



POLITECNICO DI TORINO

Corso di Laurea Magistrale in Ingegneria Informatica

Tesi di Laurea Magistrale

*Avatar Realistici per Esperienze in
Realtà Virtuale e nel Metaverso*

Believable Avatars for Virtual Reality Experiences and the Metaverse

A cura del candidato

Gianmario LUPINI

Relatore

Prof. Fabrizio LAMBERTI

Co-Relatore

Davide CALANDRA

Dicembre 2022

Abstract

In un mondo che sta sempre più progredendo verso la virtualità, in particolar modo con il continuo miglioramento dei cosiddetti Ambienti Virtuali (Virtual Environment, in inglese) e del nuovo e più recente concetto del Metaverso, l'importanza di avere un sufficiente livello di realismo in questi mondi digitali diventa un punto chiave per garantire un'esperienza piacevole agli utenti. Più nel dettaglio, molti studi in letteratura hanno mostrato l'importanza dell'avatar, ossia la "rappresentazione di sé stessi nel mondo virtuale", in quanto in grado di influenzare in maniera significativa il comportamento, le azioni e l'esperienza complessiva degli utenti. Inoltre, l'espressività stessa dell'avatar gioca un ruolo chiave in molteplici scenari virtuali, specialmente quelli nei quali è richiesta una comunicazione in tempo reale tra utenti e, quindi, si deve fare un uso intensivo di tutti quei segnali verbali e non verbali del volto e del corpo che tipicamente caratterizzano le interazioni nel mondo reale. Questo lavoro di tesi si concentra su questo ultimo aspetto, studiando e confrontando alcuni approcci per la rappresentazione del viso e del corpo dell'avatar, focalizzandosi in particolar modo sulle emozioni trasmesse, sulle espressioni facciali e sulle diverse modalità di comunicazione verbale e non verbale tipiche di una conversazione.

Indice

| | |
|--|----|
| Abstract | I |
| Elenco delle figure | VI |
| Glossario | IX |
| Acronimi | X |
| 1 Introduzione | 2 |
| 1.1 Contesto ed obiettivi | 2 |
| 1.2 Primo studio: Corpo | 4 |
| 1.3 Secondo studio: Volto | 4 |
| 1.4 Struttura della tesi | 5 |
| 2 Stato dell'arte | 7 |
| 2.1 Importanza dell'avatar in applicazioni XR | 7 |
| 2.1.1 Estetica | 7 |
| 2.1.2 Movimento nel mondo virtuale | 11 |
| 2.1.3 Alcune applicazioni delle potenzialità dell'avatar | 14 |
| 2.2 Animare il volto dell'avatar | 15 |
| 2.2.1 Il problema dell'Uncanny Valley | 16 |
| 2.3 Rappresentazione di Emozioni in VR | 17 |
| 2.3.1 Nozioni teoriche | 18 |
| 2.3.2 Affective Computing | 19 |
| 2.4 Contesti orientati al sociale | 20 |
| 2.4.1 Metaverso | 21 |
| 3 Realizzazione: Primo studio | 24 |
| 3.1 Idea di base | 24 |
| 3.2 Punto di Partenza | 24 |
| 3.2.1 Struttura generale | 25 |
| 3.2.2 Multi-user & Networking | 25 |
| 3.2.3 Inverse Kinematics (IK) | 25 |
| 3.3 Innovazioni introdotte | 26 |

| | | |
|----------|---|-----------|
| 3.3.1 | Avatar - Movimenti | 26 |
| 3.3.2 | Avatar - Logica Animazioni | 26 |
| 3.3.3 | Avatar - Mani | 28 |
| 3.4 | Fase Sperimentale | 28 |
| 3.4.1 | Idea dell'esperimento | 28 |
| 3.4.2 | Ambiente | 30 |
| 3.4.3 | Attività e procedura | 30 |
| 3.4.4 | Questionari e valutazione | 33 |
| 4 | Risultati: Primo studio | 37 |
| 4.1 | Partecipanti | 37 |
| 4.2 | Risultati | 37 |
| 4.2.1 | Embodiment | 37 |
| 4.2.2 | Social Presence | 39 |
| 4.2.3 | Immersione e Presenza | 40 |
| 4.2.4 | Domande Finali di Confronto | 41 |
| 4.3 | Discussione | 42 |
| 5 | Realizzazione: Secondo studio | 44 |
| 5.1 | Idea di base | 44 |
| 5.1.1 | Struttura generale | 45 |
| 5.2 | Avatar | 45 |
| 5.2.1 | Creazione e Caratteristiche del Modello | 45 |
| 5.2.2 | Inverse Kinematics (IK) | 46 |
| 5.3 | SALSA LipSync Suite | 47 |
| 5.3.1 | Implementazione e Adattamento all'Avatar | 48 |
| 5.3.2 | EmoteR & Eyes | 48 |
| 5.4 | HTC VIVE Pro Eye and Facial Tracker | 49 |
| 5.4.1 | Eye Tracking | 50 |
| 5.4.2 | Lip Tracking | 51 |
| 5.5 | Recording e Playback | 51 |
| 5.5.1 | Fase di Recording | 51 |
| 5.5.2 | Fase di Playback | 52 |
| 5.5.3 | Registrazione dell'audio | 52 |
| 5.6 | Problematiche | 53 |
| 5.6.1 | Compatibilità avatar Autodesk® con Facial Tracker | 53 |
| 5.6.2 | Limitazioni hardware del Facial Tracker | 54 |
| 5.7 | Fase sperimentale | 55 |
| 5.7.1 | Idea dell'esperimento | 55 |
| 5.7.2 | Ambiente | 55 |
| 5.7.3 | Attività e procedura | 56 |
| 5.7.4 | Questionari e valutazione | 57 |

| | | |
|----------|---|-----------|
| 6 | Risultati: Secondo studio | 60 |
| 6.1 | Partecipanti | 60 |
| 6.2 | Risultati | 60 |
| 6.3 | Discussione | 65 |
| 7 | Conclusioni e Sviluppi Futuri | 68 |
| 7.1 | Considerazioni finali | 68 |
| 7.2 | Sviluppi futuri | 69 |
| | Appendici | 71 |
| A.1 | Questionario completo del primo caso di studio | 71 |
| A.1.1 | Questionario “Embodiment” | 71 |
| A.1.2 | Questionario “Networked Minds and Social Presence” | 73 |
| A.1.3 | Questionario “VR-Use” | 75 |
| A.1.4 | Domande finali di confronto tra modalità | 75 |
| A.2 | Questionario completo del secondo caso di studio | 76 |
| A.3 | Copioni delle scene mostrate nel secondo caso di studio | 76 |
| A.3.1 | Scena “Happy” | 76 |
| A.3.2 | Scena “Sad” | 77 |
| A.3.3 | Scena “Angry” | 77 |
| A.3.4 | Scena “Fear” | 78 |
| A.3.5 | Scena “Surprise” | 78 |
| A.3.6 | Scena “Disgust” | 79 |
| | Bibliografia | 80 |

Elenco delle figure

| | | |
|-----|--|----|
| 1.1 | Esempi di <i>avatar</i> in due applicazioni per l'interazione virtuale, VRChat e Meta | 3 |
| 2.1 | Esempi della rappresentazione dell' <i>avatar</i> | 9 |
| 2.2 | Alcuni esempi di <i>avatar</i> e delle piattaforme dove vengono usati. L'asse orizzontale rappresenta l'estetica da "infantile" a "maturo", mentre l'asse verticale rappresenta la scala di realismo [17]. | 10 |
| 2.3 | Esempi di Full Body Motion Reconstruction tramite IK con HTC VIVE <i>headset</i> , <i>Tracker</i> e <i>controller</i> [39]. | 14 |
| 2.4 | Grafici relativi al fenomeno dell'Uncanny Valley. Sull'asse verticale è riportata l'empatia di un utente verso l'entità artificiale che si ha davanti; sull'asse orizzontale il livello di piacevolezza. Si noti come nel caso non-stazionario la curva sia molto più profonda. | 17 |
| 2.5 | Le 6 emozioni base secondo Ekman [71] [72]. | 18 |
| 2.6 | L' <i>Affective Loop</i> che caratterizza l' <i>Affective Computing</i> [74]. | 19 |
| 2.7 | Uno schema generale della struttura del Metaverso e delle sue singole componenti [95]. | 22 |
| 3.1 | Schema dei <i>blendtree</i> per il movimento realistico degli <i>avatar</i> nell'esperienza virtuale. | 27 |
| 3.2 | Schema degli stati dell' <i>Animation Controller</i> per il movimento realistico delle mani sulla base delle interazioni nell'esperienza virtuale. | 28 |
| 3.3 | Ambiente semplificato della versione modificata del progetto originale, con i nuovi ostacoli prima e dopo l'auto degli utenti (in nero). Le varie icone indicano la posizione degli oggetti interagibili, il fuoco e infine le frecce indicano i passaggi da superare in posizione accovacciata. Per lo scenario originale più dettagliato si rimanda invece a [7] | 31 |
| 3.4 | Le due rappresentazioni dell' <i>avatar</i> che si andranno a confrontare nello studio. | 31 |
| 3.5 | Alcune immagini dell'esperienza virtuale multi-utente. | 34 |

| | | |
|-----|--|----|
| 4.1 | Risultati medi (con normalizzazione a 1) delle categorie del questionario di <i>Embodiment</i> [99]. I valori statisticamente significativi sono marcati con un simbolo a stella. | 38 |
| 4.2 | Risultati medi delle categorie del questionario di <i>Networked Minds Social Presence</i> [100]. I valori statisticamente significativi ($p < 0,05$) sono marcati con un simbolo a stella. | 39 |
| 4.3 | Risultati medi delle categorie del questionario di <i>VR Use</i> [101]. I valori statisticamente significativi ($p < 0,05$) sono marcati con un simbolo a stella. | 40 |
| 4.4 | Risultati relativi alle domande dirette di confronto dell'ultima parte del questionario complessivo post-esperienza. Per le domande specifiche, si rimanda all'Appendice A.1.4. I risultati statisticamente significativi sono marcati con un simbolo a forma di stella. | 41 |
| 5.1 | Interfaccia Grafica Utente (GUI) di Autodesk® Character Generator. | 46 |
| 5.2 | Opzioni di export dell' <i>avatar</i> generato tramite Autodesk® Character Generator. | 47 |
| 5.3 | Risultato dell'applicazione all' <i>avatar</i> di SALSA LipSync Suite. Vengono mostrati due momenti diversi nel tempo mentre l'audio è riprodotto. | 48 |
| 5.4 | Funzionamento di EmoteR e di Eyes. I raggi gialli costituiscono lo sguardo dell' <i>avatar</i> e puntano ad un punto randomico generato nel rettangolo verde; a destra, l'elenco delle <i>emote</i> definite per l' <i>avatar</i> che si attivano randomicamente mentre l' <i>avatar</i> parla. | 49 |
| 5.5 | Dispositivo Head Mounted Display (HMD) HTC VIVE Pro con annesso hardware per il Facial Tracking. | 50 |
| 5.6 | Limitazioni del rilevamento del sorriso anche nel caso di altri utenti, come lo <i>youtuber</i> Frooxius. | 54 |
| 5.7 | Alcuni estratti dalle scene recitate dall' <i>avatar</i> -attore. | 56 |
| 5.8 | Vista sulla platea dal palco, con l' <i>avatar</i> -attore al centro. | 56 |
| 5.9 | Canvas per la valutazione interno all'applicazione dal punto di vista dell'utente. | 58 |
| 6.1 | Risultati relativi alle domande poste dopo la scena focalizzata sull'emozione della “ Tristezza ” del secondo caso di studio. Per le domande specifiche, si rimanda all'Appendice A.2. I risultati statisticamente significativi, se presenti, sono marcati con un simbolo a forma di stella. | 61 |
| 6.2 | Risultati relativi alle domande poste dopo la scena focalizzata sull'emozione del “ Disgusto ” del secondo caso di studio. Per le domande specifiche, si rimanda all'Appendice A.2. I risultati statisticamente significativi, se presenti, sono marcati con un simbolo a forma di stella. | 62 |

| | | |
|-----|---|----|
| 6.3 | Risultati relativi alle domande poste dopo la scena focalizzata sull'emozione della “Paورا” del secondo caso di studio. Per le domande specifiche, si rimanda all'Appendice A.2. I risultati statisticamente significativi, se presenti, sono marcati con un simbolo a forma di stella. | 63 |
| 6.4 | Risultati relativi alle domande poste dopo la scena focalizzata sull'emozione della “Felicità” del secondo caso di studio. Per le domande specifiche, si rimanda all'Appendice A.2. I risultati statisticamente significativi, se presenti, sono marcati con un simbolo a forma di stella. | 63 |
| 6.5 | Risultati relativi alle domande poste dopo la scena focalizzata sull'emozione della “Rabbia” del secondo caso di studio. Per le domande specifiche, si rimanda all'Appendice A.2. I risultati statisticamente significativi, se presenti, sono marcati con un simbolo a forma di stella. | 64 |
| 6.6 | Risultati relativi alle domande poste dopo la scena focalizzata sull'emozione della “Sorpresa” del secondo caso di studio. Per le domande specifiche, si rimanda all'Appendice A.2. I risultati statisticamente significativi, se presenti, sono marcati con un simbolo a forma di stella. | 65 |

Glossario

Blendshape Corrispettivo del muscolo facciale negli *avatar* virtuali. Animare una *blendshape* vuol dire contrarre o rilassare quel particolare muscolo facciale. Più in generale, è un insieme di vertici della *mesh*, generalmente sul viso, che si muovono e possono essere animati insieme..

Blendtree Strumento che permette di fondere tra loro diverse animazioni sulla base del valore di alcune variabili..

Headset Un headset è uno strumento da indossare in testa in modo da poter vedere immagini rimprese da una camera direttamente davanti agli occhi..

Jitter Indica la variazione di una o più caratteristiche di un segnale..

Package Collezioni di file e data provenienti da progetti di Unity o parte di essi, compressi e uniti in un unico file, come in un archivio `.zip`..

Plugin Programma non autonomo che amplia o estende le funzionalità originarie di un altro programma..

Script Estratto di codice programmato per eseguire una specifica azione oppure gestire una determinata situazione..

Tracker Dispositivo dedito al tracciamento di una parte del corpo, o, più in generale, di un qualche oggetto..

Acronimi

API Application Programming Interface.

AR Realtà Aumentata.

CFD Computational Fluid Dynamics.

CG Computer Grafica.

GUI Graphical User Interface.

HMD Head-Mounted-Display.

IK Cinematica Inversa.

MR Realtà Mista.

NPC Non-Player Character.

SALSA Simple Automated LipSync Approximation.

VR Realtà Virtuale.

XR Realtà Estesa.

Capitolo 1

Introduzione

1.1 Contesto ed obiettivi

Negli ultimi anni lo sviluppo di tecnologie di virtualizzazione o, più nello specifico, di tecnologie di *eXtended Reality* (XR), ha portato ad un intenso susseguirsi di studi volti a cercare di migliorare il più possibile tali tecnologie, apparentemente promettenti, per il futuro prossimo. In particolar modo, nasce l'esigenza di trovare un modo di distinguere e rappresentare i diversi utenti che entrano in uno stesso mondo virtuale, affinché possano interagire tra loro come farebbero nel mondo reale. Numerose ricerche, che verranno esposte nel capitolo successivo, si sono poste il problema di come rappresentare un utente nel mondo virtuale, optando alle volte con soluzioni più "ergonomiche" e minimaliste, come ad esempio mostrare solo le mani, fino a soluzioni molto più vicine alla realtà, con figure complete dalla testa ai piedi.

E' in questo contesto, o più in generale in qualsiasi contesto mediato da una macchina, che si inserisce il concetto di "avatar", definito generalmente come una qualche rappresentazione digitale che rappresenta il "sè" nel mondo virtuale [1]. Tramite gli *avatar*, gli utenti interagiscono con il mondo virtuale e con le altre entità in esso presenti, come ad esempio altri *avatar* di altri utenti. Alcuni studi hanno mostrato che, come nella realtà, gli utenti nel mondo virtuale tendono a formare giudizi o ad attribuire certe "etichette" sulla base degli *avatar* degli altri utenti [2], o addirittura influenzare i loro comportamenti durante le interazioni [3]. Sulla base di questo, studi recenti hanno esaminato quali caratteristiche dell'*avatar*, come tratti estetici, abilità, comportamenti e movimenti arrivino ad influenzare una serie di fattori importanti in campi come la comunicazione interpersonale nel virtuale [4], la comunicazione non-verbale [5] e anche l'*advertising* [6]. Questo mostra come l'aspetto, le espressioni, i movimenti e i comportamenti dell'*avatar* abbiano importanti ripercussioni sull'ambito psico-sociale e percettivo in ambienti multi-utente e di conseguenza multi-*avatar*.

Come sostenuto in [1] nella loro indagine, la ricerca focalizzata sugli *avatar* e la loro rappresentazione è ancora molto acerba, proprio a causa delle enormi

potenzialità e conseguenze che essa può portare in una qualsivoglia esperienza intermediata da una macchina. In particolare, ci sono pochi studi nello stato dell'arte volti a misurare l'impatto della rappresentazione di un *avatar* da un punto di vista più psico-sociologico e percettivo, specialmente in un'applicazione in Realtà Virtuale.

Questa tesi vuole inserirsi in questa corrente, in particolare focalizzandosi sul nostro *alter ego* nel mondo virtuale. Un *avatar*, come già anticipato e come verrà approfondito nei capitoli successivi, può avere diverse forme e strutture, che variano a seconda del caso d'uso o del contesto della particolare esperienza considerata. In Figura 1.1 è possibile osservare alcuni esempi di *avatar* in due applicazioni attualmente molto popolari per l'interazione virtuale, come VRChat¹ e Meta². Si può notare come lo stile sia decisamente diverso, evidenziando di fatto le innumerevoli possibilità di rappresentazione di un *avatar* nel mondo virtuale.

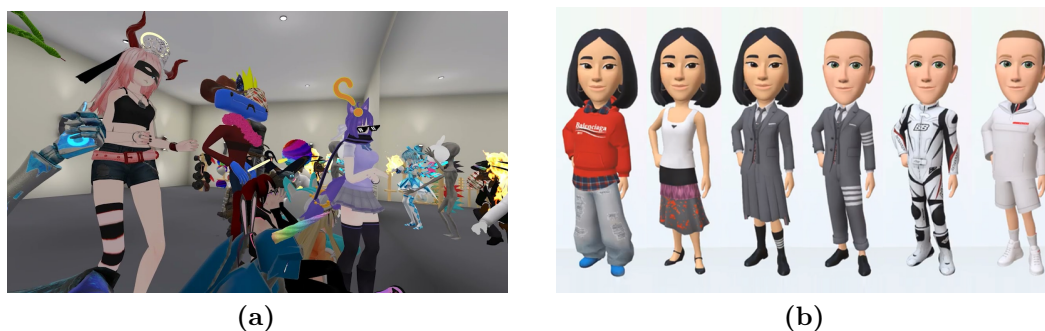


Figura 1.1: Esempi di *avatar* in due applicazioni per l'interazione virtuale, VRChat e Meta

L'idea del presente lavoro è quindi quella di focalizzarsi proprio su tale concetto, in particolar modo osservando e studiando i mezzi al momento disponibili sul mercato e l'impatto che essi hanno nella realizzazione di un *avatar* realistico e quanto più verosimile alla realtà, al reale "essere" dell'utente, soprattutto dal punto di vista dei movimenti. Verrà inoltre prestata molta attenzione su come la rappresentazione del proprio *avatar* impatti sulla percezione e sull'immersività di altri utenti all'interno della stessa esperienza, non tralasciando chiaramente l'impatto sulla percezione di sé e del proprio *avatar*.

Data la complessità del corpo umano e dei suoi movimenti, ma anche per rendere la presente tesi di più facile lettura, è stato deciso di effettuare due studi separati, seppur collegati come nozioni, per esaminare più nel dettaglio le due macro-componenti che caratterizzano un essere umano: il viso, o più in generale la

¹<https://hello.vrchat.com>

²<https://www.meta.com>

testa, ed il corpo. Effettuare due esperimenti permette di considerare con attenzione prima l'una e poi l'altra problematica in modo più indipendente.

1.2 Primo studio: Corpo

L'idea alla base del primo caso di studio è nata considerando un lavoro precedente [7]. In esso, veniva riportato uno studio relativo alla realizzazione di un ambiente virtuale che simulava una situazione di emergenza, in particolare l'incendio in galleria. Nell'esperimento, due o più utenti, ognuno con uno specifico ruolo, collaboravano tra loro nell'ambiente virtuale seguendo le vigenti norme di sicurezza e comportandosi di conseguenza per mettersi al riparo. Una valutazione finale sull'operato veniva poi mostrata ai singoli utenti in modo che potessero rendersi conto se, durante l'esperienza, avessero commesso errori potenzialmente pericolosi o che sarebbero risultati persino letali in un contesto reale. In questo modo, l'esperienza complessiva si inseriva perfettamente in un contesto di educazione e addestramento di utenti civili per situazioni pericolose, rendendolo di fatto un importante strumento per imparare a comportarsi in situazioni pericolose senza trovarsi effettivamente in pericolo. Progetti di questo tipo prendono il nome di *Serious Game*.

In tale lavoro, come molti altri simili, ogni utente era rappresentato da un *avatar*. A seconda del ruolo ricoperto nell'esperienza, l'*avatar* era differente: un *avatar* pompiere per l'utente o gli utenti che svolgevano il ruolo dei soccorritori, ed *avatar* generici per i civili presenti nell'esperienza (camionista e automobilisti). Per gli obiettivi di questa tesi si è partiti da questi *avatar*, cercando di migliorarne la resa estetica, ampliando lo spettro dei movimenti e apportando miglioramenti generali alla loro rappresentazione. Successivamente, si è deciso di modificare il progetto generale in modo da aggiornarlo alle modifiche attuate sull'*avatar* e creare quindi un esperimento collegato, pur mantenendo gli obiettivi e la struttura generale del lavoro originale. I dettagli implementativi verranno trattati nel Capitolo 3, mentre la parte relativa ai risultati verrà discussa nel Capitolo 4.

