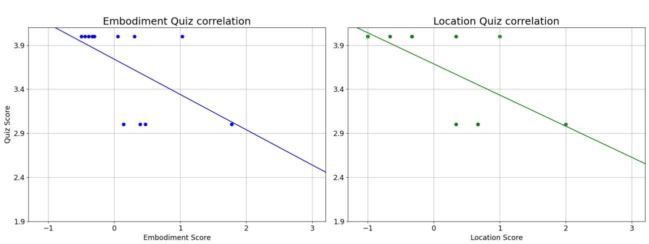

Figure 6: Simple linear regression showing embodiment, agency and location predicting quiz scores.

4.5 Embodiment and Self-Reported Measures

For 'curiosity', a Spearman's rank-order correlation test was run as scores did meet the normality assumption. There is no statistically significant relationship (rs = 0.26, n = 13, p = 0.41) between the variables. For 'interest', a Spearman's rank-order correlation test was run as scores did meet the normality assumption. There was no significant relationship (rs = 0.28, n = 13, p = 0.38) For 'follow-up intent', a Spearman's rank-order correlation test was run as scores did meet the normality assumption. There was no significant relationship (rs = 0.14, n = 13, p = .66)

5 Discussion

The present study aimed to explore the impact of embodiment in smartphone AR on user comprehension, specifically how embodiment influenced participants' performance in a comprehension quiz about Dr. Martin Luther King Jr.'s iconic speech. Descriptive analyses revealed that participants in the experimental group reported positive scores across various subcomponents of embodiment, including body ownership, agency, and location. While participants in the experimental group performed significantly better on the factual quiz than the control group participants, further analysis revealed a negative correlation between embodiment and quiz performance. Higher agency and location scores were associated with lower quiz performance, suggesting that while embodiment can enhance subjective engagement, it may also divert cognitive resources away from content comprehension. These findings can be explained through CLT, which describes how cognitive resources are divided into intrinsic, extraneous, and germane loads. Due to its rhetorical complexity and historical significance, the speech likely imposed a high intrinsic load. At the same time, the AR experience may have introduced additional extraneous load by requiring participants to process both their virtual embodiment and the surrounding digital environment. The negative regression coefficient for location supports the interpretation that participants who felt a stronger presence in the virtual audience might have devoted more cognitive effort to maintaining spatial awareness within the scene rather than focusing on speech comprehension. This aligns with previous studies in VR and AR, where heightened presence sometimes interferes with task performance due to increased cognitive demands.

Although we argue the pattern aligns with CLT rather than a pure novelty effect, we cannot exclude AR-specific novelty or presentation factors without an explicit "AR without embodiment" control. Moreover, because we used a single, static scene, readers should be cautious in generalizing to dynamic or interactive AR. In addition, using a single speech and a single photographic scene prioritizes internal validity at the cost of content diversity; motion and scene variability could modulate attention and cognitive load independently of embodiment. Testing multiple speeches and visual scenes (including moving video) will therefore be important to establish robustness across content and format. Participants in the experimental group reported higher follow-up intent, suggesting that embodiment in smartphone AR can stimulate curiosity about historical content. However, interest scores were uniformly low across conditions, which may be attributed to the way the question was framed.

The survey measured interest in the broader civil rights movement rather than the specific speech event participants experienced, which could have misaligned with their immediate engagement. These results raise broader questions about the implications of smartphone AR for education, particularly in historical or informational contexts. While AR enhances immersion and motivation, its cognitive trade-offs must be carefully managed. The observed inverse relationship between embodiment and factual recall suggests that immersive experiences should be designed to minimize extraneous cognitive load while preserving the benefits of engagement. Strategies such as structuring content delivery to balance immersion with cognitive efficiency, reducing unnecessary visual complexity, and allowing users to adjust immersion settings may help optimize learning outcomes.

6 Limitations & Future Work

When we design the study, we contrasted "No AR, No Embodiment" with "AR with Embodiment", so embodiment and AR-specific factors (e.g., novelty, interface salience) are not fully separable. A third "AR without embodiment" condition is needed to isolate embodiment's unique contribution. As stimulus, We used a single, static photograph to maximize control and reduce motion-related confounds; this limits generalizability to dynamic video, interactive 3D, or temporally/location-registered AR experiences common in practice.

Future studies should try to examine a "With AR, No Embodiment" condition; this might involve an AR experience without the use of face filters/virtual avatars. For instance, exploring a 360° AR environment of Martin Luther King Jr.'s speech through the back-facing camera on a smartphone.