1.3 Secondo studio: Volto

Per il secondo caso di studio si è intrapresa una strada diversa. L'idea nasce dal problema di come rappresentare al meglio le espressioni facciali sull'*avatar* virtuale, in particolar modo in un'ottica di espressività. Ridere, piangere, corruciarsi, ecc. sono tutti gesti semplici e quotidiani che sono frutto dei numerosi muscoli e delle numerose ossa che compongono la complessa struttura organica della testa dell'essere umano. Tali azioni esprimono qualcosa, che sia un'emozione, un'intenzione, e pertanto diventano di notevole importanza in un qualsiasi contesto sociale, sia nel mondo reale, ma anche nei corrispettivi ambienti virtuali. In quest'ultimo caso, diventa importante riuscire, tramite opportune tecnologie sviluppate negli ultimi anni, a trasporre tutti questi gesti, animazioni ed espressioni facciali anche

sull'*avatar* virtuale, in modo da mantenere il carico espressivo di un dialogo, di una reazione o di una conversazione ad un livello analogo a quello che ipoteticamente si avrebbe nel mondo reale.

Per questo studio, posto e sviluppato temporalmente dopo il precedente, è stato comunque deciso di mantenere le metodologie usate nel primo esperimento, in particolare quelle promosse dai risultati e dai test effettuati per validarlo. Seppure questo secondo studio si concentri principalmente nella parte della testa e del volto, in un ipotetico contesto sociale tale parte del corpo non è l'unica componente che possiede e veicola un particolare carico emotivo ed espressivo, ma è bensì inserita in un insieme che le permette di essere opportunamente contestualizzata. Tale insieme, come illustrato nel capitolo successivo, comprende anche la voce ed i gesti del corpo: grazie a queste tre componenti che si muovono e ruotano assieme in un qualsiasi contesto sociale, le altre persone possono capire cosa si stia cercando di trasmettere.

Sulla base di queste considerazioni, sebbene questo secondo studio sia focalizzato perlopiù sul volto, verranno comunque considerate anche le altre due componenti, trattate con il resto dell'implementazione nel Capitolo 5. I dettagli dei risultati dell'esperimento verranno invece discussi nel Capitolo 6.

1.4 Struttura della tesi

Dopo la breve introduzione riportata, la tesi sarà strutturata essenzialmente mantenendo la suddivisione di questo capitolo. In particolare, si affronteranno i due casi di studio precedentemente menzionati in ordine cronologico, a partire dai Capitoli 3 e 4, con il primo, spiegandone le implementazioni, le tecnologie usate per lo sviluppo del relativo esperimento e infine i suoi risultati. Si proseguirà poi con il secondo, nei capitoli 5 e 6, nel quale verranno esposti in modo analogo tutte le tecniche e le componenti necessarie per il secondo studio, più i suoi risultati. Infine, nel Capitolo 7, verranno tratte le conclusioni generali e verranno proposti alcuni sviluppi futuri che potranno migliorare ulteriormente quanto qui trattato.

Capitolo 2

Stato dell'arte

Negli ultimi anni la ricerca e lo sviluppo sempre maggiori incentrati sulla cosiddetta tecnologia XR (che comprende *Mixed Reality*, *MR*, *Virtual Reality*, *VR* ed *Augmented Reality*, *AR*) hanno contribuito ad un nuovo e progressivo aumento dell'interesse scientifico internazionale riguardo queste nuove tecnologie. L'avvento della pandemia da Covid-19, con tutte le relative misure di restrizione che hanno compromesso inevitabilmente ogni qualsivoglia evento pubblico o azione che prevedeva la presenza di più persone, ha rafforzato inoltre l'esigenza di studiare più approfonditamente e risolvere quei problemi che hanno caratterizzato l'XR fin dagli albori, in modo da poter rendere tali tecnologie un modo efficiente per ovviare e rendere più sopportabile un problema impendente futuro simile a quello vissuto con la pandemia.

In questo capitolo verranno quindi esaminati tutti quei lavori, quegli studi e quelle scoperte ritenuti fondamentali per il lavoro presentato in questa tesi.

2.1 Importanza dell'avatar in applicazioni XR

Possedere un *avatar* in un'applicazione XR comporta numerosi vantaggi [8]. Concentrandosi in particolar modo per le applicazioni VR, emerge già dalla definizione stessa di questa tecnologia l'esigenza di un tramite, di un mezzo con cui possiamo interagire con il mondo virtuale. Essendo la VR un'altra realtà, fittizia e diversa da quella reale, non ci si può entrare in prima persona; di qui quindi l'esigenza di avere un *alter ego* che ci rappresenti in quel mondo fittizio: da questa idea nasce il concetto di *avatar*.

2.1.1 Estetica

Un *avatar* può assumere molteplici forme, diverse tra loro. Possono essere piccoli, grandi, corporei, eterei, umanoidi, eteroformi, fotorealistici, cartooneschi, e così via: non c'è alcuna limitazione al riguardo, in quanto consistono semplicemente in entità nel mondo virtuale. Ogni forma o aspetto che possiede l'*avatar* non è influente dal

punto di vista della sua funzione: esso rimarrà comunque la controparte dell'utente o di un oggetto nel mondo virtuale.

Limitandoci ad esaminare il caso di *avatar* di utenti e non di oggetti, davanti a questa libertà pressochè infinita di caratterizzazione ci si è posti il problema se l'aspetto dell'*avatar* influisca o meno sull'esperienza virtuale. Prima di rispondere a questo quesito, è doveroso suddividere i vari tipi di *avatar* in delle sottocategorie, come proposto da A. Schafer et al. [9] e che è possibile vedere in Figura 2.1:

- **Full Body:** Gli *avatar* di questo tipo si riferiscono generalmente ad *avatar* di tipo umanoide dove tutti gli arti sono attaccati al corpo principale. E' la tipologia di *avatar* più completa e realistica, perchè non tralascia nessuna parte del corpo dell'utente.
- **Upper Body:** In questo caso vengono tagliati fuori gli arti inferiori, lasciando quindi solo testa, mani, torso e braccia. Risultano adatti per applicazioni dove non è richiesto muoversi o camminare, lasciando tuttavia maggiore realismo per la parte superiore del corpo.
- **Head&Hands:** Come dice il nome, gli *avatar* di questo tipo comprendono unicamente una testa, generalmente fluttuante, e un paio di mani, anch'esse fluttuanti e staccate dal corpo. E' una semplificazione della tipologia precedente. E' raro trovare però una rappresentazione di testa e mani esplicita in quanto potrebbe risultare disturbante vedere una testa fluttuante, pertanto spesso vengono visualizzati l'Headset e/o i controller al posto di testa e mani.
- **Hands:** In questo caso viene rimossa persino la testa fluttuante, lasciando unicamente le mani fluttuanti come *avatar*. Generalmente questa tipologia di *avatar* si usa in applicazioni VR in prima persona, dove le mani rappresentano generalmente i controller che l'utente tiene in mano, anche se questa modalità viene più spesso usata in ambito AR.

Quando si parla di Realtà Virtuale, ci sono due importanti concetti da tenere a mente: la presenza e l'immersione. Nel primo caso, si intende generalmente il livello di realismo che un soggetto esperisce dall'interazione con il mondo virtuale, nel suo rapporto con l'ambiente e nella coerenza della sua evoluzione rispetto alle aspettative ed alle previsioni, il tutto dal punto di vista psicologico. E' un concetto un po' volatile e astratto, per certi versi, e a cui molti hanno provato a dare una definizione univoca. Generalmente viene intesa come quella sensazione da parte di un utente che lo fa "sentire" all'interno dell'ambiente virtuale, come se ne facesse parte [12]. Ad esempio, se nel mondo virtuale si lancia una pallina contro un muro, ci si aspetta che quella pallina si muova nella direzione in cui è stata lanciata, con una certa traiettoria, e che urti il muro. Se ciò non succedesse, si perderebbe il senso di presenza o appartenenza all'ambiente virtuale.

Con il termine "immersione" ci si riferisce invece alla capacità dell'ambiente virtuale di coinvolgere direttamente i sensi dell'utente, isolandolo dagli stimoli dell'ambiente reale [13]. Se consideriamo un esempio dove un corpo viene lasciato

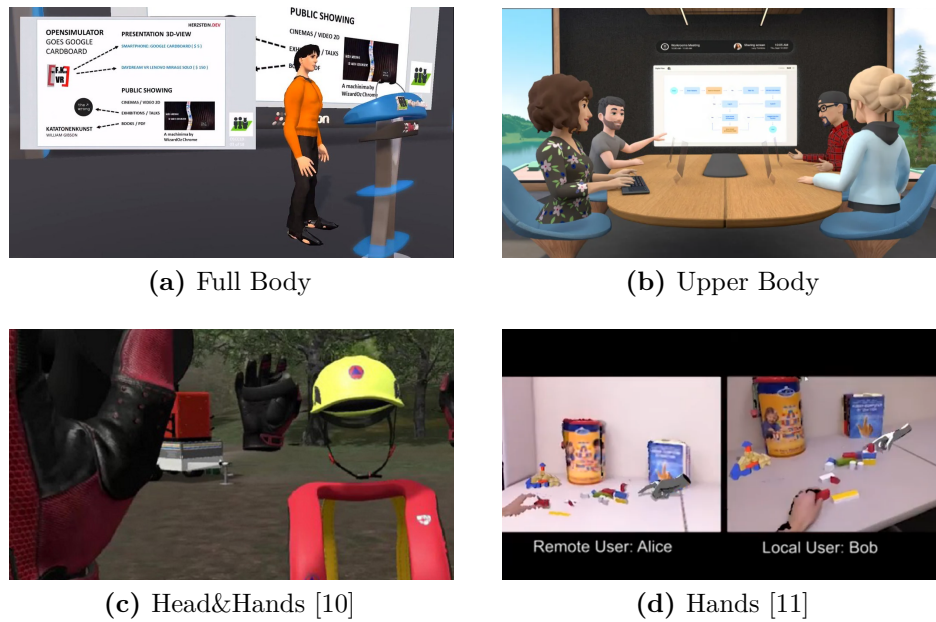


Figura 2.1: Esempi della rappresentazione dell'*avatar*.

cadere, l'immersione è data non solo dalla sensazione tattile dell'oggetto che scivola via dalla mano, dal suono prodotto all'impatto col terreno e dalle conseguenze visive dell'azione, ma anche dall'attivazione dei processi automatici legati al tentativo di riprenderlo prima che tocchi terra e si danneggi. Come si può notare, i due concetti appaiono estremamente vicini tra loro, seppur il loro significato sia nettamente diverso: se l'immersione coinvolge la sfera "corporea" con tutte le risposte dei sensi agli stimoli dell'ambiente virtuale, la presenza invece coinvolge di più l'aspetto "psicologico" dell'utente, che si convince a trovarsi davvero in un ambiente virtuale. Proprio per questi legami con la cognizione, la realtà virtuale può avere un elevato impatto sui videogiochi o applicazioni educative [14].

Ora che tutti i concetti necessari sono stati esposti, si può intuire come la tipologia dell'*avatar* influisca sull'esperienza virtuale. Avere davanti un *avatar* che si avvicini ai canoni "realistici" che di norma vediamo nel mondo reale favorisce enormemente l'immersività nel mondo virtuale, migliorando di conseguenza le prestazioni fisiche e mentali dell'utente quando si interagisce col mondo fittizio, rispetto all'avere un *avatar* molto più semplicistico che non tiene conto di molte parti del corpo. Chiunque troverebbe strano avere un *avatar* che pieghi le ginocchia o i gomiti all'indietro, o che abbia ad esempio le mani e i piedi scambiati. Per questi motivi l'aspetto dell'*avatar* è un fattore non banale da tenere in considerazione nella progettazione di un'applicazione in Realtà Virtuale, e come tale è stato centro di numerosi studi [15] [16]. Al giorno d'oggi vi sono oramai innumerevoli modi e piattaforme per creare un *avatar*, di cui alcuni esempi sono mostrati in Figura 2.2.

Oltre ad evitare dissociamenti di identità tra un utente e il proprio *alter ego*,

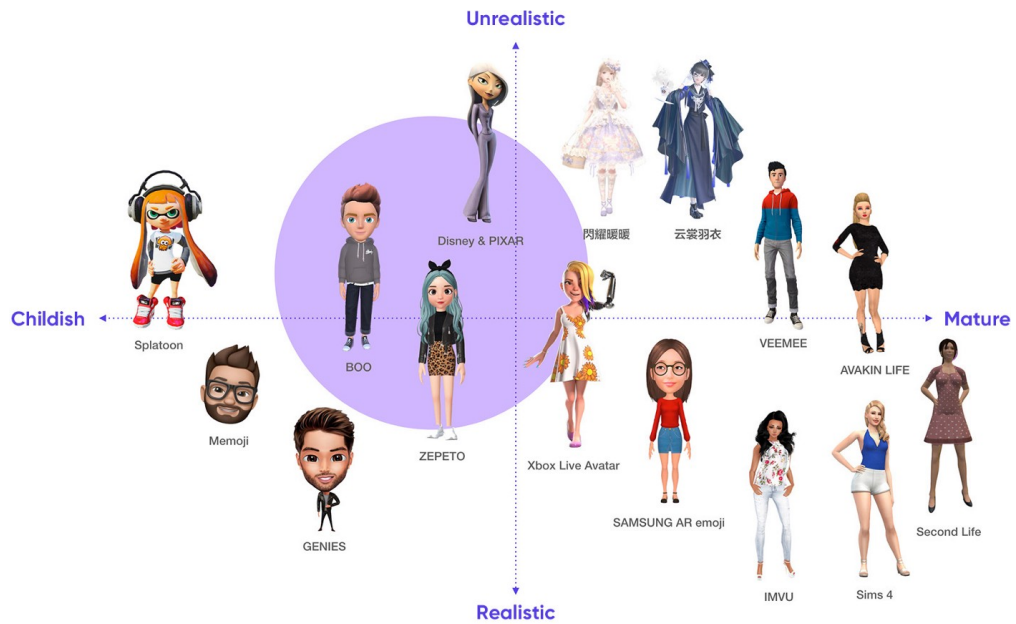


Figura 2.2: Alcuni esempi di *avatar* e delle piattaforme dove vengono usati. L'asse orizzontale rappresenta l'estetica da "infantile" a "maturo", mentre l'asse verticale rappresenta la scala di realismo [17].

un *avatar* virtuale sufficientemente realistico ha anche un forte risvolto psicologico sull'utente. E' stato mostrato come in alcune situazioni i ruoli di utente e *avatar* è come se si ribaltassero, ossia che l'*avatar* stesso arriva ad influenzare i movimenti e il comportamento dell'utente ad esso associato [18] [19]. La scelta degli *avatar* diventa quindi un fattore non banale, che va soppesato e ponderato bene a seconda del caso di studio che si intende intraprendere ma anche al tipo di utenti a cui tale studio è rivolto.

Inoltre, come già citato precedentemente, gli *avatar* stessi possono svolgere un ruolo notevolmente importante in particolari ambiti. Si pensi ad esempio a delle applicazioni VR multi-utente educative, dove un addestratore o un supervisore devono mostrare ad un apprendista le mansioni che deve imparare nel futuro posto di lavoro: vedere un *avatar* compiere le azioni stimola l'azione dei neuroni specchio, quindi è fondamentale che l'*avatar* sia caratterizzato il più dettagliatamente possibile per poterlo imitare al meglio e ridurre quindi al minimo le differenze con un'analoga esperienza nel mondo reale. Numerosi studi simili, sia in ambito pedagogico che dal punto di vista della VR sono stati fatti e sono tuttora portati avanti [20] [21] [22].

Per i fini di questo progetto di tesi, d'ora in avanti l'*avatar* verrà sempre inteso di tipologia *Full Body*, a meno che non sia specificato altrimenti. Inoltre, è bene ricordare che come da titolo della tesi ciò che verrà perseguito sarà un alto livello di realismo, sia che esso riguardi l'estetica, i movimenti o quant'altro.

2.1.2 Movimento nel mondo virtuale

Messo da parte per il momento il lato estetico dell'*avatar*, ci si concentra ora sul come rappresentare al meglio i suoi movimenti nel mondo virtuale. Come già accennato nella precedentemente sezione, i movimenti del nostro *alter ego* virtuale devono generalmente rispettare alcuni canoni intrinseci che derivano dalle nostre esperienze nel mondo reale: a meno di uno scenario horror, un utente generico non giudicherà accurato e verosimile un *avatar* che muove le articolazioni in modo strano o si comporta in modo anomalo rispetto a sè stesso o alla sua concezione degli altri: ciò romperebbe infatti quei sensi di immersione e presenza che tanto sono fondamentali nel mondo della VR. Presenteremo qui di seguito alcuni studi e tecnologie rivolte a tale problematica.

Locomozione nell'ambiente virtuale

La prima domanda che ci si può porre è la seguente: visto che l'*avatar* è il nostro *alter ego* nell'ambiente virtuale, come ci possiamo muovere in tale ambiente? Nel mondo reale ci basta camminare da un posto all'altro, ad esempio, quindi è corretto aspettarsi che si possa fare anche nel mondo virtuale? La risposta è sì, tuttavia non è una risposta assoluta. La totalità degli *headset* al momento presenti sul mercato tiene traccia della posizione della testa dell'utente all'interno della stanza associata all'*headset*, e riesce a trasporre questa posizione anche nell'ambiente virtuale, creando una sorta di rapporto 1 a 1 tra stanza reale dove si trova l'utente e la stanza virtuale dove si trova il suo *alter ego*. Ogni movimento posizionale che verrà fatto nella realtà (ad esempio un passo) verrà trasposto nel mondo virtuale, dove anche l'*avatar* farà un passo nella stessa direzione per aggiornare le proprie coordinate a quelle dell'utente nel mondo reale. Svitati studi hanno infatti dimostrato che il metodo di locomozione più naturale è dato dalla possibilità per gli utenti di camminare nel mondo reale mentre sono immersi nel mondo virtuale, mantenendo quindi quel rapporto 1 a 1 posizionale tra i due ambienti [23] [24] [25].

Non sempre però è possibile questo rapporto 1 a 1 tra ambiente reale e virtuale. Spesso capita infatti che si è ricorsi alla tecnologia della VR proprio per esigenze di spazio, quindi si potrebbe essere costretti a rimanere sul posto nel mondo reale. Molti *headset* in commercio, specialmente quelli più accessibili ai consumatori, agiscono effettivamente su un'area ristretta nel mondo reale che potrebbe limitare l'esperienza dell'utente in ambienti virtuali ben più grandi ed estesi. In tali contesti, sono state ideate nuove tecniche di locomozione, alcune tra queste l'*Arm Swinging* o il Teletrasporto, che permettono all'utente "congestionato" di muoversi comunque liberamente nell'ambiente virtuale.

Una recente panoramica sulle tecniche di locomozione più usate nella letteratura è stata proposta da Cannavò et al. [26], che effettuano questa suddivisione:

- **Repositioning Systems:** Queste soluzioni controbilanciano il movimento dell'utente tramite un qualche dispositivo che li mantiene in una posizione

fissa. Rientrano in questa categoria lo Cyberith's Virtualizer [27] e il KatVR Kat Walk.

- **Proxy Gestures:** In questo caso gli utenti si muovono nell'ambiente virtuale tramite opportuni movimenti fatti con la parte superiore o inferiore del corpo. Alcuni esempi sono le interfacce *leaning*, che sfruttano l'inclinazione del torace per la locomozione virtuale [28], oppure il *walking-in-place* (WIP) che fornisce una sensazione più realistica del camminare, mantenendo però l'utente sul posto [29], oppure infine l'*Arm Swinging*, dove la locomozione è ottenuta tramite ritmici movimenti su e giù delle braccia, simulando quanto avviene in una normale camminata [30].
- **Redirected Walking:** Qui vengono incluse tutte quelle tecniche che influenzano il percorso dell'utente tramite l'ambiente fisico manipolando gli stipoli offerti dall'ambiente virtuale stesso, come ad esempio cambiare ripetutamente il punto di vista dell'utente tramite impercettibili trasformazioni alla mappatura tra movimento reale e virtuale [31].

Da questo si può concludere come vi sia un elevato interesse in materia da parte del mondo accademico e non, che cerca continuamente di evolversi per aumentare il più possibile gli orizzonti di questa tecnologia e renderla accessibile in svariati scenari che di norma risulterebbero come restrittivi.

Ora ci concentreremo sul movimento dell'*avatar* stesso, a prescindere dall'ambiente virtuale. Verrà esaminato lo stato dell'arte più recente relativamente alle tecniche di ricostruzione motoria dell'*avatar*, ossia tutte quelle tecniche che permettono un movimento corretto e sufficientemente realistico degli arti, della testa e del corpo dell'*alter ego* virtuale.

Full Body Motion Reconstruction

In un contesto virtuale, riuscire ad avere una ricostruzione nell'*avatar* dei movimenti di tutto il nostro corpo reale è un fattore necessario per garantire un maggior livello di immersione e una più naturale interazione con il mondo fittizio [32]. Inoltre, associare e tracciare i movimenti dell'utente sul suo rispettivo *avatar* fa in modo che anche la presenza ne benefici [33]. In altre parole, l'immersione richiede un corpo virtuale, mentre la presenza richiede che l'utente si identifichi con questo corpo virtuale, riconoscendone i movimenti come propri [34].