Within smartphone AR, different software implementations can also exist. The present study utilized a simple face-swap onto a static image. However, more dynamic and immersive implementations are also possible (e.g., with video, fully explorable 3D environments). In addition, different effects can be delivered through face filters, such as trying on different outfits, hats, or uniforms to embody other persons. As previously discussed, artificial intelligence (AI) and machine learning (ML) may also soon make it possible to embody and control hyper-realistic virtual avatars [52]. How these different implementations will affect embodiment and its subsequent effects is an interesting area of research for future studies. The dependency of an ad-hoc static image with specific visual conditions (face crowd, not occluded) may limit the broader applicability of the method. However, recent advancements in generative AI technologies offer a potential solution by allowing the creation or modification of original images to ensure they are more compatible with the AR face filter setup, enabling more flexible and adaptive applications of the method.

7 Conclusion

This study explored the impact of embodiment in smartphone AR on user engagement, comprehension, and performance, specifically focusing on the relationship between perceived embodiment and cognitive outcomes. Using a controlled scenario based on a historical event and leveraging a smartphone AR face filter, we showed that smartphone AR can effectively induce a sense of embodiment, even with minimal hardware requirements. Participants in the experimental group demonstrated higher overall quiz scores than the control group, showcasing the potential of smartphone AR for enhancing user interaction. However, our findings also revealed a more complex dynamic between embodiment and task performance. While participants in the experimental group performed better on the quiz overall, those with higher levels of perceived embodiment, particularly in the dimensions of location, tended to perform worse. This suggests that the immersive qualities of embodiment may inadvertently divert cognitive resources away from content comprehension, particularly in tasks requiring focused attention. These findings contribute to the growing body of research on AR and embodiment, highlighting both the opportunities and challenges of integrating embodied experiences into smartphone AR applications. While smartphone AR offers a more accessible platform than fully immersive virtual reality, designers must carefully consider the cognitive trade-offs associated with embodiment. For applications that prioritize learning or information retention, it may be necessary to moderate the degree of embodiment to avoid cognitive overload. Our study therefore targets a controlled, passive, single-scene framing to probe embodiment's cognitive trade-offs on smartphones; broader claims about interactive/mobile AR require dynamic, multi-scene stimuli and multiple content domains.

Acknowledgments

This study was carried out within the MICS (Made in Italy – Circular and Sustainable) Extended Partnership and re- ceived funding from the European Union Next-GenerationEU (Piano Nazionale di ripresa e resilienza (PNRR) – Missione 4 Componente 2, Investimento 1.3 - D.D. 1551.11-10-2022, PE00000004).

References

[1] Alissa N Antle, Paul Marshall, and Elise Van Den Hoven. 2011. Workshop on embodied interaction: theory and practice in HCI. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*. 5–8.

[2] Domna Banakou, Raphaela Groten, and Mel Slater. 2013. Illusory ownership of a virtual child body causes overestimation of object sizes and implicit attitude changes. *Proceedings of the National Academy of Sciences* 110, 31 (2013), 12846–12851.

[3] Domna Banakou, Parasuram D Hanumanthu, and Mel Slater. 2016. Virtual embodiment of white people in a black virtual body leads to a sustained reduction in their implicit racial bias. *Frontiers in human neuroscience* (2016), 601.

[4] Domna Banakou, Sameer Kishore, and Mel Slater. 2018. Virtually being Einstein results in an improvement in cognitive task performance and a decrease in age bias. *Frontiers in psychology* (2018), 917.

[5] Jessica Barker. 2020. Making-up on mobile: The pretty filters and ugly implications of Snapchat. *Fashion, Style & Popular Culture* 7, 2-3 (2020), 207–221.

[6] Adrián Borrego, Jorge Latorre, Mariano Alcañiz, and Roberto Llorens. 2019. Embodiment and presence in virtual reality after stroke. A comparative study with healthy subjects. *Frontiers in neurology* 10 (2019), 1061.

[7] Dimitris Chatzopoulos, Carlos Bermejo, Zhanpeng Huang, and Pan Hui. 2017. Mobile augmented reality survey: From where we are to where we go. *Ieee Access* 5 (2017), 6917–6950.

[8] William Farr, Sara Price, and Carey Jewitt. 2012. An introduction to embodiment and digital technology research: Interdisciplinary themes and perspectives. (2012).

[9] Carlos Flavián, Sergio Ibáñez-Sánchez, and Carlos Orús. 2019. The impact of virtual, augmented and mixed reality technologies on the customer experience. *Journal of business research* 100 (2019), 547–560.

[10] Carlos Flavián, Sergio Ibáñez-Sánchez, and Carlos Orús. 2019. Integrating virtual reality devices into the body: Effects of technological embodiment on customer engagement and behavioral intentions toward the destination. *Journal of Travel & Tourism Marketing* 36, 7 (2019), 847–863.