Per fare ciò, negli anni si sono susseguite diverse tecnologie di tracciamento del corpo, distinguendosi in quelle con un apparato di sensoristica [35] e quelle che invece non ne necessitavano, affidandosi ad esempio a delle camere, come il KinectTM. Caserman et al. [36] hanno mostrato nella loro indagine che non c'è una tecnologia vincitrice assoluta della diatriba. In particolare, le tecnologie di tracciamento che fanno uso di sensoristica risultano essere sì più accurate, tuttavia richiedono accessori esterni che possono creare disagio all'utente (come una tuta attillata), oppure ancora che il loro costo sia piuttosto elevato [37]. Viceversa,

le tecniche senza sensoristica da un lato non richiedono particolari hardware o comportamenti *ad hoc* dell'utente, ma dall'altro sono spesso caratterizzati da *Jitter*, inconsistenze e occlusione di alcune parti del corpo.

Per una ragione o per l'altra, quindi, questi due tipi di soluzione possono risultare poco adatti nel caso di applicazioni commerciali o casi di studio dove serve una sorta di "via di mezzo" tra le due. E' qui quindi che entra in scena la tecnica dell'*Inverse Kinematics* (IK), che prende come sensori generalmente testa, mani e talvolta piedi, e stima con opportuni algoritmi la posizione del resto del corpo e delle rimanenti articolazioni.

Inverse Kinematics (IK)

L'approccio dell'IK consiste principalmente nell'avere un insieme di *end point*, o punti estremi, a partire dai quali un qualche algoritmo stima la posizione e la rotazione dei rispettivi arti e delle parti che li compongono usando equazioni cinematiche e la definizione dello scheletro interno all'oggetto [38]. Ad esempio, una soluzione con IK può prendere come *input* il valore posizionale e rotazionale di una mano, generalmente dati dalla posizione e dalla rotazione del *controller* che si tiene in mano, e della testa, data dalla posizione e dalla rotazione dell'*headset* stesso, per stimare la posizione del torace, poi successivamente della spalla, del braccio e infine dell'avambraccio, tali che si raggiunga l'*end point* della mano preso in ingresso con una posa realisticamente credibile.

Sulla base di ciò, si può intuire come il principale problema di queste soluzioni con IK sia il come scegliere il valore posizionale e rotazionale di tutte le articolazioni necessarie per arrivare alla posizione e all'orientamento di un determinato *end-point*. Inoltre, l'algoritmo "risolutore" deve essere veloce al punto che l'utente non percepisca alcun ritardo nei movimenti, deve essere in grado di risolvere il problema di posizionamento per più *end points* contemporaneamente e allo stesso tempo il risultato deve essere un movimento realistico e verosimile [39]. Per queste ragioni, un qualsivoglia algoritmo risolutore dovrà tenere conto di alcuni valori limite delle articolazioni (ad esempio il gomito non può ruotare in un certo modo), e potrà scegliere se intraprendere un approccio di calcolo di tipo analitico o numerico, anche se il primo caso è in generale non considerato a causa dei potenziali lunghi tempi di calcolo che lo rendono inadatto ai requisiti sopra descritti [40].

Il come risolvere e migliorare l'algoritmo di IK è tuttora tema chiave della letteratura, al centro di numerosi studi e soluzioni [33], [41], che hanno portato nel tempo ad una progressiva ottimizzazione delle logiche coinvolte e alla definizione di un paradigma di programmazione molto viabile e dai risultati più che accettabili, come ad esempio quelli raggiunti nello studio in Figura 2.3.

Per il progetto su cui verte la presente tesi è stato quindi scelto un approccio di questo tipo per ricostruire i movimenti dell'*avatar*, usando l'algoritmo risolutore messo a punto dal *Package* di FinalIK dallo *Unity Asset Store*. Maggiori dettagli al riguardo verranno poi discussi nel Capitolo 3.



Figura 2.3: Esempi di Full Body Motion Reconstruction tramite IK con HTC VIVE headset, Tracker e controller [39].

2.1.3 Alcune applicazioni delle potenzialità dell'avatar

In questa sezione verrà esaminato lo stato dell'arte che ha applicato i risultati e le conclusioni esposte dai lavori delle sezioni precedenti, in particolar modo soffermandosi su tutti i casi d'uso che si sono susseguiti nel corso del tempo che hanno mostrato l'importanza del ruolo che l'*avatar* ha in contesti simili a quello che si era pensato di riprendere per il primo caso di studio, ossia uno scenario orientato all'educazione civile per le emergenze, o più in generale, "*Serious Games*".

Per il caso particolare della simulazione di emergenze nella VR a fini di addestramento, negli ultimi anni sono emersi numerosi studi che hanno sottolineato come tale tecnologia si presti in modo molto efficace a tali applicazioni [42], [43], [44]. Riprendendo l'indagine effettuata da Lovreglio et al. [43], tra gli scenari di emergenza più popolari implementati nella Realtà Virtuale ci sono gli incendi, tra cui anche gli incendi in galleria, in cui si inserisce il primo caso di studio sviluppato per questa tesi. Da un'analisi più attenta però si nota che la maggior parte di tali esperienze manchi una qualche componente multi-utente, che invece è presente nel caso di studio su cui si è deciso di lavorare. Sulla linea di ciò, un altro studio è stato fatto e portato avanti recentemente [45] che ha raccolto le ricerche fatte su tali scenari multi-utente e ha poi analizzato le varie componenti che sono fondamentali da inserire per avere un'esperienza coerente quanto più possibile con la realtà e di conseguenza un apprendimento più completo.

Tra questi aspetti, uno forse dei più critici e importanti specie in ambito di casi d'uso collaborativi multi-utente, è dato dal realismo che possiedono gli *avatar* usati per rappresentare i partecipanti stessi, come dimostrato dall'elevato numero di lavori portati avanti sugli effetti che una specifica modalità di rappresentazione di un *avatar* ha sulla percezione degli utenti [46]. Ad esempio, Roth et Al. [47] hanno investigato gli effetti dell'avere *avatar Full Body* poco realistici, andando ad usare dei manichini, proponendo un esperimento volto ad investigare in particolare gli effetti di ridotta informazione sociale e di più ristretti canali comportamentali. Al fine di ciò, le interazioni, sia verbali che non verbali, venivano eseguite sia

nel mondo virtuale che in quello reale, confrontandole poi in termini di presenza sociale, presenza generica, focalizzazione dell'attenzione e performance. I loro risultati hanno mostrato che la mancanza di realismo dell'*avatar* umanoide ha ostacolato e limitato le interazioni sociali, compromettendo di conseguenza anche la performance complessiva degli obiettivi assegnati ai partecipanti, sebbene d'altro canto ciò compensava in qualche modo la mancanza di quei segnali comportamentali che generalmente si notano in una persona davanti a sè, come sguardo o espressioni facciali.

Negli studi di Kasapakis et Al. [48] e Benrachou et Al. [49], sono stati proposti esperimenti caratterizzati da un approccio generalmente simile sebbene sviluppati con casi d'uso e tecnologie differenti. Nel primo caso, gli autori si sono focalizzati sul cercare di capire quanto utile fosse effettivamente l'aver un *avatar* realistico in un ambiente virtuale multi-utente educativo, come una classe. In tale classe virtuale, il docente possedeva una rappresentazione dell'*avatar* realistica, comprendente anche movimenti facciali e degli occhi, mentre quella degli studenti non era di forma antropomorfa al fine di evitare di attirare l'attenzione dell'insegnante. I risultati mostrano che avere *avatar* realistici per quei soggetti più "importanti" all'interno di un particolare contesto porti a migliorare significativamente l'esperienza per tutti gli utenti che stanno partecipando, sottolineando quindi il potenziale che un *avatar* realistico può avere nelle applicazioni virtuali collaborative. Nel secondo studio sopra menzionato, invece, il contesto era più verso uno scenario riabilitativo, usando la sensoristica del KinectTM per cercare di individuare delle posture del corpo umano scorrette o da migliorare nel percorso riabilitativo. Il sistema da loro realizzato si è rivelato essere flessibile e preciso, confermando che una rappresentazione completa del corpo umano all'interno del mondo virtuale possa essere molto importante per una moltitudine di applicazioni.

Per concludere, nonostante l'interesse nello stato dell'arte verso il legame sempre più forte tra Realtà Virtuale e scenari collaborativi, educativi o di emergenze, risulta come l'investigazione sugli impatti che effettivamente ha una rappresentazione *avatar* rispetto ad un'altra è relativamente limitato, e di qui la voglia di investigare con i casi di studio proposti nella presente tesi.

2.2 Animare il volto dell'avatar

Si prosegue dunque all'ultima parte relativa all'animazione di un *avatar*: il volto. Finora è stato considerato come si può animare la locomozione di un *avatar* nell'ambiente virtuale, per poi passare a come di fatto animare le parti stesse che compongono l'*avatar* come gli arti, lasciando fuori quindi l'ultimo tassello necessario per avere un *avatar* con, sulla carta almeno, tutte le caratteristiche di cui necessita per essere verosimile ad una persona in carne ed ossa del mondo reale.

L'animazione del volto è una tematica che è stata affrontata spesso negli ultimi anni, perchè come già ribadito ultimo tassello chiave per avere un *avatar* con tutto ciò che occorre per simulare una persona vera e propria. Grazie alla VR si riesce a

raggiungere infatti quella illusione dovuta alla presenza e all'immersione che il corpo dell'*avatar* diventa proprio dell'utente. Tuttavia, per una totale immedesimazione nell'*avatar* l'utente ha bisogno di avere anche pieno controllo sul proprio volto, e questo fatto contrasta nitidamente con la tecnologia base della Realtà Virtuale che prevede l'indossare un *Head Mounted Display* (HMD) che copre buona parte del proprio volto e che quindi permette un tracciamento facciale molto limitato [50].

E' quindi inevitabile che ciò si ripercuota anche nella capacità dell'*avatar* di trasporre correttamente non solo le espressioni del viso dell'utente corrispettivo, ma anche le emozioni stesse, visto che quest'ultime sono manifestate tramite movimenti del corpo e del viso [51]. Questo argomento verrà poi discusso in una sezione successiva.

Per ovviare quindi al problema dell'ostruzione del visore, sono stati portati avanti studi vari che hanno cercato di effettuare un tracciamento di ciò che l'HMD non copre, in particolare i muscoli facciali della parte bassa del viso [52] [53] oppure, in modo forse più semplicistico, la bocca e le labbra [54]. Altri ancora, tramite opportuni hardware inseriti all'interno dell'HMD, sono riusciti a tracciare il movimento di occhi, palpebre e sopracciglia [55] [56].

In conclusione, seppur tracciando separatamente le varie parti che compongono il viso, si può ottenere un risultato sufficientemente e potenzialmente vicino a quello dei volti reali, garantendo quindi una corretta trasposizione di espressioni e di emozioni dell'utente nel mondo virtuale.

2.2.1 Il problema dell'Uncanny Valley

Avere la possibilità di poter animare potenzialmente le espressioni facciali su un *avatar* o più in generale un agente virtuale non vuol dire però che fare ciò sia immediato e semplice. Quando si lavora con agenti umanoidi, sia che essi siano *avatar* virtuali o robot reali con fattezze umane, bisogna tenere in considerazione il problema dell'Uncanny Valley. Citato per la prima volta in uno studio giapponese nella seconda metà del Novecento, il problema dell'Uncanny Valley consiste nell'ipotesi che l'atteggiamento di una persona nei confronti di un'entità simile ad un essere umano ad un certo livello di realismo cambia completamente passando da empatia a repulsione [57]. Una rappresentazione grafica di questo fenomeno può essere vista in Figura 2.4.

L'idea di base di Mori, l'autore della pubblicazione sovramenzionata, è che, quando si osserva una persona artificiale, sia essa un robot con fattezze umane, un qualche *avatar* umanoide o un qualsivoglia oggetto realizzato in Computer Grafica (CG) si abbia, mano a mano che aumenta il realismo, una forte sensazione di disagio nel guardarlo, come se ci si rendesse conto che qualcosa non vada in ciò che si ha davanti. La sensazione sembra derivare dal fatto che l'osservatore non sa se giudicare ciò che sta vedendo come vivente o meno, o come oggetto animato o meno, e di qui viene inconsciamente generata una sensazione di inquietudine o appunto, di disagio. Lasciando da parte tutte le concezioni filosofiche che ne derivino, come ad esempio ciò che si considera "umano" da ciò che non lo è o simili,

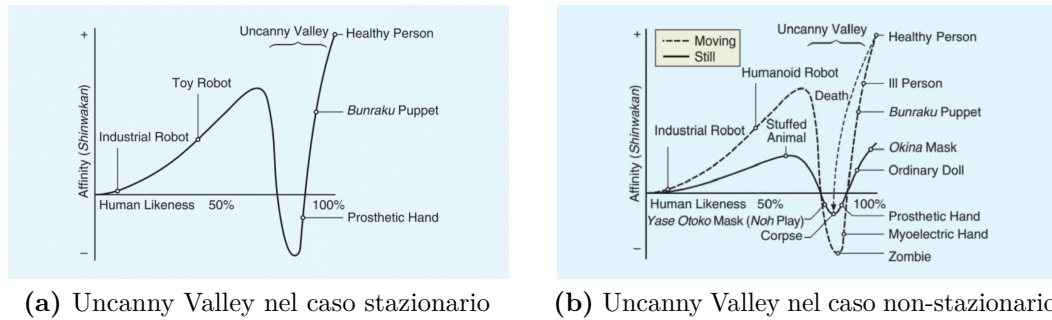


Figura 2.4: Grafici relativi al fenomeno dell'Uncanny Valley. Sull'asse verticale è riportata l'empatia di un utente verso l'entità artificiale che si ha davanti; sull'asse orizzontale il livello di piacevolezza. Si noti come nel caso non-stazionario la curva sia molto più profonda.

il problema dell'Uncanny Valley è nato pur sempre come un'ipotesi teorica, mai di fatto dimostrata in pratica. Alcuni studi hanno osservato che il mancato successo di alcune opere cinematografiche di animazione sia stato dovuto a questo fattore [58], tuttavia allo stesso tempo ci sono stati studi che hanno dimostrato che di fatto non esiste una Uncanny Valley [59] [60] o altri che hanno esaminato tutti i fronti in gioco per avere una risposta [61]. Con il passare del tempo, il termine Uncanny Valley è stato associato piuttosto ad un effetto, un effetto visivo e psicologico che influenza la nostra percezione di un'entità artificiale realisticamente umana. Quali siano però le effettive cause di tale effetto rimane oggetto di studio [62].

Quale sia la risposta della sua esistenza trascende gli obiettivi di questa tesi, tuttavia, ricollegandosi al discorso di questo capitolo, bisogna comunque tenere in considerazione che avere la possibilità di animare le espressioni facciali e il corpo non implica direttamente l'avere un *avatar* realistico e paragonabile ad un essere umano. Dopotutto, l'avere risultati esteticamente strani che suscitano vaghe sensazioni di disagio non è atipico quando si lavora con *avatar* realistici. Bisogna fare attenzione, specie in un periodo dove l'evoluzione del realismo nelle applicazioni è sempre crescente, a non incappare in quell'ipotetico effetto dell'Uncanny Valley che rende un *avatar* o una qualsivoglia persona artificiale soltanto una specie di bambolotto espressivo, ma vuoto e inquietante.

2.3 Rappresentazione di Emozioni in VR

Come già accennato nella precedente sezione, animare il volto rende possibile anche, in linea teorica, avere un mezzo per poter rappresentare le emozioni su un *avatar* virtuale. Il problema della rappresentazione delle emozioni è stato sempre al centro di studi, sia in ambito più umanistico con studi di fattura psico-socio-etologica [51] [63], che in ambito più tecnico, come la rappresentazione delle emozioni su robot,

avatar o simili [64]; in quest'ultimo caso ci si riferisce alla problematica con il nome di *Affective Computing* [65], che verrà esaminato più nel dettaglio qui di seguito.

2.3.1 Nozioni teoriche

Le emozioni sono una parte fondamentale della vita umana e sono di norma correlate ed espresse tramite comunicazione non verbale [66], comportamento [67] e cambiamenti fisici o mentali [68]. Studi in ambito di neuroscienze hanno mostrato il ruolo che le emozioni hanno con la creatività [65], la cognizione [63], il prendere decisioni [69] e più in generale con l'attività cerebrale [70], e per queste ragioni diventa importante per le macchine e i computer avere dati a sufficienza per dapprima identificare e poi esprimere le emozioni, in modo da creare efficienti interazioni uomo-macchina [65].

Particolarmente importante nello stato dell'arte è il lavoro di Paul Ekman [71], che è diventato con il passare del tempo una base solida per qualsivoglia studio sulle emozioni. Secondo l'autore, a giudicare dalle possibili espressioni facciali che una persona può fare si possono definire 6 emozioni di base: gioia, rabbia, tristezza, disgusto, paura e sorpresa. Da queste emozioni base, che coinvolgono ognuna un set unico di muscoli facciali, è possibile, secondo l'autore, derivare tutte le altre. Una rappresentazione visiva di queste emozioni base si può osservare in Figura 2.5.

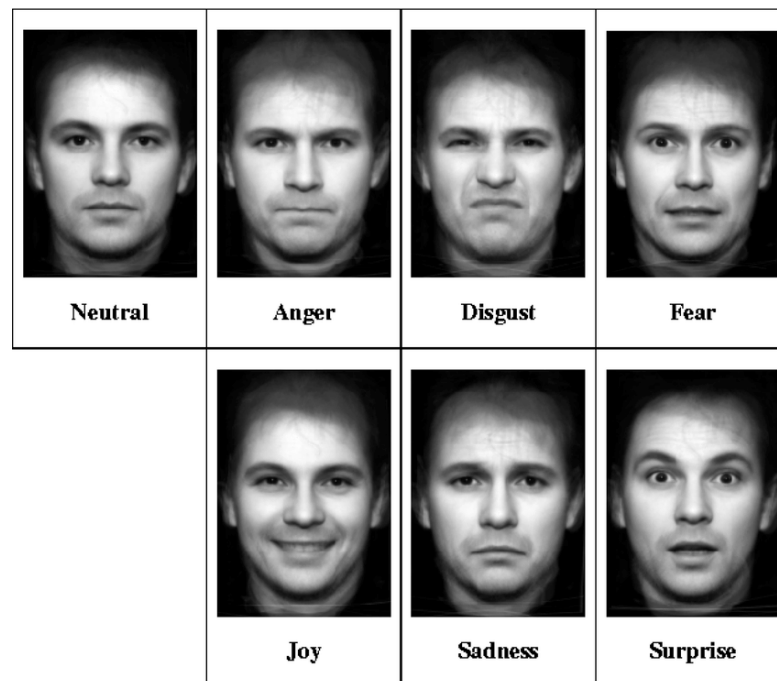


Figura 2.5: Le 6 emozioni base secondo Ekman [71] [72].

Queste nozioni teoriche vanno tenute a mente quando nel progetto ci si preoccuperà di fare in modo che l'*avatar* rappresenti correttamente certe emozioni. In

particolare, bisognerà tenere in considerazione quell'insieme di gesti, di espressioni, di atteggiamento che nel loro insieme vengono associate ad una certa emozione.

2.3.2 Affective Computing

L'*Affective Computing* è emerso negli ultimi anni come un vero e proprio paradigma di computazione che riguarda nella fattispecie la modellazione delle emozioni, il loro riconoscimento e la loro sintesi [73]. Paiva et al. hanno approfondito questa tematica concettualizzando di fatto l'*Affective Computing* come in Figura 2.6: c'è una sorta di ciclo tra l'elicitazione di un'emozione, il suo riconoscimento e la generazione del comportamento associato, ossia l'insieme di quei gesti, espressioni che servono per la sua sintesi finale [74].

Tutto ciò enfatizza il concetto secondo il quale vi è il bisogno di capire a fondo quali siano i meccanismi che governano le emozioni e cosa esse comportano. E' essenziale quindi, per ottenere risultati validi, realizzare un ambiente efficace per le emozioni, e ciò inevitabilmente dipende da un qualche meccanismo efficace di induzione emotiva [75].

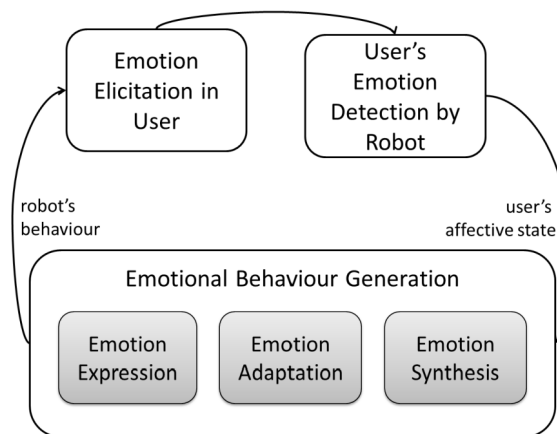


Figure 1.1. *Affective Loop of Emotional Robots.*

Figura 2.6: L'*Affective Loop* che caratterizza l'*Affective Computing* [74].

Emotion Elicitation

Per *Emotion Elicitation* si intende tutto quell'insieme di meccanismi e processi che in un certo contesto sono volti a far scaturire una certa emozione [73]. Tali meccanismi, o "stimoli", devono essere tali da essere organizzabili in un modo tale che è possibile offrire la stessa esperienza per tutti i partecipanti all'elicitazione. Generalmente ci sono due approcci per indurre una particolare emozione: attivo e passivo [76].

L'elicitazione di tipo passivo consiste nell'osservatore che assiste in modo passivo ad un evento emotivo (ad esempio guardare filmati o immagini, ascoltare musica,

ecc.), mentre in un metodo attivo l'osservatore sta attivamente partecipando nell'attività emotiva, generalmente tramite l'uso di un qualche tipo di *avatar* immerso in un ambiente virtuale e/o interattivo. Come si può intuire, in quest'ultimo caso l'esperienza emotiva del partecipante è molto più realistica e potente. Inoltre, la presenza di contenuto interattivo, oltre a far scaturire emozioni più genuine, è altresì importante per interazioni sociali più naturali [77] e ci sono numerosi lavori nello stato dell'arte che hanno incluso tali metodi in ambienti VR [78] [79] o direttamente su *avatar* che rispondevano con comportamenti vari o espressioni sul volto [80] [81] [82].