[11] Carlos Flavián, Sergio Ibáñez-Sánchez, and Carlos Orús. 2021. User responses towards augmented reality face filters: Implications for social media and brands. In *Augmented Reality and Virtual Reality*. Springer, 29–42.

[12] Adélaïde Genay, Anatole Lécuyer, and Martin Hachet. 2022. Being an Avatar "for Real": A Survey on Virtual Embodiment in Augmented Reality. *IEEE Transactions on Visualization and Computer Graphics* 28, 12 (2022), 5071–5090. doi:10.1109/TVCG.2021.3099290

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

[14] Mar Gonzalez-Franco, Anthony Steed, Steve Hoogendyk, and Eyal Ofek. 2020. Using facial animation to increase the enfacement illusion and avatar self-identification. *IEEE transactions on visualization and computer graphics* 26, 5 (2020), 2023–2029.

[15] Arvid Guterstam, Zakaryah Abdulkarim, and H Henrik Ehrsson. 2015. Illusory ownership of an invisible body reduces autonomic and subjective social anxiety responses. *Scientific reports* 5, 1 (2015), 1–8.

[16] Chris Harrison, Shilpa Ramamurthy, and Scott E Hudson. 2012. On-body interaction: armed and dangerous. In *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction*. 69–76.

[17] Chiao Ling Huang, Yi Fang Luo, Shu Ching Yang, Chia Mei Lu, and An-Sing Chen. 2020. Influence of students' learning style, sense of presence, and cognitive load on learning outcomes in an immersive virtual reality learning environment. *Journal of Educational Computing Research* 58, 3 (2020), 596–615.

[18] Mina C Johnson-Glenberg. 2018. Immersive VR and education: Embodied design principles that include gesture and hand controls. *Frontiers in Robotics and AI* (2018), 81.

[19] Martin Luther King Jr. 2010. 'I Have a Dream'Speech, in Its Entirety. *March on Washington for Jobs and Freedom, Washington, DC, August. https://www.npr.org/////i-have-a-dream-speech-in-its-entirety* (2010).

[20] Stefania La Rocca, Silvia Gobbo, Giorgia Tosi, Elisa Fiora, and Roberta Daini. 2023. Look at me now! Enfacement illusion over computer-generated faces. *Frontiers in Human Neuroscience* 17 (2023), 1026196.

[21] Serena Lee-Cultura and Michail Giannakos. 2020. Embodied interaction and spatial skills: A systematic review of empirical studies. *Interacting with Computers* 32, 4 (2020), 331–366.

[22] Joanne Leong, Pat Pataranutaporn, Yaoli Mao, Florian Perteneder, Ehsan Hoque, Janet M Baker, and Pattie Maes. 2021. Exploring the Use of Real-Time Camera Filters on Embodiment and Creativity. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 316, 7 pages. doi:10.1145/3411763.3451696

[23] Nai Li and Henry Been-Lirn Duh. 2012. Cognitive issues in mobile augmented reality: an embodied perspective. In *Human factors in augmented reality environments*. Springer, 109–135.

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

[25] Elhassan Makled, Florian Weidner, and Wolfgang Broll. 2022. Investigating User Embodiment of Inverse-Kinematic Avatars in Smartphone Augmented Reality. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 666–675. doi:10.1109/ISMAR55827.2022.00084

[26] Paul Marshall and Eva Hornecker. 2013. Theories of Embodiment in HCI. *The SAGE handbook of digital technology research* 1 (2013), 144–158.

[27] Matteo Martini, Konstantina Kilteni, Antonella Maselli, and Maria V Sanchez-Vives. 2015. The body fades away: investigating the effects of transparency of an embodied virtual body on pain threshold and body ownership. *Scientific reports* 5, 1 (2015), 1–8.

[28] Chontira Nimcharoen, Stefanie Zollmann, Jonny Collins, and Holger Regenbrecht. 2018. Is That Me?—Embodiment and Body Perception with an Augmented Reality Mirror. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. 158–163. doi:10.1109/ISMAR-Adjunct.2018.00057

[29] Chontira Nimcharoen, Stefanie Zollmann, Jonny Collins, and Holger Regenbrecht. 2018. Is that me?—Embodiment and body perception with an augmented reality mirror. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 158–163.

[30] Sofia Adelaide Osimo, Rodrigo Pizarro, Bernhard Spanlang, and Mel Slater. 2015. Conversations between self and self as Sigmund Freud—A virtual body ownership paradigm for self counselling. *Scientific reports* 5, 1 (2015), 1–14.

[31] Tabitha C Peck, Sofia Seinfeld, Salvatore M Aglioti, and Mel Slater. 2013. Putting yourself in the skin of a black avatar reduces implicit racial bias. *Consciousness and cognition* 22, 3 (2013), 779–787.