Emotion Representation

Collegando quanto detto nelle precedenti sezioni, una volta indotta un'emozione ad un utente, come la si rappresenta su un *avatar* virtuale? In linea generale, quando si usa un *avatar* come “mediatore” e come responsabile di convogliare le emozioni, queste si manifestano principalmente con le espressioni facciali [83] e con gesti delle braccia e delle mani [84]. Studi recenti come il sovramenzionato [83], oppure [85] [86], sono riusciti a rappresentare su un *avatar* le emozioni basilari come felicità, sorpresa, paura, tristezza e altre, sia con *avatar* maschili che femminili, evidenziando tutti i cambiamenti nelle varie parti del viso coinvolte con le emozioni stesse. Questi studi dimostrano che un *avatar* in VR opportunamente fatto in modo da rappresentare emozioni può avere enormi implicazioni negli ambiti della interazione sociale o dello sviluppo di abilità sociali, che verranno esaminate nella prossima sezione.

2.4 Contesti orientati al sociale

Tutto ciò che è stato detto finora va applicato in un qualche contesto, e per gli obiettivi di questa tesi tale contesto è in particolare il contesto sociale. Le emozioni sono parte integrante nella vita degli esseri umani, specialmente nella comunicazione interpersonale. Sebbene all'apparenza questo argomento possa sembrare sconnesso dalla presente tesi, in realtà è da considerare perché la comunicazione è strettamente collegata alle emozioni. Nell'atto del comunicare, non viene usata solo la bocca per le parole, ma anche una serie di gesti ed espressioni che manifestano un certo tono, danno una certa tonalità alla conversazione e soprattutto trasmettono le emozioni di chi parla, o di chi ascolta [87]. Lavori recenti come [88] sottolineano che si sente il bisogno, specialmente dal punto vista delle scienze linguistiche e psico-sociologiche, di estendere le correnti capacità tecnologiche facendo in modo di tracciare e renderizzare in *real-time* tutto quell'insieme di *non-verbal behaviours*, come ad esempio i gesti con le mani e le espressioni del viso, appunto, e mostrarle su di un *avatar* in un ambiente virtuale mentre indossa un HMD.

L'interesse della comunità scientifica più “umanistica” della sfera psico-sociale verso la VR o più in generale la “virtualizzazione” ha portato alla pubblicazione di

studi e indagini che sottolineano come la VR possa essere un mezzo alternativo alla realtà per poter effettuare esperimenti, visto che si riesce a garantire, anche nell'ambiente virtuale, certi fattori che fino a poco tempo fa si era soliti indagare solo nel mondo e nelle interazioni reali [89] [90]. Un'altra corrente di lavori, sempre in ambito psico-linguistico, ha trovato nella VR un modo per poter effettuare esperimenti che nella realtà sarebbe complicato effettuare sia per esigenza di risorse, sia per imparzialità [91] [92]. In quest'ultimo lavoro, Heyselaar et Al. mostrano come sia possibile valutare tramite opportuni test e questionari la percezione e l'espressività dell'*avatar* che si ha di fronte in un ambiente virtuale multi-utente, offrendo quindi di fatto un'estensione delle nozioni di espressività non più solo al mondo reale ma anche al mondo virtuale e agli *avatar*.

2.4.1 Metaverso

Si conclude il capitolo facendo un piccolo *excursus* sul Metaverso, concetto tornato alla ribalta nell'ultimo anno nella comunità accademica e secondo molti il futuro prossimo della tecnologia e dell'internet [93].

Il Metaverso viene definito come “l'universo della post-realtà, un ambiente multi-utente perpetuo e persistente che fonde la realtà fisica con la virtualità digitale. E' basato sulla convergenza di tecnologie che abilitano interazioni multisensoriali con gli ambienti virtuali, gli oggetti digitali e persone tramite tecnologie di VR e AR. Di qui il Metaverso emerge come una rete interconnessa di ambienti sociali, immersivi e connessi che si basa sulla multi-utenza” [94]. In altre parole, il Metaverso è un ambiente virtuale dove più utenti interagiscono, ovviamente tramite i rispettivi *avatar*, connettendosi ad uno spazio comune a prescindere dalla loro distanza nel mondo reale. Nel Metaverso le persone interagiscono, sia con l'ambiente stesso sia tra di loro, pertanto la comunicazione nel mondo virtuale torna ad essere un argomento di fondamentale importanza: renderizzarla in modo realistico e verosimile permetterà di avere conversazioni virtuali più realistiche e rappresenta quindi una sfida-chiave al centro dei prossimi sviluppi futuri del Metaverso [95] [96]. La Figura 2.7 mostra la complessità del Metaverso e delle sue componenti, tra le quali è presente anche la componente sociale su cui ci si è soffermati nella precedente sezione.

In conclusione, il progetto della presente tesi si inserisce quindi perfettamente nel quadro dello stato dell'arte esposto in questo capitolo: dall'uso e dall'aspetto dell'*avatar*, dalle tecnologie usate per animare il suo corpo e il suo viso per renderlo espressivo e infine con l'aspirazione che possa essere un progetto in linea con i prossimi sviluppi futuri sui mondi virtuali e sul Metaverso.

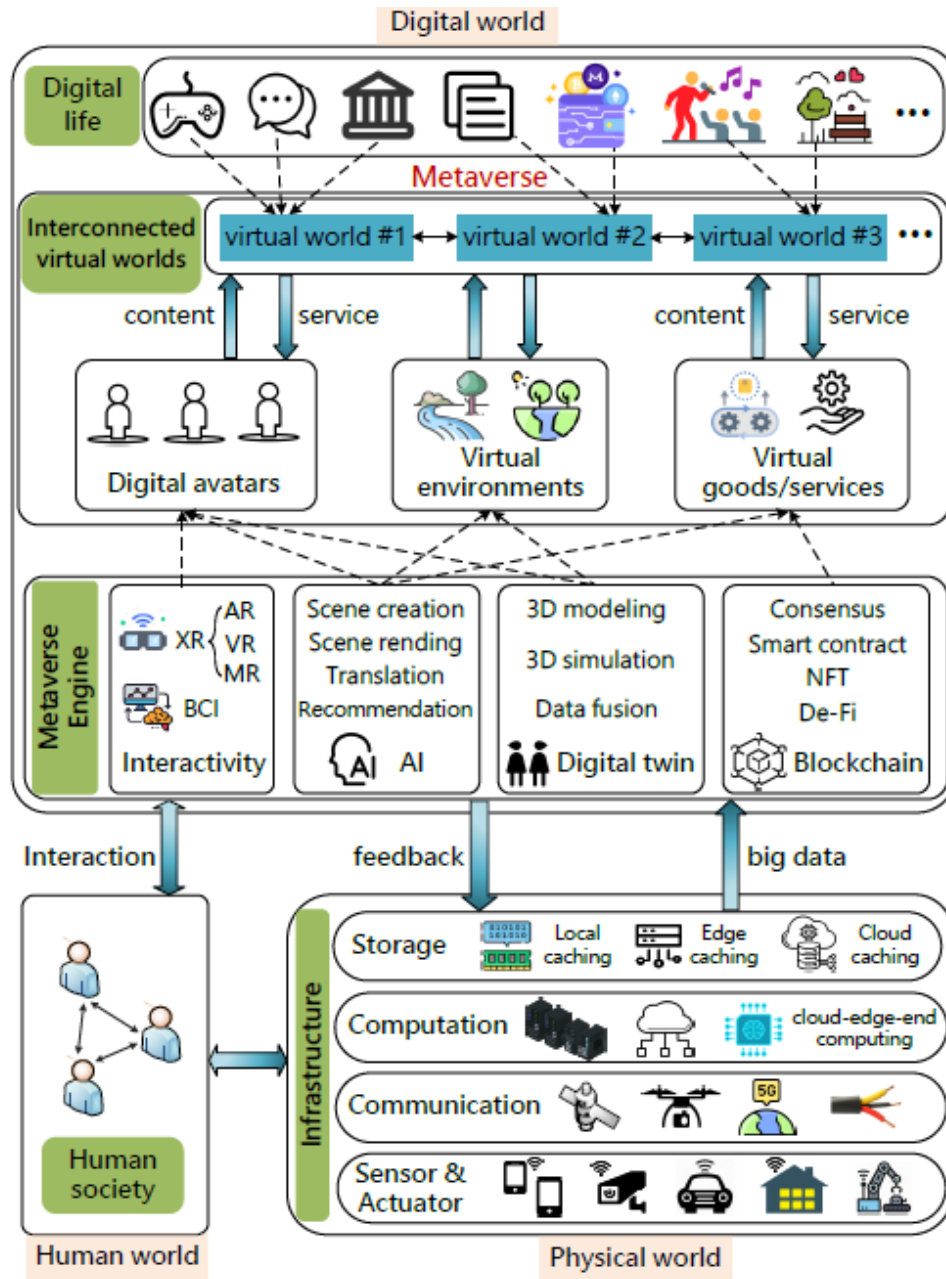


Figura 2.7: Uno schema generale della struttura del Metaverso e delle sue singole componenti [95].

Capitolo 3

Realizzazione: Primo studio

3.1 Idea di base

In questo capitolo della presente tesi verrà illustrato più nel dettaglio l'esperimento, o caso d'uso, scelto per questo primo studio, spiegando anche le tecnologie e le modalità usate per realizzarlo.

Come già accennato nell'introduzione, il presente caso di studio deriva da un lavoro precedente che si è voluto migliorare e aggiornare in ottica rappresentazione *avatar*. Dopo un attento ragionare, è stato deciso di mantenere la linea generale del progetto di base [7] in ottica *Serious Game* ed esperienza educativa per la gestione delle emergenze e per l'addestramento di civili per situazioni critiche come quella raffigurata nel progetto. Allo stesso tempo, per non lasciare tutto invariato e con l'intenzione di tenere aggiornato il caso di studio con gli sviluppi fatti in questo studio della presente tesi, si è deciso di ampliare un po' l'esperienza virtuale, aggiungendo alcuni ostacoli e *task* affrontabili con le nuove rappresentazioni introdotte, mantenendo comunque la plausibilità e la vicinanza con la realtà, e rimuovere quegli elementi che magari erano diventati più secondari per il nuovo obiettivo dello studio, che verranno dettagliati maggiormente qui nel seguito.

Verranno ora trattate e riepilogate tutte le tecnologie e le modalità che si sono ritenute necessarie per applicare i miglioramenti previsti e permettere quindi l'idea sopra menzionata.

3.2 Punto di Partenza

In questa sezione verrà fatta una breve panoramica sul progetto originale, menzionando le tecnologie e le funzionalità più importanti che erano state implementate e che sono state mantenute nel presente progetto di studio.

3.2.1 Struttura generale

L'applicazione è stata creata tramite il *Game Engine* Unity, versione 2018.4.36f. La scelta di questo software, tra le varie alternative disponibili sul mercato, è stata dovuta principalmente alle innumerevoli funzionalità offerte dal suo *Asset Store* e soprattutto dall'ampia gamma di dispositivi supportati, rendendolo quindi uno strumento molto flessibile per la progettazione. È stato usato inoltre lo *SteamVR Software Development Kit* (SDK), che garantiva il funzionamento dell'applicazione su qualsiasi hardware di realtà virtuale compatibile con OpenVR.

A causa dell'elevata scala dell'ambiente virtuale in cui è possibile muoversi, nel progetto originale sono state incluse anche modalità di locomozione stazionaria [26] per abbattere i limiti imposti dalla dimensione della stanza nel mondo reale, di certo non paragonabile a quella di un tunnel stradale. Tra queste, è stato deciso di mantenere la tecnica dell'*arm Swinging*, in quanto non richiedeva hardware o sensoristica addizionale e si è già mostrata superiore alle altre in alcuni studi precedenti [97] [98].

3.2.2 Multi-user & Networking

Per quanto concerne la parte multi-utente, è stato scelto di usare l'*Application Programming Interface* (API) ad alto livello U-NET fornita da Unity per creare una struttura di rete di tipo *client-server*. Come *host* è possibile scegliere una delle due macchine che ospiteranno un utente oppure una macchina a parte, nel caso si voglia ridurre il carico computazionale sui due *client* che manderanno in esecuzione l'esperienza. Per la comunicazione nello spazio virtuale è stato implementato un sistema VOIP, grazie al *package* Dissonance Voice Chat presente sull'*Asset Store* di Unity¹.

Per quanto concerne la sincronizzazione tra le macchine collegate, sono stati selezionati alcuni oggetti che avevano la necessità di essere sincronizzati tra i due utenti, impostando che la loro posizione e la rotazione si sincronizzasse con una frequenza di 60Hz, interpolando matematicamente tra due *frame* successivi così da rendere il tutto più fluido.

3.2.3 Inverse Kinematics (IK)

Lavorare su Unity offre la possibilità di inserire comodamente svariati *package* che aggiungono funzionalità al progetto che si sta sviluppando. Tra questi c'è anche FinalIK², un *Plugin* che permette di associare ad un *avatar* un algoritmo risolutore di IK. Una volta importato nel progetto, sarà necessario andare a prendere i vari *script* offerti dal software che corrispondono ai requisiti di progetto. Nel caso di questa tesi, verranno usati gli *script* di tipologia VR IK, ossia IK in Realtà Virtuale.

¹assetstore.unity.com/packages/tools/audio/dissonance-voice-chat-70078

²assetstore.unity.com/packages/tools/animation/final-ik-14290

Il passo successivo consiste nell'associare gli *script* all'*avatar*, predisponendo correttamente le ossa del suo scheletro nei parametri degli *script*, in particolare lo script VR IK, e impostare anche i target delle mani e della testa, che corrisponderanno rispettivamente alle posizioni dei *controller* e del visore di VR. A questo punto, già indossando l'HMD e impugnando i *controller*, l'*avatar* si muoverà in modo simile a come ci muoviamo noi nel mondo reale.

3.3 Innovazioni introdotte

In questa sezione verranno trattati più nel dettaglio tutti gli aggiornamenti e i miglioramenti introdotti rispetto allo studio originale, specialmente sul lato degli *avatar* e sulla loro rappresentazione.

3.3.1 Avatar - Movimenti

Gli *avatar* del progetto originale sono stati creati tramite Adobe Fuse e poi importati su Mixamo³ e animati tramite le animazioni presenti su tale sito. Visto che il livello di realismo puramente estetico di tali *avatar* non era basso, è stato deciso di mantenerli, tuttavia aggiornando in linea generale il loro insieme di animazioni di camminata e di corsa con un insieme più realistico e accurato.

Poichè inoltre il movimento all'interno dello spazio virtuale è stato fatto prevalentemente tramite *Arm Swinging*, si è implementata una logica che in pratica individua quando l'utente sta applicando l'*Arm Swinging* e in tale condizione spegne il componente di VR IK relativo alle mani precedentemente spiegato, lasciando che le braccia vengano animate puramente dalle animazioni impostate nel proprio *Animator Controller*. Questo è stato necessario perchè altrimenti, se fosse rimasto acceso l'algoritmo di IK durante l'*Arm Swinging*, dove l'utente fa freneticamente su e giù con le braccia, l'algoritmo avrebbe continuato a risolvere le articolazioni delle braccia continuamente in accordo ai movimenti reali, che non erano affatto realistici per una corsa. Al fine di aumentare quindi il realismo complessivo, qualora l'utente avrebbe premuto i pulsanti dei *grip* sui *controller* avrebbe attivato la modalità di *Arm Swinging*, e in tale modalità il suo *avatar* si sarebbe comunque mosso in modo realistico sulla base di un'animazione e non più dell'algoritmo di IK. Una volta rilasciati quindi i *grip*, l'IK veniva quindi riattivata e mappata con la posizione delle mani dell'utente nel mondo reale.

3.3.2 Avatar - Logica Animazioni

Oltre alle modifiche sopra enunciate, è stata modificata e potenziata la logica dell'*Animator Controller* collegata agli *Avatar*, espandendola da un semplice *Blendtree* in due livelli ad un *blendtree* a tre livelli, sulla base dell'altezza in cui la testa

³<https://www.mixamo.com>

dell'utente si trovava rispetto alla posizione base.

Il parametro dell'altezza, preso ad ogni istante di tempo, influenzava il *blendtree* più esterno, che si occupava dello scegliere quale *blendtree* secondario fosse più opportuno usare. Se l'altezza corrente era molto vicina a quella di base, allora si procedeva nel *blendtree* contenente le animazioni in posizione “eretta”, mentre via via che ci si allontanava dall'altezza base veniva coinvolto un secondo *blendtree* contenente stavolta varie animazioni di corsa e camminata però in posizione “accovacciata”. Con questa logica, mostrata in Figura 3.1, il *blendtree* più esterno interpolava costantemente, sulla base dell'altezza corrente dell'utente, tra i *blendtree* secondari, i quali a loro volta interpolavano tra di loro le animazioni direzionali che contenevano sulla base della componente direzionale della velocità dell'utente e del suo orientamento.

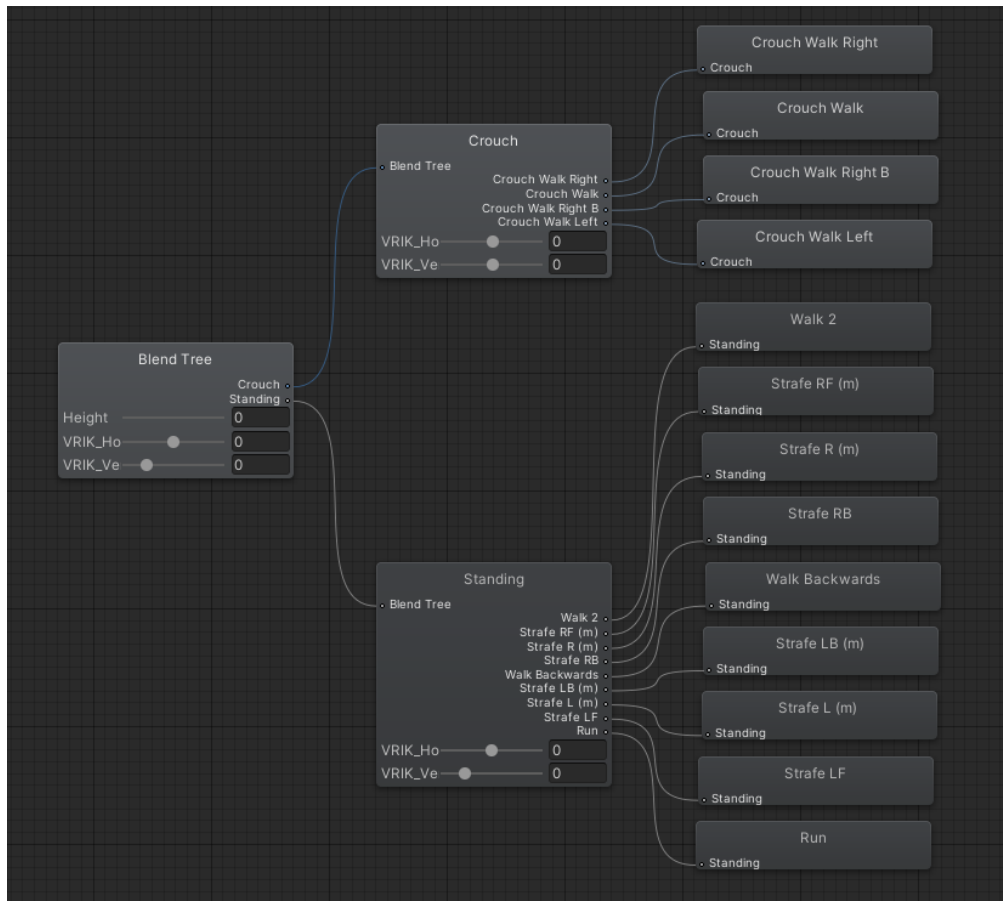


Figura 3.1: Schema dei *blendtree* per il movimento realistico degli *avatar* nell'esperienza virtuale.

3.3.3 Avatar - Mani

Per animare le mani è stata seguita una logica simile alla precedente, tuttavia in questo caso si è agito nella parte “a stati” dell’*Animation Controller* e non nei *blendtree*. L’intenzione era che le mani si animassero sulla base delle interazioni che l’utente faceva nel mondo virtuale: se l’utente premeva un pulsante, la mano doveva entrare nello stato del “puntare” con l’indice, se invece doveva prendere il telefono la sua mano doveva chiudersi a pugno, e così via. Per realizzare questo sistema è stato creato un *Animation Controller* ad hoc, con la struttura a stati mostrata in Figura 3.2. Come si può osservare, ogni stato è stato collegato con tutti gli altri, così che da ognuno era possibile transizionare su qualsiasi altro senza processi intermedi. Le transizioni da uno stato all’altro sono state gestite tramite operatori booleani modificati opportunamente dagli *script* in fase di esecuzione del programma.

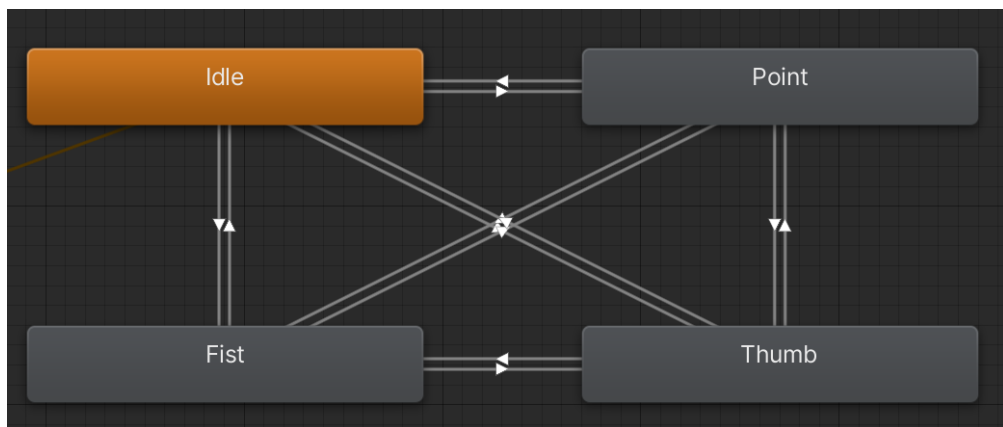


Figura 3.2: Schema degli stati dell’*Animation Controller* per il movimento realistico delle mani sulla base delle interazioni nell’esperienza virtuale.

3.4 Fase Sperimentale

3.4.1 Idea dell’esperimento

Lo scenario originale, orientato maggiormente al multi-utente e quindi comprendente una moltitudine di ruoli possibili e impersonabili nell’esperienza (come ad esempio passeggeri di un’auto, camionisti, vigili del fuoco), ognuno con il proprio insieme di obiettivi e di *task* da portare a termine, è stato semplificato da questo punto di vista, visti gli obiettivi puramente orientati agli *avatar* degli utenti e alla loro rappresentazione. In particolare, è stato deciso di limitarsi ai soli ruoli dei passeggeri dell’auto, ottenendo in questo modo un’esperienza per due utenti con lo stesso contesto e con lo stesso insieme di *task* da completare. Inoltre, il ruolo di passeggero dell’auto non richiede conoscenze pregresse, al contrario degli altri due ruoli, e

avendo la stessa situazione e partendo dallo stesso punto (all'interno della stessa auto) entrambi gli utenti hanno modo di osservarsi a vicenda e interagire più spesso tra loro rispetto ad una configurazione differente, offrendo al contempo più elasticità e flessibilità per l'esecuzione della procedura.

In aggiunta, proprio per non compromettere la visibilità degli utenti all'interno dell'esperienza e/o deviarli dagli obiettivi effettivi da completare nella stessa, si è deciso di disabilitare il fumo, realizzato tramite *Computational Fluid Dynamics* (CFD) e riarrangiare alcuni elementi dell'esperienza originale, come ad esempio limitare i ruoli di alcuni *Non-Player Characters* (NPC) solamente a comparse estetiche all'interno dell'esperienza (come ad esempio il camionista che viene trovato seduto all'interno di uno dei rifugi alla fine dell'esperienza).

Sulla base di quanto detto finora, si riassume quindi l'esperienza complessiva pianificata nei seguenti passaggi:

1. Entrambi gli utenti, rappresentati da *avatar* civili, iniziano l'esperienza all'interno della stessa vettura, in viaggio dall'Italia verso la Francia, con la radio accesa che trasmette gli usuali messaggi per chi si trova all'interno del tunnel del Frejus.
2. Dopo un certo intervallo di tempo, dove l'auto continua il suo viaggio nel tunnel e i due utenti possono parlare tra loro, un camion in fiamme fermo compare nella carreggiata opposta a quella di viaggio. Da questo momento, il passeggero può accedere tramite *controller* alla brochure di sicurezza data dal personale del tunnel all'ingresso che mostra le normative da seguire in caso di incendio, mentre il guidatore può attivare i freni dell'auto sempre tramite i propri *controller*, scegliendo a quale distanza dall'incendio far fermare l'auto.
3. Appena l'auto è completamente ferma, gli utenti possono interagire con le parti interne dell'auto, in particolare con le chiavi nel cruscotto, il pulsante dedito alle luci di emergenza e le maniglie delle portiere, da cui possono uscire dall'auto.
4. Una volta fuori, i due utenti possono iniziare a collaborare per seguire le normative di emergenza, come attivare i bottoni SOS nelle vicinanze per segnalare il problema oppure interagire con gli strumenti presenti nelle alcove laterali del tunnel (telefono, estintore, ecc.). Nel frattempo, un altro incidente è successo alle spalle del veicolo iniziale, di fatto ostacolando il raggiungimento di uno dei due rifugi più vicini (l'altro è già ostacolato dal camion in fiamme). In questo modo entrambe le vie verso un rifugio saranno ostacolate da veicoli che vanno a fuoco, seppur con diversa entità.
5. Nel caso in cui i due utenti optino per raggiungere il rifugio più vicino, però dietro il camion in fiamme, possono scegliere se provare a spegnerlo con l'estintore fornito nelle alcove laterali al tunnel oppure correre direttamente verso il rifugio. In ogni caso, dietro il camion gli utenti troveranno alcune assi

cadute dal carico del camion che hanno parzialmente bloccato il cammino, superabili accovacciandosi e passandoci sotto. Se invece gli utenti optano per allontanarsi dal camion e dirigersi invece verso il rifugio dalla parte opposta, troveranno l'incidente tra vetture precedentemente menzionato che blocca la via. In particolare, a lato delle due auto ci sarà uno stretto passaggio parzialmente bloccato da una sbarra metallica caduta dal soffitto del tunnel, superabile anche questa tramite accovacciamento passandoci sotto. In questa direzione l'estintore più vicino è interno al rifugio, quindi eventualmente, se gli utenti lo ritengono adeguato, possono prelevare l'estintore e tornare verso l'incidente e cercare di domare le fiamme. All'interno del rifugio, gli utenti possono interagire con un armadietto che contiene acqua e un kit di primo soccorso, con cui volendo possono offrire assistenza ad un NPC lì presente.

6. In ambo i casi, una volta all'interno del rifugio e aver chiesto aiuto tramite bottone SOS o col telefono, la simulazione termina.

3.4.2 Ambiente

In questa sezione tratteremo brevemente lo scenario dell'esperienza, descrivendolo a grandi linee e fornendo alcune immagini per permettere al lettore di immaginarselo correttamente.

Per il caso di studio in sé è stata ripresa l'ambientazione dello studio originale, con opportune modifiche già menzionate nella precedente sezione. In particolare, è stato aggiunto l'incidente alle spalle del veicolo degli utenti e il carico del camion disperso sulla strada nell'altra direzione. Queste modifiche sono state apportate così da avere una motivazione verosimile per testare l'estensione dell'animatore degli avatar dell'esperienza, che appunto comprende animazioni di accovacciamento.

Per il resto, lo scenario originale è rimasto invariato, con tutte le funzionalità e gli elementi che già possedeva. Un'immagine esplicativa tuttavia semplificata dello scenario può essere osservata in Figura 3.3.

3.4.3 Attività e procedura

Ora verranno enunciate le attività e la struttura dell'esperimento, usando gli elementi e le logiche precedentemente discusse.

Poiché il fine ultimo dello studio è confrontare le due modalità di rappresentazione di un *avatar* virtuale, è stato deciso di far fare agli utenti l'esperienza due volte, una in una modalità e l'altra nella corrispettiva (*Full Body* e *VR Kit*), il cui ordine veniva selezionato casualmente. In Figura 3.4 è possibile vedere la rappresentazione del proprio *avatar* e di quello dell'altro utente nelle due modalità.

Come prima cosa, ai partecipanti è stato inizialmente chiesto di compilare un piccolo questionario demografico mirato a conoscere alcuni dati anagrafici come l'età o il genere, ma anche l'esperienza pregressa dell'utente con la VR e altre tecnologie o componenti usati nello studio. Una volta compilato, ad ogni partecipante è stato

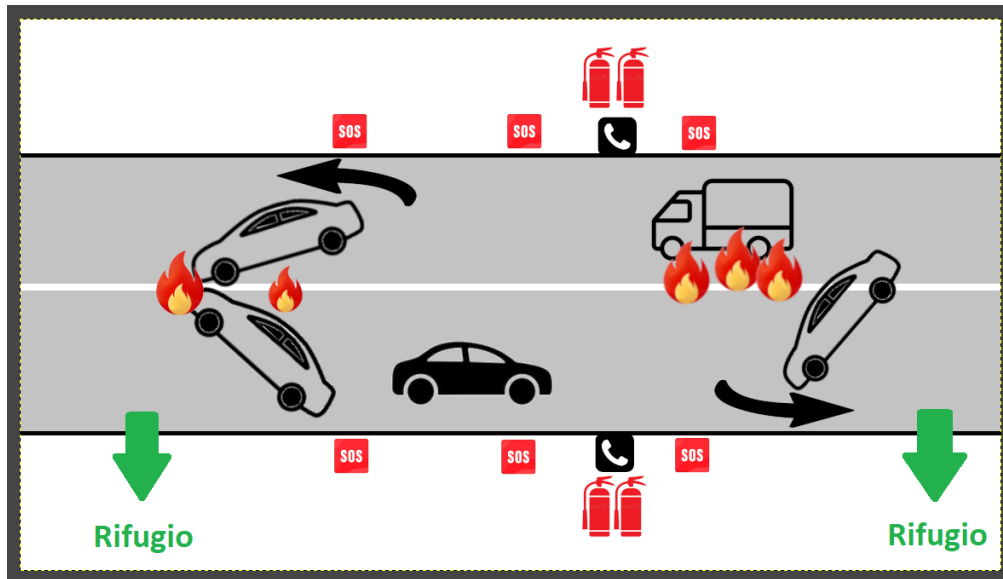


Figura 3.3: Ambiente semplificato della versione modificata del progetto originale, con i nuovi ostacoli prima e dopo l'auto degli utenti (in nero). Le varie icone indicano la posizione degli oggetti interagibili, il fuoco e infine le frecce indicano i passaggi da superare in posizione accovacciata. Per lo scenario originale più dettagliato si rimanda invece a [7]

raccontato l'esperimento a grandi linee, spiegando cosa avrebbero dovuto fare una volta dentro l'esperienza.

Prima di iniziare ognuna delle due esperienze previste, ogni partecipante ha selezionato, dall'interno dell'applicazione virtuale, uno dei due avatar disponibili per il ruolo di civile, uno di genere maschile e uno di genere femminile. Dopo la selezione, agli utenti veniva mostrato davanti a sé l'avatar scelto, in modo simile ad uno specchio, così da vedere come sarebbero stati visibili dall'altra persona nella



(a) Modalità *Full Body*.



(b) Modalità *VR Kit*.

Figura 3.4: Le due rappresentazioni dell'avatar che si andranno a confrontare nello studio.

stessa esperienza virtuale. Ovviamente, nel caso della modalità *VR Kit* la scelta era piuttosto ininfluyente, visto che non c'era alcun corpo, ma solo l'*headset* al posto della testa e i *controller* al posto delle mani.

Una volta definito il proprio *avatar*, al partecipante veniva richiesto di selezionare come ruolo il guidatore, per poi, grazie alla funzionalità di rete fornita dall'applicazione, aspettare un secondo utente, controllato da uno degli sperimentatori, in modo che si unisse alla stessa sessione e partecipasse quindi alla stessa esperienza virtuale del partecipante. Poichè il ruolo del guidatore era già stato preso dal partecipante, il secondo utente, una volta entrato nell'applicazione, sarebbe comparso direttamente come passeggero nella stessa auto del partecipante, dando così inizio ufficialmente all'esperienza.

La presenza di uno sperimentatore è stata decisa per garantire il più possibile che tutti i partecipanti portassero a termine l'esperienza in modo equo e imparziale, sfruttando il ruolo e la conoscenza dello sperimentatore stesso per il procedere con le azioni da fare e dare quindi il contributo educativo all'esperienza. Lo sperimentatore, per tutte le esperienze, si è attenuto ad un insieme predefinito di azioni, cercando di seguirlo in tutti gli esperimenti. Visto inoltre che ogni partecipante avrebbe dovuto portare a termine l'esperienza due volte, ognuna in una particolare modalità di rappresentazione dell'*avatar*, al fine di evitare che durante la seconda esecuzione sapesse già cosa attendersi e quindi compromettere la valutazione, è stato deciso di variare leggermente le due esperienze. Più in particolare, lo sperimentatore nella prima esperienza:

1. Compiva azioni e interazioni mentre era osservato dal partecipante all'esperimento.
2. Suggeriva all'utente di dirigersi verso il rifugio più vicino, ossia quello oltre il camion in fiamme.
3. Si accertava che l'utente interagisse con le varie alcove presenti lungo il tragitto e con gli oggetti in esse contenute.
4. Mostrava all'utente come superare i principali ostacoli di questa variante, ossia l'incendio divampante e le assi di legno che bloccavano il passaggio e che richiedevano al partecipante di chinarsi per superarle.
5. Assecondava moderatamente le richieste e le iniziative del partecipante, a seconda che quanto proposto fosse in linea con le misure di sicurezza da adottare in quel contesto. Ad esempio, se il partecipante si proponeva di provare a spegnere l'incendio usando l'estintore trovato all'interno di una delle alcove, lo sperimentatore lo permetteva, prendendone uno anche lui stesso, tuttavia facendo poi, dopo un po' di tempo, desistere il partecipante dal continuare a provare a spegnerlo e a dirigersi quindi nel rifugio così da proseguire con il resto dell'esperimento.

Passando ora alla lista delle azioni compiute nella seconda esperienza, lo sperimentatore:

1. Suggestiva, una volta fuori dall'auto, di allontanarsi dal camion e dirigersi verso il rifugio più vicino in quella direzione, notando che nel frattempo è successo un altro terribile incidente dietro al veicolo di partenza.
2. Anche qui, mostrava all'utente come aggirare in modo sicuro le due vetture coinvolte nell'incidente dalla parte opposta del camion, cercando di farlo stare lontano dal fuoco potenzialmente pericoloso. Guidava inoltre l'utente al superamento della sbarra metallica tramite l'accovacciarsi, oltre ad eventuali altri ostacoli.
3. Incoraggiava l'utente a correre più velocemente (vista la distanza maggiore a cui era posto il rifugio) simulando una sensazione di pericolo e l'urgenza di allontanarsi il prima possibile dall'incidente in fiamme ormai alle loro spalle.
4. Dopo aver raggiunto finalmente il rifugio, proponeva stavolta attivamente all'utente di prendere un estintore e provare ad estinguere il piccolo incendio divampato a seguito dell'incidente delle due vetture. Anche in questo caso, dopo alcuni secondi passati a provare a spegnerlo, suggeriva all'utente di lasciar perdere e di tornare all'interno del rifugio ed aspettare i soccorsi, soccorrendo eventualmente l'NPC lì presente con acqua o kit di pronto soccorso.

In Figura 3.5 sono mostrati alcuni estratti dell'esperienza.

In ogni caso, le linee guida e le tracce sopra descritte non sono state considerate in modo stretto e seguite alla lettera; vanno più intese come una sorta di copione indicativo da seguire, eventualmente tralasciando alcuni passaggi a seconda dello spirito di collaborazione del partecipante all'interno dell'esperienza.

Dopo ognuna delle due esperienze, al partecipante veniva richiesto di compilare un questionario, descritto più nel dettaglio nella sezione seguente.

3.4.4 Questionari e valutazione

Per gli obiettivi preposti dello studio in esame, i partecipanti hanno valutato l'esperienza con un questionario post-esperienza da rispondere da un punto di vista più soggettivo. Ciò è stato deciso perchè ciò che si è proposti di fare è valutare quale configurazione dell'*avatar* un utente medio preferisca e quali effetti, sia positivi che negativi, essa preclude negli altri aspetti che intercorrono in un'esperienza virtuale. Non è listata come sezione a parte, tuttavia è presente comunque un piccolo insieme di domande demoscopiche al fine di avere qualche informazione sul partecipante specifico, come ad esempio genere, età ed esperienza con la VR.

Il questionario è composto da quattro sezioni principali:

- **Sezione I:** Nella prima sezione è stato scelto di mettere un questionario volto soprattutto a valutare l'impersonazione nel proprio *avatar* nell'esperienza



(a) Utenti all'interno dell'auto iniziale.



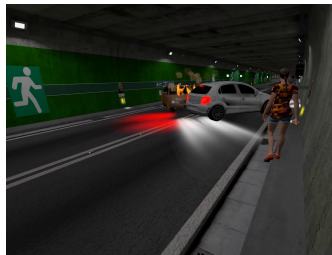
(b) Usciti dall'auto, con le luci di emergenza e il camion in fiamme in lontananza.



(c) Interazione con bottone SOS.



(d) Interazione con il telefono per chiamare i soccorsi.



(e) Incidente avvenuto alle spalle dell'auto iniziale.



(f) Incidente avvenuto nella zona del camion in fiamme.



(g) Passaggio stretto che richiede l'accovacciamento degli utenti.



(h) Interazione con l'estintore.



(i) Interazione finale con l'NPC camionista.

Figura 3.5: Alcune immagini dell'esperienza virtuale multi-utente.

virtuale. Per fare ciò, è stato scelto di usare l'*Embodiment Questionnaire* [99], che verte tra le varie cose su fattori come possesso del proprio corpo virtuale, controllo dei movimenti del proprio avatar o sensazioni tattili percepite.

- **Sezione II:** La seconda sezione è stata riservata ad un questionario più orientato verso l'esperienza multi-utente e la percezione degli altri utenti nello stesso ambiente condiviso, in particolar modo in termini di mutua coscienza, comprensione e assistenza, tra le altre cose. Le domande offerte dal *Networked Minds Social Presence Questionnaire* [100] si sono rivelate consone agli obiettivi e pertanto sono state proposte ai partecipanti nel post-esperienza, eccezion fatta per una categoria di domande, originariamente

inclusa nel questionario, che non è stata ritenuta adatta al caso di studio qui in discussione: tale categoria riguardava domande relative all'empatia con gli altri utenti dell'ambiente virtuale, che seppur importante nella sua particolarità non è propriamente fulcro di questo lavoro di tesi.

- **Sezione III:** All'interno della terza sezione è stato deciso di inserire parte del questionario *VR-Use* [101], in particolar modo le sezioni riguardanti l'immersione e la presenza percepite nell'esperienza virtuale da parte degli utenti.
- **Sezione IV:** Infine, nell'ultima sezione, sono state proposte alcune domande dirette di confronto, da noi concepite, atte a far in modo che il partecipante potesse esprimere una propria preferenza tra le due tecniche di rappresentazione di *avatar* proposte nello studio.

E' possibile esaminare più nel dettaglio tutte le domande a cui hanno risposto i partecipanti allo studio nel post-esperienza all'interno dell'Appendice A.1.

Capitolo 4

Risultati: Primo studio

4.1 Partecipanti

Per quanto riguarda l'insieme dei partecipanti che sono stati scelti per sperimentare l'applicativo, si hanno 15 individui (14 uomini, 1 donna) di età compresa tra 24 e 67 anni. La maggior parte di essi, nella parte demoscopica, ha risposto di possedere un'esperienza medio-alta con videogiochi, VR e applicazioni multi-utente, tuttavia pochissimi di loro possedevano una qualche esperienza in ambito *Serious Games*, in particolar modo per scenari collegati ad emergenze.

4.2 Risultati

In questa sezione verranno enunciati i risultati ottenuti dalle risposte ai questionari precedentemente definiti, analizzati statisticamente in modo da trovare correlazioni tra i dati. Le domande sono state ripetute sia quando l'utente ha vissuto l'esperienza con l'*avatar Full Body*, sia con l'altra modalità, ovvero con il *VR Kit*, così da avere due insiemi di risposte analoghi e perfettamente confrontabili.

Sui dati in sè, in termini più statistici e tecnici, è stato usato il test Shapiro-Wilk per analizzare la normalità dei dati, visto il numero relativamente stretto dei partecipanti. Tuttavia, poichè i dati sono risultati essere non normalmente distribuiti, è stato deciso di optare per l'uso del test non parametrico con segno Wilcoxon, con una significatività del 5% ($p < 0.05$), per studiare le differenze statistiche nei campioni posseduti.

4.2.1 Embodiment

Per ciò che riguarda la Sezione I dei questionari, ossia quella dedicata all'impersonazione del proprio *avatar*, i risultati sono riportati in Figura 4.1.

Come si può notare osservando l'istogramma, le differenze tra le due modalità esaminate non sono sostanziali, seppur mostri una leggera superiorità complessiva della modalità *Full Body*. Ci sono tuttavia differenze statisticamente rilevanti e

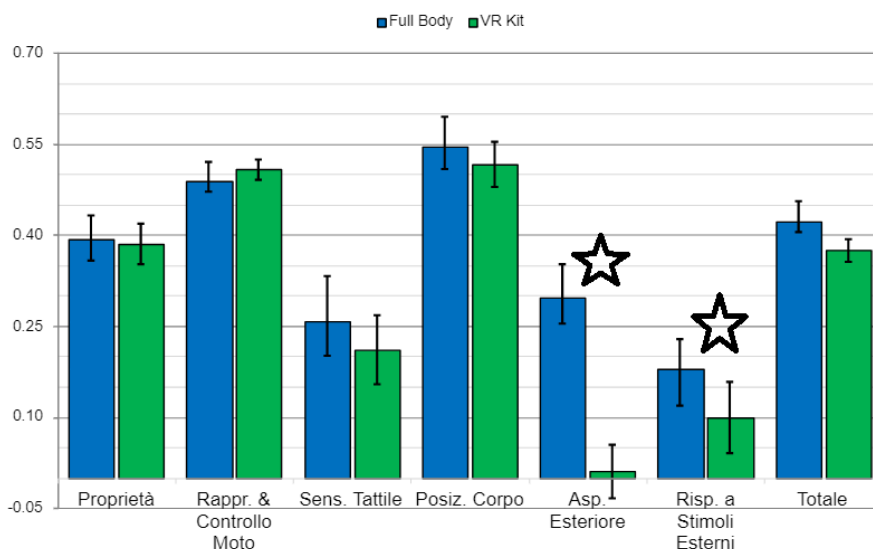


Figura 4.1: Risultati medi (con normalizzazione a 1) delle categorie del questionario di *Embodiment* [99]. I valori statisticamente significativi sono marcati con un simbolo a stella.

significative per due delle sottoscale del questionario, in particolar modo per quanto concerne l'aspetto esteriore dell'*avatar* (0,3 vs 0; $p < 0,001$) e la risposta a stimoli esterni (0,18 vs 0,1; $p = 0,003$), dove in entrambi i casi la modalità *Full Body* si è mostrata superiore della controparte *VR Kit*. Questa serie di risultati grossomodo equilibrati si allinea perfettamente allo stato dell'arte già menzionato nei precedenti capitoli, dove è stato visto che non ci sono differenze significative tra il mostrare o meno l'*avatar* dell'utente. quando è in prima persona [102].