[32] Giuseppina Porciello, Ilaria Bufalari, Ilaria Minio-Paluello, Enrico Di Pace, and Salvatore Maria Aglioti. 2018. The 'Enfacement' illusion: A window on the plasticity of the self. *Cortex* 104 (2018), 261–275. doi:10.1016/j.cortex.2018.01.007 Special Section: The body and cognition: the relation between body representations and higher level cognitive and social processes.

[33] Catherine Preston and H Henrik Ehrsson. 2014. Illusory changes in body size modulate body satisfaction in a way that is related to non-clinical eating disorder psychopathology. *PloS one* 9, 1 (2014), e85773.

[34] Loom.ai Product. 2021. Loom.ai. https://loomai.com/product

[35] Juan Sebastian Rios, Daniel John Ketterer, and Donghee Yvette Wohn. 2018. How users choose a face lens on Snapchat. In *Companion of the 2018 ACM conference on computer supported cooperative work and social computing*. 321–324.

[36] Nina Rosa, Jean-Paul van Bommel, Wolfgang Hürst, Tanja Nijboer, Remco C Veltkamp, and Peter Werkhoven. 2019. Embodying an extra virtual body in augmented reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 1138–1139.

[37] Nina Rosa, Jean-Paul van Bommel, Wolfgang Hürst, Tanja Nijboer, Remco C. Veltkamp, and Peter Werkhoven. 2019. Embodying an Extra Virtual Body in Augmented Reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 1138–1139. doi:10.1109/VR.2019.8798055

[38] Nina Rosa, Remco C Veltkamp, Wolfgang Hürst, Tanja Nijboer, Carolien Gilbers, and Peter Werkhoven. 2019. The supernumerary hand illusion in augmented reality. *ACM Transactions on Applied Perception (TAP)* 16, 2 (2019), 1–20.

[39] Claudia Schrader and Theo J Bastiaens. 2012. The influence of virtual presence: Effects on experienced cognitive load and learning outcomes in educational computer games. *Computers in Human Behavior* 28, 2 (2012), 648–658.

[40] Mel Slater. 2009. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364, 1535 (2009), 3549–3557.

[41] Mel Slater, Xavi Navarro, Jose Valenzuela, Ramon Oliva, Alejandro Beacco, Jacob Thorn, and Zillah Watson. 2018. Virtually being Lenin enhances presence and engagement in a scene from the russian revolution. *Frontiers in Robotics and AI* (2018), 91.

[42] Odd Rune Stalheim and Hege Merete Somby. 2024. An embodied perspective on an augmented reality game in school: pupil's bodily experience toward learning. *Smart Learning Environments* 11, 1 (2024), 24.

[43] Statista. 2024. Number of smartphone users worldwide from 2014 to 2029. https://www.statista.com/forecasts/1143723/smartphone-users-in-the-world

[44] Anthony Steed, Ye Pan, Zillah Watson, and Mel Slater. 2018. "We wait"—the impact of character responsiveness and self embodiment on presence and interest in an immersive news experience. *Frontiers in Robotics and AI* (2018), 112.

[45] Holotech Studios. 2021. FaceRig. https://facerig.com/

[46] Pawel Tacikowski, Jens Fust, and H Henrik Ehrsson. 2020. Fluidity of gender identity induced by illusory body-sex change. *Scientific reports* 10, 1 (2020), 1–14.

[47] Ana Tajadura-Jiménez, Domna Banakou, Nadia Bianchi-Berthouze, and Mel Slater. 2017. Embodiment in a child-like talking virtual body influences object size perception, self-identification, and subsequent real speaking. *Scientific reports* 7, 1 (2017), 1–12.

[48] Jeeliz (WebAR.rocks). 2021. jeelizFaceFilter. https://github.com/jeeliz/jeelizFaceFilter

[49] Mika Westerlund. 2019. The emergence of deepfake technology: A review. *Technology Innovation Management Review* 9, 11 (2019).

[50] Jin-Feng Wu, Jiao Dong, Yinglu Wu, and Ya Ping Chang. 2024. Shopping through mobile augmented reality: The impacts of AR embedding and embodiment attributes on consumer-based brand equity. *Information Management* 61, 6 (2024), 103999. doi:10.1016/j.im.2024.103999

[51] Ye Yuan and Anthony Steed. 2010. Is the rubber hand illusion induced by immersive virtual reality?. In *2010 IEEE Virtual Reality Conference (VR)*. IEEE, 95–102.

[52] Egor Zakharov, Aliaksandra Shysheya, Egor Burkov, and Victor Lempitsky. 2019. Few-shot adversarial learning of realistic neural talking head models. In *Proceedings of the IEEE/CVF international conference on computer vision*. 9459–9468.