Procedendo ad una analisi più nel dettaglio delle singole voci, emerge che i partecipanti hanno preferito come rappresentazione del proprio corpo la modalità dell'*avatar Full Body* (1,4 vs 0,27; $p = 0,042$) tuttavia, allo stesso tempo, si è percepita maggiormente la sensazione di avere più di un corpo (0,73 vs -0,6, $p = 0,01$), e ciò riflette i risultati di altri studi recenti che indicano che avere un *avatar Full Body* può avere conseguenze sia positive che negative in termini di impersonamento, specialmente in base a quanto sia fedele e accurato [103]. Ciò è notevole anche nelle altre risposte, come ad esempio quella che mostra come, in modalità *Full Body*, i partecipanti hanno sentito come se i movimenti virtuali del proprio *avatar* influenzassero i loro movimenti reali (-0.13 vs -1.13; $p = 0,0488$), rispetto ovviamente alla controparte *VR Kit*.

Riferendosi alla parte iniziale dell'esperienza, dove ad ogni utente viene mostrato davanti a sé il proprio *avatar*, come se fosse davanti ad uno specchio, la modalità *Full Body* sembrerebbe essere stata percepita come il proprio corpo più della corrispettiva modalità *VR Kit* (0,73 vs -0,6, $p = 0,0412$). Inoltre, secondo i partecipanti, con la modalità *Full Body* si è avuta una maggiore sensazione che il proprio corpo reale stesse diventando un corpo "avatar" (0,8 vs -0,6, $p = 0,003$), che sentivano che il

loro corpo reale iniziava a cercare di assumere la posizione del corpo virtuale (0,2 vs -1,13, $p = 0,009$), che ad un qualche punto dell'esperienza la loro rappresentazione virtuale ha iniziato a somigliare maggiormente al loro corpo reale in termini di forma e altre caratteristiche visive (0,13 vs -2,0; $p = 0,003$), e che hanno avuto la sensazione che stessero indossando vestiti diversi da quelli con cui erano entrati nel laboratorio per l'esperienza (0 vs -2,26; $p = 0,005$). Infine, i partecipanti si sono sentiti più in "pericolo" con la modalità *Full Body*, in particolar modo sentendo che gli elementi virtuali potenzialmente pericolosi come il fuoco o altri oggetti potessero arrecare loro danno (1,2 vs 0,33, $p = 0,015$), e che il fuoco gli desse l'impressione di essere più dannoso (-0,07 vs -0,53, $p = 0,015$).

4.2.2 Social Presence

Passando in questa sezione ad esaminare i risultati ottenuti dal questionario *Networked Minds Social Presence*, si presenta una prima panoramica in Figura 4.2. Solo dall'osservare l'istogramma dato, è possibile notare come in questo caso le differenze siano nettamente più marcate rispetto al precedente.

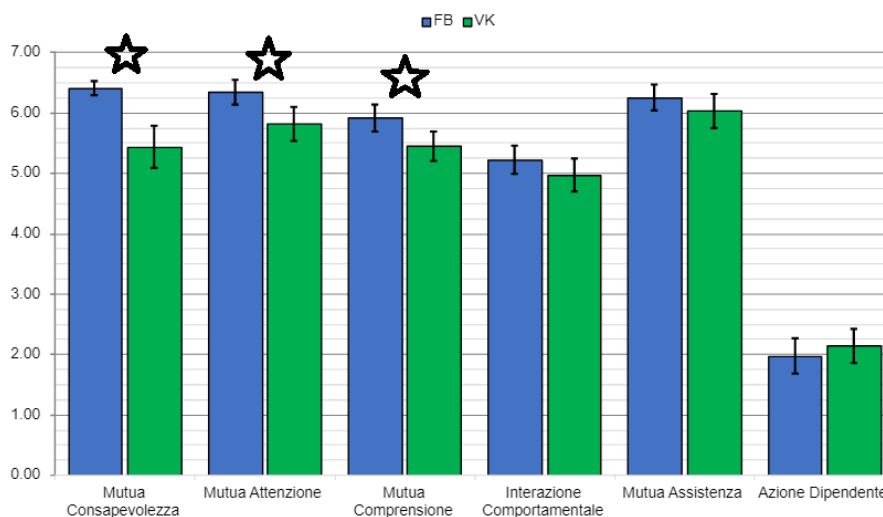


Figura 4.2: Risultati medi delle categorie del questionario di *Networked Minds Social Presence* [100]. I valori statisticamente significativi ($p < 0,05$) sono marcati con un simbolo a stella.

La modalità *Full Body* è stata ritenuta migliore in termini sia di mutua consapevolezza (6,4 vs 5,43; $p = 0,003$), sia di mutua attenzione (6,34 vs 5,81; $p = 0,003$), e sia di mutua comprensione (5,91 vs 5,44; $p = 0,047$). Tale risultato sottolinea come la rappresentazione di un altro utente tramite *Full Body*, all'interno dello stesso spazio virtuale condiviso, migliori in generale la cooperazione tra utenti. Più nel dettaglio, con la rappresentazione dell'*avatar* dell'altro utente sopra menzionata, si nota come i partecipanti si sono resi conto maggiormente della presenza dell'altra persona (1,33 vs 3,0; $p = 0,001$), hanno percepito che l'altra persona fosse più

consapevole nei loro confronti (6,33 vs 5,4; $p = 0,023$) e che loro stessi fossero più consapevoli nei propri riguardi all'interno dell'ambiente virtuale (6,4 vs 4,93; $p = 0,002$).

Proseguendo oltre, i partecipanti si sono in generale sentiti meno soli (1,2 vs 1,93; $p = 0,046$), e hanno percepito l'altra persona allo stesso modo meno solitaria (1,4 vs 2,2; $p = 0,039$) con la modalità *Full Body* rispetto la controparte *VR Kit*. Avere il corpo rappresentato nella sua interezza ha inoltre aiutato i partecipanti a prestare maggiore attenzione all'altra persona (6,33 vs 5,47; $p = 0,0312$) e, al contrario, con la modalità *VR Kit* si è riscontrata una maggiore tendenza ad ignorare l'altra persona rispetto alla modalità *Full Body* (1,67 vs 2,73; $p = 0,0117$).

Infine, con la modalità *Full Body* gli utenti hanno avuto l'impressione di poter meglio esprimere le proprie opinioni (6,00 vs 5,27; $p = 0,0027$), e allo stesso modo capire le opinioni dell'altra persona presente nell'esperienza con loro (5,93 vs 5,13; $p = 0,04888$).

4.2.3 Immersione e Presenza

Per quanto concerne la Sezione III del questionario post-esperienza, i risultati sono mostrati in Figura 4.3.

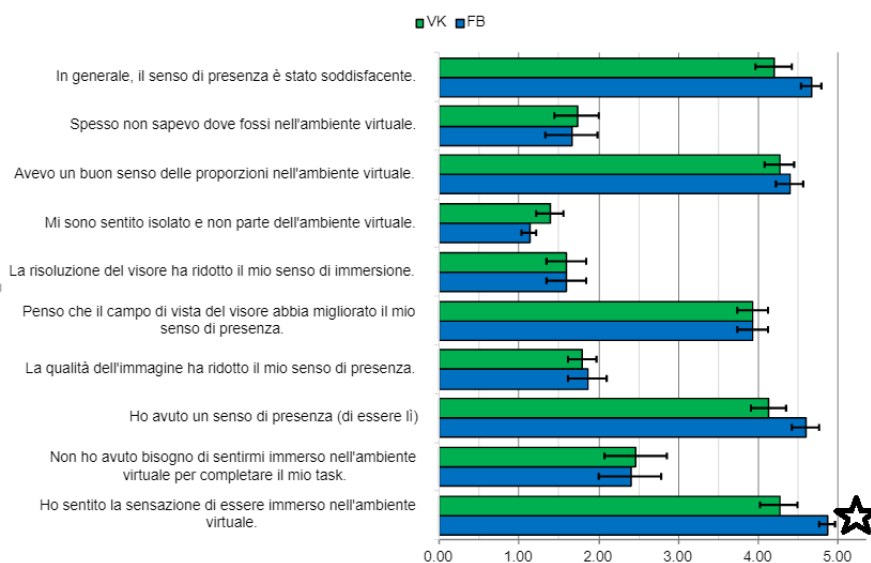


Figura 4.3: Risultati medi delle categorie del questionario di *VR Use* [101]. I valori statisticamente significativi ($p < 0,05$) sono marcati con un simbolo a stella.

Si noti come la modalità *Full Body* sia stata ritenuta migliore statisticamente parlando della controparte *VR Kit* solo per quanto concerne l'immersività nell'esperienza (4,86 vs 4,26; $p = 0,031$), lasciando di fatto poco da dire per quanto riguarda la presenza. Sulla base di alcuni commenti ricevuti dagli stessi partecipanti al termine dell'esperienza, ciò potrebbe essere dovuto al fatto che in uno scenario volto all'addestramento per emergenze come questo, avere un senso di realismo diventa

fondamentale al fine di garantire una corretta formazione e un sufficiente livello di immersione per dare il proprio meglio all'interno dell'esperienza; pertanto, tra le due modalità confrontate quella più vicina al reale è effettivamente la modalità *Full Body*, spiegando quindi la sua prevaricazione, seppur leggera, sulla controparte.

4.2.4 Domande Finali di Confronto

Si esaminerà ora la Sezione IV del questionario post-esperienza, contenente domande puramente di confronto tra le due modalità esaminate dove il partecipante ha espresso le sue opinioni relativamente a vari aspetti relativi all'esperienza virtuale appena vissuta. I risultati sono mostrati in Figura 4.4.

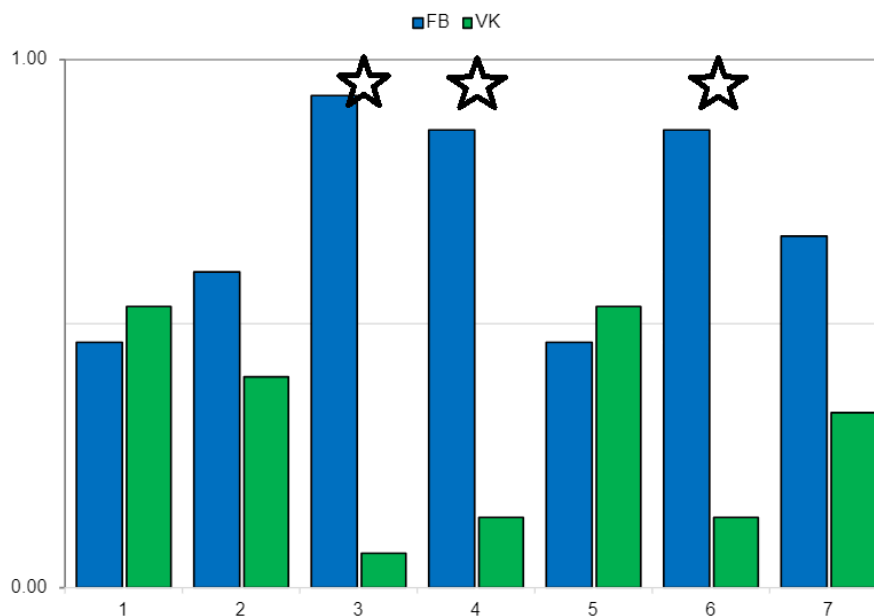


Figura 4.4: Risultati relativi alle domande dirette di confronto dell'ultima parte del questionario complessivo post-esperienza. Per le domande specifiche, si rimanda all'Appendice A.1.4. I risultati statisticamente significativi sono marcati con un simbolo a forma di stella.

Il risultato che emerge come più emblematico e interessante è che la modalità *Full Body* risulta significativamente migliore della modalità *VR Kit* per quanto concerne la rappresentazione estetica dell'altro utente (93,3% vs 6,66%; $p = 0,001$), le interazioni multi-utente (86,66% vs 13,33%; $p = 0,01$) e anche per l'esperienza complessiva (86,66% vs 13,33%, $p = 0,010$). E' però importante notare che nessuna delle due tecniche ha prevalso per quanto riguarda la rappresentazione del proprio *avatar*, e ciò è in linea con i risultati ottenuti nella Sezione I del questionario, quella relativa all'impersonazione del proprio *avatar* durante l'esperienza multi-utente.

4.3 Discussione

Ricapitolando, in questo caso di studio sono state confrontate due tecniche di rappresentazione degli *avatar* degli utenti in un contesto multi-utente di tipo educativo e gestione delle emergenze. La prima tecnica usata, usata frequentemente in applicazioni commerciali a singolo utente, consiste nel mostrare unicamente il *VR Kit*, ossia l'HMD al posto della testa e i *controller* al posto delle mani. La seconda tecnica, chiamata *Full Body*, fa invece uso di un insieme di algoritmi di IK e di *animation blending* per muovere e operare una ricostruzione completa e umanoide dell'*avatar* dell'utente, usando in particolar modo gli *end point* delle mani e della testa e i loro orientamenti come base per gli algoritmi sopra menzionati.

Per valutare le loro prestazioni è stato condotto uno studio che ha coinvolto 15 partecipanti, in modo che ognuno potesse esprimere la propria preferenza o il proprio gradimento per alcuni aspetti specifici chiesti nelle domande del questionario (Appendice A.1). I risultati sono vari, a seconda della domanda posta: per quanto concerne il proprio utente, non è emerso un chiaro vincitore riguardo l'*embodiment* o la preferenza generale, sebbene la modalità *Full Body* è stata percepita migliore della controparte *VR Kit* per quanto riguarda l'apparenza estetica, la risposta agli stimoli esterni e l'immersione. Spostandoci invece sui risultati relativi all'*avatar* dell'altro utente, la modalità *Full Body* è stata ritenuta in grado di veicolare meglio una serie di fattori di presenza sociale, quali la mutua consapevolezza, la mutua attenzione e la mutua comprensione. In ultimo, ma non meno importante, è stato mostrato che tale modalità è stata preferita rispetto al *VR Kit* per quanto riguarda la rappresentazione dell'altro utente nello stesso spazio virtuale, sia complessivamente, sia per quanto concerne estetica e interazioni multi-giocatore.

Questi risultati suggeriscono che nell'ambito particolare della simulazione delle emergenze nella Realtà Virtuale potrebbe essere più benefica una rappresentazione degli *avatar* degli altri utenti presenti nella stessa esperienza analoga alla modalità *Full Body*, purchè ovviamente nel complesso si raggiunga un livello sufficientemente credibile e realistico. Per quanto concerne invece il proprio *avatar* non sono stati riscontrati risultati chiari, e ciò probabilmente è dovuto al fatto che talvolta le articolazioni dell'*avatar* umanoide fossero un po' troppo strane e lontane dalla realtà al punto da distrarre i partecipanti dall'esperienza, come riportato da alcuni di loro. In situazioni simili, sarebbe stato preferibile non vederlo per niente a quel punto, in modo da non compromettere quel senso di presenza e di immersione che si era creato dall'inizio dell'esperienza virtuale. Questa potrebbe essere la motivazione del fatto per cui la modalità *VR Kit* si sia mostrata più valida in queste particolari situazioni.

Dalla discussione sopra riportata si evince come nel complesso, soppesando l'interesse del questionario, la modalità *Full Body* sia uscita come "vincitrice", seppur non schiacciante, dal confronto diretto con *VR Kit*. Sulla base di ciò, tali conclusioni sono state prese come ipotesi per il secondo caso di studio presentato in questa tesi, che vedrà quindi l'uso di un *avatar Full Body* e tutte le tecniche di locomozione e animazione ad esso collegate e usate nel primo caso di studio.

Capitolo 5

Realizzazione: Secondo studio

In questo capitolo verranno esposti il procedimento e le tecnologie usate nel secondo studio del progetto di tesi. Verranno discussi tutti quegli aggiustamenti e quelle impostazioni che hanno permesso a tutti i singoli componenti di funzionare tra loro e fare ciò che era richiesto.

5.1 Idea di base

L'idea del secondo caso di studio, come già ripetuto nei precedenti capitoli, nasce dove si conclude il primo. Più in particolare, nel primo caso di studio sono state confrontate due modalità di rappresentazione del corpo di un *avatar* e del loro ruolo in un contesto educativo, confronto che ha determinato un vincitore e che pertanto verrà usato come base per il secondo caso di studio. Oltre a ciò, archiviata con il primo caso di studio la parte del corpo, il secondo sarà focalizzato invece sul volto dell'*avatar* e sulla sua espressività e comunicabilità.

Sulla base delle ipotesi appena menzionate, si è pensato ad inserire o, più in generale, includere in qualche modo all'interno del progetto una componente emotiva, visto anche il *focus* sull'espressività dell'*avatar*. Sebbene questi elementi si inseriscano bene in un'ottica applicativa puramente sociale e/o metaverso, basata sulla comunicazione tra utenti e quindi sulla conseguente trasmissione di emozioni, di sensazioni e di un certo carico emotivo, lo studio è stato inoltre deciso di farlo in modalità *single player*, simulando di fatto la compagnia di un'altra persona tramite l'*avatar* che svolgerà un ruolo di NPC. Tale scelta è stata fatta puramente per lo scopo dello studio, qualitativo e soggettivo, basato sulle opinioni dei partecipanti, come verrà descritto più in seguito.

L'idea di base è che, volendo valutare quale tecnologia possieda più potenziale espressivo e che gli utenti preferiscano vedere usata sull'*avatar* altrui, non fosse necessario avere necessariamente un altro utente all'interno dello stesso spazio

virtuale, in quanto sarebbero poi sorte problematiche di come elicitare al meglio le emozioni durante la conversazione tra i due utenti e ciò avrebbe diversificato ogni esperienza perdendo di fatto una sorta di equità per tutti i partecipanti. Per ovviare a ciò, è stato deciso di usare un attore e registrare la sua performance, così che poi, in fase di test, ogni partecipante potesse assistere alla stessa registrazione e quindi dare una valutazione sulle stesse basi di quelle degli altri partecipanti.

Le successive sezioni spiegheranno quindi tutti i componenti necessari al raggiungimento di questa idea e alla realizzazione dell'esperimento di studio.

5.1.1 Struttura generale

Per la progettazione del risultato finale, è stato scelto di utilizzare l'ultima versione LTS di Unity 2021.3.11f1 al momento della stesura, predisponendo ovviamente il progetto con i *package* Unity necessari al funzionamento della VR. La scelta è stata fatta anche tenendo in conto il precedente caso di studio, cercando di omologare i due flussi di lavoro il più possibile.

5.2 Avatar

L'*avatar* è stato il primo tassello che è stato aggiunto al progetto, visto che la maggior parte delle tecnologie che verranno poi usate hanno bisogno di un *avatar* per funzionare. Tenendo a mente tutte le nozioni e le informazioni esposte nel Capitolo 2, l'*avatar* deve essere scelto in modo che soddisfi i seguenti requisiti:

- Deve essere di tipologia *Full Body*, in quanto i sensi di presenza ed immersione sono maggiori con questa tipologia;
- Deve essere realistico, evitando quindi un design infantile o poco lontano da quello di una persona reale;
- Deve possedere uno scheletro completo e un insieme di *Blendshape* facciali ben definite: entrambi i componenti sono essenziali per una corretta animazione del corpo e delle espressioni facciali; nel gergo della CG si dice che l'*avatar* deve essere dotato un *rig*.

Ci sono molteplici software di modellazione oramai disponibili, e tutti, o perlomeno la stragrande maggioranza di essi, permette di soddisfare i requisiti sopra descritti. Ai fini di questo progetto di tesi, è stato scelto di utilizzare il tool Autodesk® Character Generator, disponibile all'utente dopo l'acquisto di una licenza Autodesk® 3ds Max® o Maya®.

5.2.1 Creazione e Caratteristiche del Modello

La creazione dell'*avatar* tramite il tool offerto da Autodesk® è un processo piuttosto diretto e semplice, paragonabile a quello di creazione personaggio di molti videogiochi di ruolo recenti usciti sul mercato.

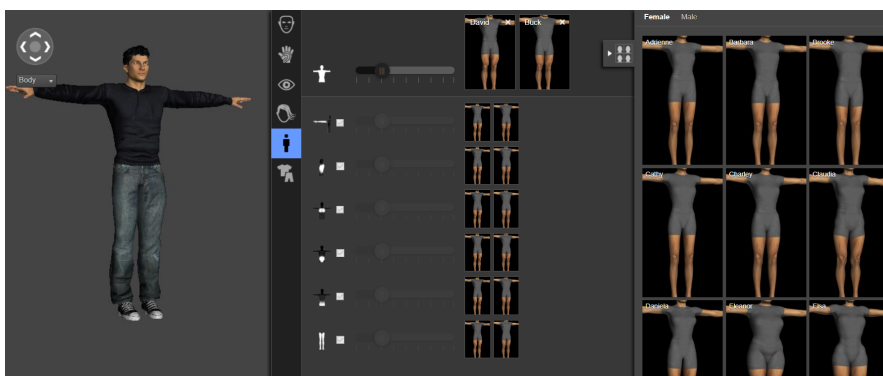


Figura 5.1: Interfaccia Grafica Utente (GUI) di Autodesk® Character Generator.

Come si può vedere nella schermata mostrata in Figura 5.1, il software offre una moltitudine di impostazioni per poter personalizzare un *avatar*. E' possibile scegliere il modello generale da un *preset* già fatto, oppure personalizzarne manualmente il colore della pelle, degli occhi o dei capelli, la struttura ossea del viso, la muscolatura del corpo e anche il vestiario. In generale non si sono poste particolari limitazioni sulla personalizzazione del personaggio, fuorchè quella già menzionate precedentemente. L'unico accorgimento è stato, visto il focalizzarsi della presente tesi sull'espressività del viso o più in generale della parte superiore del corpo di un *avatar*, lo scegliere un taglio di capelli e i vestiti il più possibile semplici, così da limitare al minimo eventuali distrazioni da parte dell'utente che lo osserverà.

Una volta terminato di personalizzare il modello, il software dà la possibilità di generarlo e di conseguenza scaricarlo per poterlo usare. Le opzioni di generazione sono mostrate in Figura 5.2.

Dopo aver scelto un nome e l'altezza del modello, vi sono una serie di opzioni per quanto riguarda geometria, texture, scheletro e risoluzione del modello dell'*avatar*. Continuando a perseguire la via dell'*avatar* realistico, sono state scelte in questo progetto le opzioni più orientate verso un'alta rappresentazione dell'*avatar*: in particolare, un'alta risoluzione, *normal* e *specular map*, *blendshape* facciali e geometria quadrangolare. Per quanto riguarda l'esportazione vera e propria, sono stati generati sia il file *.fbx* per Unity, visto che il progetto della tesi sarà sviluppato su tale *Game Engine*, sia il file *.mb* per Autodesk® Maya®, così eventualmente da poterlo modificare in caso di necessità.

Una volta esportato il modello, la parte relativa all'estetica dell'*avatar* è per il momento archiviata. Il prossimo passo sarà quindi importarlo nel *Game Engine* di Unity e iniziare ad impostare le animazioni e i vari *Script* a cui sarà soggetto.

5.2.2 Inverse Kinematics (IK)

Come ripetuto nei precedenti capitoli e anche all'inizio di questo, visti i risultati conseguiti con il primo studio è stato deciso di mantenere le logiche *Full Body* del primo esperimento, e tra queste rientra anche il *package* FinalIK. Per una

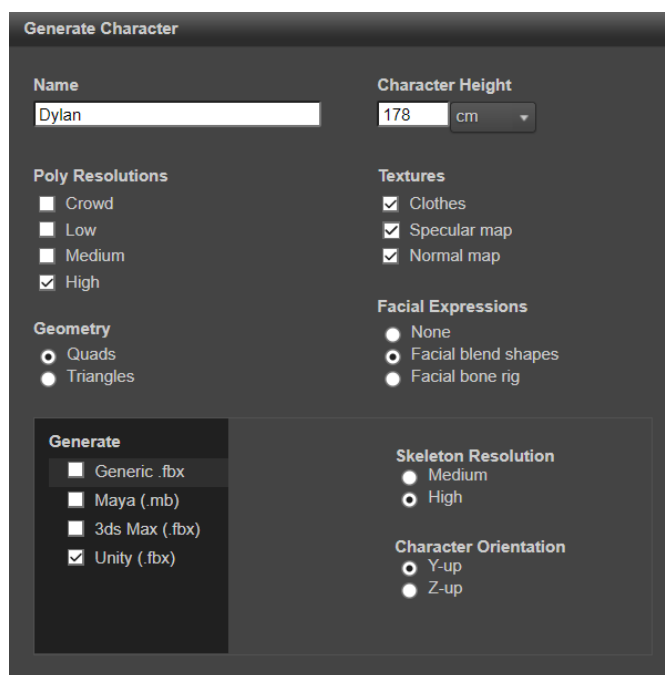


Figura 5.2: Opzioni di export dell'*avatar* generato tramite Autodesk® Character Generator.

spiegazione più dettagliata al riguardo, si raccomanda al lettore di tornare al Capitolo 3, dove è stato presentato per il primo caso di studio.

5.3 SALSA LipSync Suite

Ora che l'*avatar* e il suo movimento sono stati impostati, è il momento delle espressioni facciali e dell'animazione del volto. Poichè nell'esperimento verranno confrontate diverse modalità di rappresentazione delle emozioni o più in generale dell'espressività dell'*avatar*, ci saranno due tecnologie differenti che verranno usate per animare le *blendshape* del volto. La prima di queste che verrà trattata è quella offerta da SALSA LipSync Suite¹, progettata dallo studio indipendente Crazy Minnow².

Simple Automated LipSync Approximation (SALSA) è un sistema che permette di sincronizzare le labbra ad un file audio in ingresso in *real-time*, senza il bisogno di associare i visemi del viso al suono o fare il *baking* delle animazioni. Ciò permette di creare un procedimento semplice e lineare che in poco tempo offre risultati di sincronizzazione labbra-audio accurati e notevoli. Il formato audio in ingresso può essere di diversi tipi, da audio registrati, al microfono e persino il *text-to-speech*. Al

¹<https://assetstore.unity.com/packages/tools/animation/salsa-lipsync-suite-148442>

²<https://crazyminnowstudio.com>

suo interno usa un algoritmo procedurale e approssimativo che si basa sulle onde sonore del file in ingresso e tramite opportune tecniche automatizza il processo di movimento delle labbra per una qualche combinazione che sia verosimile e in linea con il file audio.

5.3.1 Implementazione e Adattamento all'Avatar

Anche SALSA è un apposito *package* di Unity, pertanto per includerlo basterà importarlo nella cartella del progetto stesso, in modo analogo a come è stato fatto per FinalIK nella sezione precedente. Dopo l'importazione, basterà associare all'*avatar* gli *script* relativi e per farlo basterà andare nel menù **Component** di Unity, poi **Crazy Minnow Studio**, poi **SALSA LipSync** e infine **SALSA**. Unity metterà autonomamente gli *script* necessari all'*avatar* per poter iniziare a lavorare. Seguendo un *workflow* del genere, il passo successivo consisterebbe nel definire sull'*avatar* i vari visemi che userà l'algoritmo e associarli ai fonemi che richiede SALSA, tuttavia gli sviluppatori hanno rilasciato degli speciali *add-on*, chiamati *One-Click* che se l'*avatar* rispetta opportune condizioni fanno tutto questo lavoro al posto del programmatore. Per questo progetto è stato usato l'*add-on* apposito per *avatar* generati tramite Autodesk® Character Generator, e quindi la configurazione dell'*avatar* per SALSA è stata automatizzata e pronta subito all'uso pratico. In Figura 5.3 sono riportate alcune schermate che mostrano il risultato mentre viene riprodotto un file audio parlato.

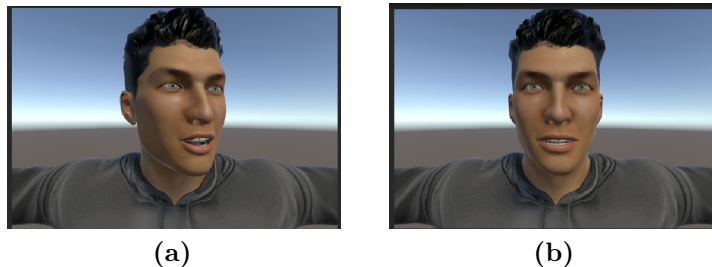


Figura 5.3: Risultato dell'applicazione all'*avatar* di SALSA LipSync Suite. Vengono mostrati due momenti diversi nel tempo mentre l'audio è riprodotto.

5.3.2 EmoteR & Eyes

Oltre al software base per la sincronizzazione delle labbra, gli sviluppatori di SALSA LipSync Suite hanno reso disponibili due utilità aggiuntive da affiancare al software principale per dare un qualcosa in più ai volti animati. Queste due utilità aggiuntive sono EmoteR e Eyes.

EmoteR è in grado di generare casualmente alcuni visemi di norma associabili a certi stati d'animo od emozioni, come ad esempio il sollevare le sopracciglia, il corruciamento, l'exasperazione di alcuni tratti facciali o il focalizzare lo sguardo.

Tutti i visemi, tecnicamente chiamati *emote*, sono personalizzabili e c'è anche la possibilità di aggiungerne, qualora lo si ritenesse necessario. Lo strumento permette anche di definire sotto quali condizioni vengono attivati e a quali tempistiche devono sottostare, quindi visualizzati sul viso dell'*avatar*, anche queste completamente definibili dal programmatore.

Eyes si focalizza invece sugli occhi, offrendo sia la possibilità di focalizzare lo sguardo dell'*avatar* su un punto specifico del suo campo visivo, sia nel generare continuamente, dopo ogni intervallo di tempo, un punto randomico in una zona definita così da simulare i piccoli, mutevoli cambi di sguardo che di norma una persona fa quando parla, aggiungendo quel pizzico di realismo in più al resto dei moduli che male non fa.

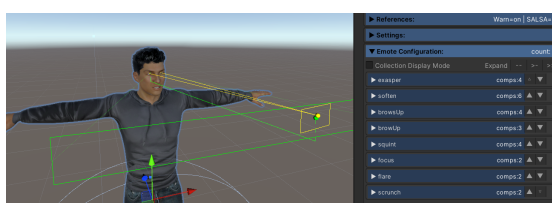


Figura 5.4: Funzionamento di EmoteR e di Eyes. I raggi gialli costituiscono lo sguardo dell'*avatar* e puntano ad un punto randomico generato nel rettangolo verde; a destra, l'elenco delle *emote* definite per l'*avatar* che si attivano randomicamente mentre l'*avatar* parla.

5.4 HTC VIVE Pro Eye and Facial Tracker

Si passa ora a trattare dell'altra tecnologia adita all'animazione delle espressioni facciali dell'*avatar*. In questo caso verrà incorporato all'HMD HTC VIVE Pro Eye un hardware della stessa casa di produzione che consiste in una sorta di videocamera ad infrarossi da connettere fisicamente all'*headset* in modo che registri la parte inferiore del viso dell'utente (Figura 5.5³). In questo modo, prelevando i dati registrati dei movimenti del volto, è possibile trasferirli sull'*avatar*, così da ottenere di fatto un tracciamento del viso vero e proprio e quindi, in linea teorica, il potere di rappresentare qualsiasi espressione che l'utente fa nel mondo reale in quello virtuale.

Da alcuni test effettuati in laboratorio, è emerso che il Facial Tracking tramite questa tecnologia perda un po' di precisione ed efficacia qualora l'utente indossi occhiali e/o abbia una barba relativamente folta. Ciò è dovuto principalmente al fatto che il tracciamento delle due parti, come vedremo poi qui di seguito, viene fatto tramite camere e tecniche di *computer vision*, e quindi eventuali ostacoli al

³<https://www.somagnews.com/>

rilevamento delle forme precise di occhi e bocca, proprio come occhiali e barba, possono influenzare sulle prestazioni complessive e sul risultato finale.



Figura 5.5: Dispositivo Head Mounted Display (HMD) HTC VIVE Pro con annesso hardware per il Facial Tracking.

5.4.1 Eye Tracking

Il tracciamento degli occhi viene effettuato da delle camere interne all'HMD che registrano i movimenti degli occhi stessi e delle sopracciglia, ossia in pratica la parte del volto che è coperta dall'*headset*.

La mappatura tra occhi reali e occhi dell'*avatar* avviene tramite uno *script*, che si occupa, come menzionato nella precedente sezione, di mappare le *blendshape* del viso reale che vede muoversi in quelle dell'*avatar* nel mondo virtuale. Oltre a ciò, gestisce anche tutto il processo relativo allo sguardo: tramite tecniche di *computer vision*, le camere interne all'*headset* riescono a trovare e tracciare la pupilla e quindi trasmettere le informazioni al software di progetto di dove si sta guardando, così da animare conseguentemente anche gli occhi e lo sguardo del corrispettivo *avatar* in modo che siano analoghi.

Lo *script* definisce anche delle curve pilota che gestiscono le animazioni relative alle sopracciglia. Tali curve verranno elaborate sulla base dei dati ricevuti dal tracciamento e quindi verrà selezionata l'enfasi opportuna del movimento delle stesse. Le curve sono personalizzabili sia tramite codice che tramite manipolazione diretta sulla curva stessa, e permettono quindi di adattarsi alle esigenze dell'utente che viene tracciato.

5.4.2 Lip Tracking

Per quanto concerne le labbra, il procedimento è pressochè simile a quello già visto per il tracciamento degli occhi. Anche in questo caso, con un opportuno *script*, viene fatta l'associazione delle *blendshape* dallo spazio HTC a quelle dell'*avatar*, così che il programma sappia quali attivare mentre traccia il viso.

In questo caso la camera che si occupa del mandare i dati non si trova all'interno del visore, ma è un hardware esterno che va attaccato al lato inferiore dell'*headset* in modo tale che punti verso la parte inferiore del volto dell'utente rimasta scoperta. In questo modo quindi vengono tracciate le labbra, la bocca, la mandibola e la lingua, e di conseguenza questi movimenti vengono mappati in corrispettive animazioni con le *blendshape* dell'*avatar*.

5.5 Recording e Playback

In questa sezione verrà trattato brevemente il *package* Unity Record&Play⁴ creato da Eli Davis.

Il *package* offre la possibilità di registrare, in fase di esecuzione, gli spostamenti e le rotazioni di vari `GameObject` definiti dall'utente, per poi riprodurre tali spostamenti in una seconda fase detta di "Playback", in modo analogo a come si registrava un film su una videocassetta e poi si premeva il tasto "Play" per riprodurlo di nuovo. Tali funzionalità si inseriscono bene nel contesto applicativo con l'*avatar*-attore che si ha in mente per questo studio, come già accennato all'inizio di questo capitolo. Tuttavia, ai fini dell'esperimento le funzionalità offerte dal *package* non erano sufficienti, pertanto sono state apportate alcune modifiche atte ad espandere le sue potenzialità. Più in particolare, è stato necessario introdurre una logica che registrasse non solo le trasformate degli oggetti ma anche le *blendshape* che su un determinato `GameObject` erano definite. Questa operazione è stata fondamentale, in quanto le animazioni del volto, come già visto nelle precedenti sezioni, si basano su *blendshape* sia con la Suite di SALSA Lip-Sync che con l'HTC VIVE Pro Eye and Facial Tracker. Per fare ciò si è cercato di mantenere la logica base degli *script* forniti dal *package*, aggiungendo ai dati salvati anche un *array* di valori `float` delle *blendshape* nel corso degli istanti di tempo.

5.5.1 Fase di Recording

La funzionalità di registrazione offerta dal *package* si basa su un sistema simile ad alcuni protocolli di rete, dove un insieme di `GameObject` si associa, o si iscrive, ad un particolare oggetto `Recorder`. Tale oggetto, d'altro canto, richiede che ogni oggetto a lui iscritto posseda una particolare componente chiamata `SubjectBehavior`: tale componente va vista come il tassello che collega un particolare oggetto al suo

⁴<https://assetstore.unity.com/packages/tools/animation/record-and-play-148294>

registratore. Qualora un oggetto fosse iscritto a due registratori, avrà due componenti **Subject Recorder**, una per registratore. E' in questa componente infatti che risiedono tutte le informazioni di associazione tra gli oggetti e un particolare registratore, informazioni essenziali per quando poi verrà fatta la riproduzione della registrazione.

Una volta avviata una registrazione, è possibile salvare anche degli eventi personalizzati durante la stessa: se ad esempio mentre si registra si vuole tenere traccia di quante collisioni hanno fatto gli oggetti coinvolti nella registrazione e quando, basta mettere la chiamata alla funzione dedicata al salvataggio degli eventi all'interno della funzione base di Unity **OnCollision**. In altre parole, salvare un evento è come marcare un particolare istante di tempo nella registrazione: è come dire "Al secondo 13 è successo un qualche evento.". La gestione poi di questi eventi registrati verrà spiegata nella sezione relativa al *playback*. Inoltre, è possibile salvare anche alcuni metadati sotto forma di stringhe in concomitanza con gli eventi, così da poterli differenziare tra loro e distinguerli in qualche modo.

Quando poi si vuole interrompere la registrazione, basterà chiamare la funzione di "Stop" che si occuperà di salvare la registrazione come un file `.asset` di tipo **Recording**.

5.5.2 Fase di Playback

Una volta che si hanno le registrazioni come file `.asset`, bisogna usare le funzionalità del *package* dedite alla riproduzioni di tali oggetti. Per farlo, bisogna innanzitutto instanziare un oggetto con una componente **Playback Behavior**, che gestirà la riproduzione mentre si è in esecuzione. Tale componente permette di chiamare funzioni che possono iniziare la riproduzione, metterla in pausa o fermarla, a seconda delle esigenze dell'utente.

La prima cosa che va gestita è l'aver degli **actor** opportuni. Più nel dettaglio, i file `.asset` della registrazione contengono sì i dati posizionali, rotazionali e delle *blendshape* degli oggetti tracciati, tuttavia non tengono effettivamente traccia dell'estetica dell'oggetto registrato. Bisogna quindi, tramite *script*, creare una funzione in grado di andare ad individuare e riassociare, nella componente **Playback Behavior** gli oggetti originali ai corrispettivi registrati così che le trasformazioni nel tempo siano coerenti con i movimenti originali.

La seconda cosa da gestire è dire al **Playback Behavior** cosa fare quando, riproducendo la registrazione, incontra un evento salvato dall'utente. Tramite una funzione apposita e grazie ai metadati salvati per distinguere i vari eventi è possibile quindi chiamare funzioni apposite a seconda dell'evento trovato. Ciò sarà utile per l'audio, come verrà spiegato nella sezione successiva.

5.5.3 Registrazione dell'audio

La registrazione dell'audio è stata gestita a parte, seppur integrata con le funzionalità del *package* che si sta trattando. Il problema alla base è fare in modo di garantire

che la registrazione delle posizioni, delle rotazioni e dei valori delle *blendshape* dei `GameObject` registrati sia contemporanea a quella dell'audio proveniente dal microfono collocato sulla parte bassa dell'HMD. Questo tassello è importante perchè per registrare le scene che interpreterà l'attore è altresì importante registrare la sua voce e ciò che dice, visto che come ribadito nei precedenti capitoli il tono della voce è uno degli aspetti principali con cui si veicolano emozioni ed espressività in generale.

Per poter quindi garantire una corretta sincronizzazione tra queste due componenti separatamente registrate è stato deciso di usare la funzionalità offerta dagli eventi descritta precedentemente. In particolare, nel momento in cui si avvia la funzione di inizio registrazione, si salva subito un evento che segnala che nel momento della riproduzione poi in quell'istante dovrà riprodurre anche l'audio. In questo modo le due registrazioni vengono riprodotte di pari passo, e quindi l'audio viene sincronizzato con le posizioni, rotazioni e le *blendshape* degli oggetti registrati mentre si evolvono nel tempo seguendo la registrazione.

5.6 Problematiche

In questa sezione verranno spiegate le principali difficoltà riscontrate durante la fase di sviluppo del progetto in esame.

5.6.1 Compatibilità avatar Autodesk® con Facial Tracker

Per poter effettuare il tracciamento del volto, gli script HTC dedicati al tracciamento del viso quantificano i movimenti rilevati dalla telecamera ad infrarossi sull'*hardware* esterno sulla base di un insieme di *blendshape*, internamente definite, comprendente movimenti di occhi, sopracciglia, bocca e lingua. Tuttavia, come è stato affermato nella sezione di creazione dell'*avatar*, un modello uscente da Autodesk® Character Generator possiede già il suo set di *blendshape*, definite da Autodesk® stesso. Prima di proseguire quindi è fondamentale confrontare i due insiemi per capire se vi è compatibilità, ossia, in altre parole, se le *blendshape* definite e richieste da HTC hanno un qualche corrispettivo in quelle definite da Autodesk®.

Dal confronto però emerge che i due insiemi sono solo parzialmente compatibili, visto che ci sono *blendshape* analoghe e quindi associabili, ma allo stesso tempo vi sono anche *blendshape* HTC che non hanno un analogo nel set di Autodesk®. Per ovviare a ciò e quindi risolvere tale incompatibilità, è stato aperto l'*avatar* su Autodesk® Maya® e sono state create le *blendshape* mancanti così da aggiungerle a quelle già presenti. Con l'incompatibilità risolta, l'*hardware* di Facial Tracking di casa HTC può ora funzionare con l'*avatar* Autodesk®.

5.6.2 Limitazioni hardware del Facial Tracker

una volta risolta la mappatura tra i due insiemi di *blendshape* della precedente lezione, sono stati portati avanti alcuni test prestazionali per valutare la qualità del rilevamento e del tracciamento dell'HTC VIVE Pro Eye e Facial Tracker.

Da questa analisi è emerso che la mappatura delle *blendshape* era accurata e precisa, e per il tracciamento della mascella, degli occhi e dello sguardo non vi era alcun problema nè di tipo estetico nè di tipo funzionale. Altri movimenti invece, come ad esempio il sorriso o più in generale i movimenti delle estremità della bocca, erano molto più limitati e a fatica venivano rilevati. Si è pensato inizialmente a qualche problema associativo di *blendshape* o simile, tuttavia dopo anche una ricerca sulla rete è emerso che anche altri utenti che lo hanno provato hanno notato limitazioni per quanto concerne il sorriso. Un esempio è mostrato in Figura 5.6. Dopo un'analisi più tecnica e interna è emerso che il tracciamento del sorriso di fatto funzionava, tuttavia per qualche ragione non è enfatizzato a sufficienza nelle *blendshape* del sorriso, che su una scala da 1 a 100 non superavano mai il 30 e pertanto venivano mostrate solo poco e parzialmente.



Figura 5.6: Limitazioni del rilevamento del sorriso anche nel caso di altri utenti, come lo *youtuber* Frooxius.

Un'altra limitazione riscontrata con l'HTC VIVE Pro Eye e Facial Tracker riguarda invece le sopracciglia. Le funzionalità dell'Eye Tracker riescono a tracciare alcuni dei movimenti delle sopracciglia, come quelli più bassi, tuttavia perde di efficacia mano a mano che l'espressione diventi più complessa, come il corrucciamento o alzare le sopracciglia. Usando anche l'*avatar sample* fornito dall'SDK dell'HTC VIVE si sono riscontrati gli stessi problemi, quindi non dovrebbe essere un problema di *avatar*, bensì di tracciamento.

In ogni caso, poichè questo studio è stato pensato per avere un'analisi qualitativa di due tecnologie per le espressioni facciali in ottica consumatore e quindi rivolta ad utenti generici che generalmente non metterebbero mano agli *script* forniti da HTC

o da SALSA, si è deciso di lasciare queste limitazioni e vedere se effettivamente hanno un qualche impatto nell'espressività mostrata dall'*avatar*-attore durante le scene registrate per l'esperimento. Si rimanda in tal senso al capitolo successivo, dove vengono esposti i risultati e le discussioni sugli stessi.

5.7 Fase sperimentale

5.7.1 Idea dell'esperimento

Riprendendo l'idea esposta all'inizio di questo capitolo, e sulla base di tutto il processo di realizzazione menzionato nelle sezioni precedenti, si è arrivati alla definizione di un *iter* sperimentale che ben si accosti agli obiettivi prefissati.

In particolare, riguardo l'*avatar*-attore e il contenuto emotivo da includere in qualche modo, si è deciso di prendere in esame 6 scene, ognuna ruotante attorno ad una delle emozioni base di Ekman [71] (rabbia, felicità, tristezza, paura, disgusto e sorpresa). La scelta di focalizzarsi su 6 scenette separate è dovuta al voler vedere come le due modalità esaminate in questo caso di studio, ossia il Facial Tracker di HTC Vive Pro e SALSA Lip-Sync Suite, si comportano separatamente per ogni emozione. Questo è dovuto principalmente al fatto che, come ribadito nello studio di Ekman sopra menzionato, queste sei emozioni possono essere considerate di "base" proprio perchè possiedono un insieme delle caratteristiche espressive sul viso diverso rispetto alle altre. Nel caso in cui le scene avessero avuto un contenuto emotivo più diversificato, si rischiava di avere una valutazione più complessiva e, pertanto, meno specifica per gli obiettivi posti. Il canovaccio, o copione, delle scene è consultabile nell'Appendice A.3. Come linee guida per i copioni, ci si è attenuto al fare in modo che siano più o meno tutte equivalenti: in particolare, in ognuna è stato inserito un momento più intenso, più esasperato dell'emozione rappresentata, e in tutte si è cercato comunque di mantenere la sensazione e la percezione complessiva all'interno del campo dell'emozione rappresentata, sulla base delle scale di positività e valenza che spesso vengono usate per definire le emozioni percepite in studi sul tema [104]. Inoltre, si è cercato di renderle più "sociali" possibile e di non farle più lunghe di un minuto ognuna, così da non uscire dal contesto prefissatosi delle applicazioni virtuali sociali e Metaverso. In Figura 5.7 alcuni estratti delle scenette registrate a cui hanno assistito e che hanno valutato i partecipanti.

Una volta realizzate le registrazioni delle scenette con l'attore e salvate opportunamente grazie alla componente del progetto *Record&Play* spiegata nel Capitolo 4, tutte le parti sono pronte per la procedura.

5.7.2 Ambiente

Vista l'idea scelta dell'*avatar*-attore che recita delle scene, si è deciso di contestualizzarla usando come ambientazione il palco scenico di un teatro moderno con



Figura 5.7: Alcuni estratti dalle scene recitate dall'*avatar*-attore.

elementi *époque*, la cui *mesh* è stata presa dal sito Sketchfab⁵ e il cui autore è l'Enterprise "Connections XR"⁶. Sono state poi aggiunte poi alcune luci sopra il palco così da simulare dei riflettori e aumentare così l'immersività della scena. In Figura 5.8 sono mostrate due inquadrature dalla posizione in cui si troveranno i partecipanti una volta iniziata l'esperienza.



Figura 5.8: Vista sulla platea dal palco, con l'*avatar*-attore al centro.

5.7.3 Attività e procedura

Come nel precedente caso di studio, anche qui l'obiettivo ultimo è il confronto tra due modalità di rappresentazione, seppur in un contesto diverso. Per questa ragione, la linea generale è rimasta simile a quella del primo esperimento fatto.

⁵<https://sketchfab.com>

⁶<https://sketchfab.com/connectionsxr>

Più in particolare, dopo alcune brevi domande conoscitive demografiche rivolte al partecipante come età, genere, esperienza generale con la VR (da 1 - Quasi nulla, a 5 - Quasi quotidiana) e infine esperienza con Metaverso e/o Realtà Virtuali orientate al sociale (da 1 - Quasi nulla, a 5 - Quasi quotidiana), al partecipante veniva presentato lo studio e le indicazioni per l'esperienza. A tutti i partecipanti è stato detto che lo studio si focalizzava sull'espressività e la comunicabilità dell'*avatar*, in particolar modo della parte superiore del corpo, dove i tre fattori comunicativi principali che intervengono in una qualsiasi conversazione, anche nel mondo reale (tono della voce, gesti delle braccia e delle mani ed espressioni facciali), erano più evidenti. Nessuno dei partecipanti era al corrente all'inizio dell'esperienza che le due modalità differivano solo nella rappresentazione delle espressioni facciali e non nelle altre componenti, così da non creare alcun *bias* e far loro considerare sempre l'insieme delle componenti comunicative. Stava poi a loro, durante l'esperienza, accorgersi se effettivamente l'espressività del viso era diversa oppure rimasta equivalente cambiando da una modalità all'altra.

Terminata la spiegazione dell'esperienza, ai partecipanti veniva fatto indossare l'HMD HTC VIVE Pro e venivano fatti posizionare opportunamente nella realtà così che nel mondo virtuale si trovassero ad una distanza ottimale per osservare l'*avatar*-attore. Ad ogni partecipante è stata data la possibilità di muoversi liberamente durante la riproduzione delle scene, qualora lo ritenessero opportuno per osservare meglio l'*avatar*-attore, oppure anche di sedersi su una sedia se l'esperienza iniziava a dar loro fastidio a causa proprio dei movimenti. Una volta in posizione, veniva riprodotta una scena casuale tra le 6, in una delle due modalità (HTC VIVE Pro *Facial Tracker* o SALSA Lip-Sync Suite) anch'essa scelta in modo casuale. Terminata la scena in quella modalità, veniva mostrata subito successivamente anche l'altra modalità della stessa scena, così da garantire che i partecipanti avessero la mente fresca per cogliere le eventuali sottili differenze. Una volta conclusasi anche la seconda modalità della particolare scena, venivano poste al partecipante alcune domande di confronto sulle due modalità, chiedendo principalmente quale delle due avesse preferito relativamente a certi ambiti e, infine, complessivamente.

Dopo aver quindi spiegato la modalità di risposta e raccolto tutte le risposte, si procedeva con la seconda delle 6 scene, anch'essa scelta casualmente, con una delle due modalità, anche qui casuale. Tale *iter* è stato poi ripetuto fino alla conclusione, nelle due modalità, della sesta scenetta, terminando così di fatto l'esperienza complessiva.

5.7.4 Questionari e valutazione

Come già accennato nella precedente sezione. l'esperienza valutativa avveniva tra una scena riprodotta e l'altra, pertanto al fine di non scomodare eccessivamente il partecipante con il rimuovere e il rimettere l'HMD dopo ognuna delle 6 scene, è stato scelto un approccio che ha coniugato il virtuale e l'orale, ossia porre le domande a voce mentre il partecipante vedeva nel suo campo visivo virtuale la griglia di valutazione. Questa soluzione ha permesso così che il partecipante all'esperienza

non si togliesse mai l'HMD, e quindi si limitassero di conseguenza tutti gli eventuali fastidi dovuti al caso.

Per le domande specifiche poste dopo aver visto entrambe le modalità di ogni scena si rimanda all'Appendice A.2. Le domande sono state create *ad hoc* per l'esperienza, volendo misurare in modo generale dapprima l'agio degli utenti con una particolare rappresentazione, quindi più dal punto di vista psicologico, per poi proseguire poi con aspetti puramente estetici come realismo, piacevolezza, naturalezza, verosimiglianza. Sono state poi inserite anche un paio di domande più relative all'espressività e al veicolamento emotivo, lasciando poi come ultima domanda quella di preferenza complessiva.

Ogni domanda presupponeva come risposta un numero da 1 a 5, con le due modalità viste per ogni scena agli estremi di questa scala: una valutazione di "1" esprime una completa favorevolezza verso la prima modalità vista (A) per l'ambito chiesto nella domanda, mentre una valutazione di "5" esprime una completa preferenza per la seconda modalità vista (B); valutazioni come "2" o "4" esprimono invece una preferenza di una delle due modalità (risp. A o B) tuttavia meno marcata; infine, una valutazione di "3" esprime un'equivalenza tra le due modalità viste, oppure anche l'eventualità che il partecipante non abbia colto alcuna differenza mentre guardava le due modalità. In Figura 5.9 è mostrata la griglia che il partecipante aveva davanti a sé nel mondo virtuale al momento della valutazione.



Figura 5.9: Canvas per la valutazione interno all'applicazione dal punto di vista dell'utente.

Capitolo 6

Risultati: Secondo studio

6.1 Partecipanti

Per il test hanno deciso di partecipare su base volontaria 31 persone (18 uomini, 13 donne) di età compresa tra i 21 e i 68 anni. Buona parte dell'insieme ha risposto, nella parte demoscopica, di avere poca esperienza con tecnologie VR, mentre la stragrande maggioranza ha ritenuto di non possedere esperienza in ambito Metaverso o Realtà Virtuali sociali. Di queste esperienze ne sono state rimosse tre, di cui una per impossibilità di proseguimento e due perchè ritenute incoerenti nel loro complesso, forse a causa di una mancata comprensione della modalità di risposta alle domande.

Ciò porta quindi il campionario netto a 28 elementi totali.

6.2 Risultati

In questa sezione verranno esposti i risultati relativi al secondo caso di studio della presente tesi, discutendo delle significatività ottenute e altresì di quelle mancate che ci si aspettava di ottenere. L'indagine dei risultati è stata fatta ottenendo le medie delle risposte date ad ogni singola domanda, separatamente per scena. E' stata poi testata la normalità dei dati, che si sono rivelati come distribuiti non normalmente. Sulla base di ciò, è stato eseguito il test non parametrico con segno Wilcoxon, mantenendo la significatività del 5% come nel precedente caso di studio.

Partendo dalla scena più incentrata sull'emozione della "Tristezza", i risultati medi dei punteggi sono mostrati in Figura 6.1. Come si può notare fin da subito, la superiorità dell'HTC VIVE Pro Eye e Facial Tracker è stata significativa su tutte e otto le domande poste ai partecipanti. In particolare, tale tecnologia ha convinto maggiormente per quanto concerne l'espressività generale dell'*avatar* (3,93 vs 2,07; $p = 0,0016$) e il veicolamento delle emozioni (4,00 vs 2,00; $p = 0,001$). Seguono anche il resto delle categorie, come piacevolezza complessiva (3,78 vs 2,22; $p = 0,0056$), sensazione di essere a proprio agio (3,71 vs 2,29; $p = 0,008$), realismo (3,64 vs 2,36; $p = 0,004$) e giudizio complessivo (3,82 vs 2,18; $p = 0,009$), arrivando

a quelle leggermente minori, seppur ancora significative, come la naturalezza delle espressioni (3,64 vs 2,36; $p = 0,0066$) e la verosimiglianza (3,57 vs 2,43; $p = 0,014$). Questi risultati sono in linea con quanto ci si aspettava, visto che l'emozione della tristezza è maggiormente veicolata dal movimento degli occhi e della bocca [71], che l'HTC VIVE Pro Eye e Facial Tracker fa in modo piuttosto accurato sulla base di quanto visto durante la fase di progettazione.

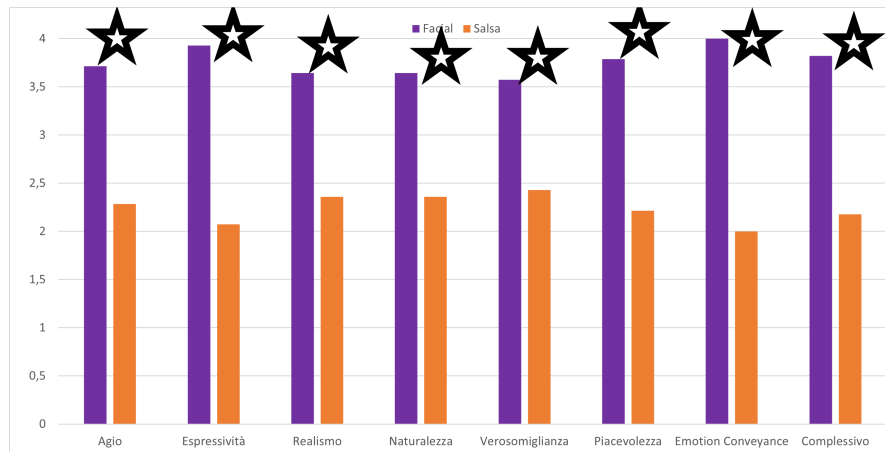


Figura 6.1: Risultati relativi alle domande poste dopo la scena focalizzata sull'emozione della “**Tristezza**” del secondo caso di studio. Per le domande specifiche, si rimanda all'Appendice A.2. I risultati statisticamente significativi, se presenti, sono marcati con un simbolo a forma di stella.

Si procede quindi verso la scena incentrata sull'emozione del disgusto, i cui risultati alle domande poste sono mostrati in Figura 6.2. Anche in questo caso si nota la significativa vittoria dell'HTC VIVE Pro Eye e Facial Tracker, seppur non in modo completo come nella precedente. In questa scena, i partecipanti hanno trovato la modalità dove veniva usato il Facial Tracker più realistica (3,67 vs 2,33; $p = 0,015$), più naturale (3,75 vs 2,25; $p = 0,008$), più verosimile nel complesso (3,75 vs 2,25; $p = 0,0087$), più piacevole (3,71 vs 2,29; $p = 0,02$), che mettesse più a proprio agio (3,57 vs 2,43; $p = 0,04$) e come preferenza complessiva (3,75 vs 2,25; $p = 0,0039$). Sono mancate le significatività per quanto concerne l'espressività e il veicolamento del carico emotivo della scena, che probabilmente sarebbe stata raggiunta con qualche partecipante aggiuntivo vista la loro vicinanza alla significatività statistica (risp. $p = 0,0585$ e $p = 0,07$). Alcuni partecipanti hanno inoltre sostenuto, dopo aver effettuato l'esperienza, che per alcune emozioni, in particolare quella del disgusto e della sorpresa, a causa della loro natura sono meno esposte temporalmente delle altre: un'espressione arrabbiata, triste o felice può essere mantenuta più facilmente durante un discorso o una conversazione, ad esempio, ma per quanto riguarda disgusto e sorpresa diventa più difficile mantenerla allo stesso modo in quanto emozioni più legate “al momento”.

Proseguendo oltre, i risultati della terza scena, relativa all'emozione della paura,

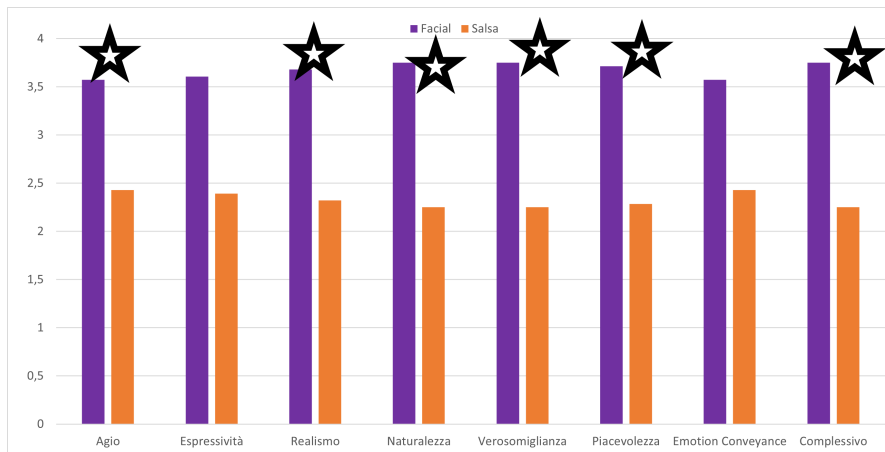


Figura 6.2: Risultati relativi alle domande poste dopo la scena focalizzata sull’emozione del “**Disgusto**” del secondo caso di studio. Per le domande specifiche, si rimanda all’Appendice A.2. I risultati statisticamente significativi, se presenti, sono marcati con un simbolo a forma di stella.

sono riportati in Figura 6.3. Nel complesso si ha anche qui una superiorità dell’HTC VIVE Pro Eye e Facial Tracker, seppur tale superiorità sia effettivamente significativa solo in quattro delle otto categorie di risposta. In particolare, i partecipanti hanno ritenuto la modalità con il Facial Tracker come più naturale nelle espressioni (3,60 vs 2,40; $p = 0,009$), più verosimile (3,53 vs 2,47; $p = 0,02$), come più capace di veicolare il contenuto emotivo della scena (3,57 vs 2,43; $p = 0,037$) e come preferenza generale tra le due modalità (3,60 vs 2,40; $p = 0,016$) rispetto alla controparte di SALSA Lip-Sync Suite. Per le altre categorie, sebbene ci sia una superiorità in linea con le aspettative iniziali del progetto, tale superiorità non è risultata significativa, seppur per poco nel caso dell’espressività ($p = 0,056$) o per il realismo percepito ($p = 0,06$). Anche qui, ciò probabilmente sarebbe potuto diventare un risultato statisticamente significativo con un maggior numero di test e di partecipanti.

Venendo ora alla quarta scena, quella incentrata sull’emozione della felicità, si mostrano i risultati in Figura 6.4. Anche qui, come nella precedente, si ha una superiorità non eccessiva dell’HTC VIVE Pro Eye e Facial Tracker. In particolar modo tale differenza con l’altra modalità, quella con SALSA Lip-Sync Suite, è risultata significativa per quanto concerne il realismo percepito (3,46 vs 2,54; $p = 0,027$), la verosimiglianza generale (3,46 vs 2,54; $p = 0,039$), la piacevolezza estetica (3,64 vs 2,36; $p = 0,016$) e infine la preferenza generale tra le due modalità (3,50 vs 2,50; $p = 0,045$). Si noti come in questa scena, per quanto concerne l’espressività si è avuto un pressochè pareggio tra le due modalità: ciò è stato dovuto probabilmente al fatto che l’HTC VIVE Pro Eye e Facial Tracker, come già menzionato nel Capitolo 5, faceva fatica a rilevare correttamente quando il viso dell’utente tracciato sorridesse. Questa teoria è in linea con alcuni dei commenti dei

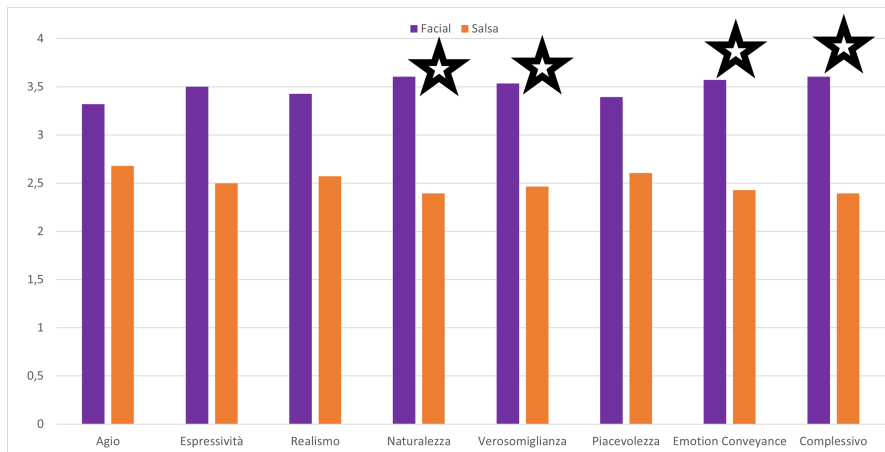


Figura 6.3: Risultati relativi alle domande poste dopo la scena focalizzata sull'emozione della **“Pauro”** del secondo caso di studio. Per le domande specifiche, si rimanda all'Appendice A.2. I risultati statisticamente significativi, se presenti, sono marcati con un simbolo a forma di stella.

partecipanti che infatti hanno visto l'*avatar* poco sorridente nonostante il contesto e, pertanto, meno espressivo.

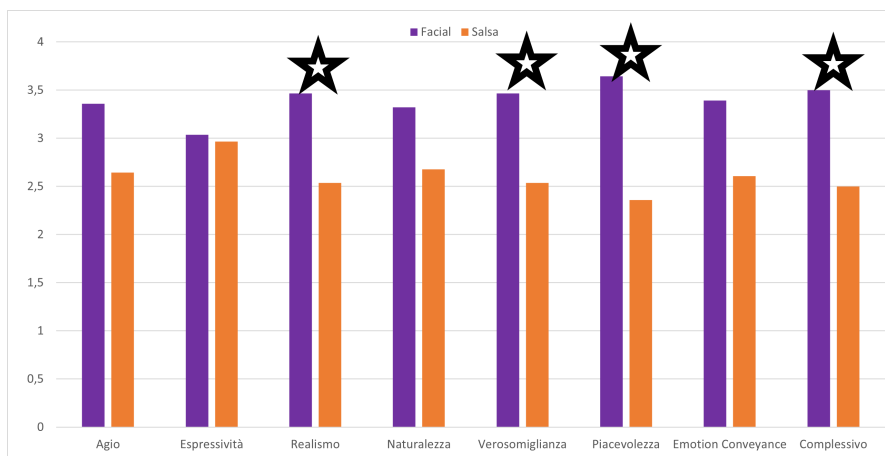


Figura 6.4: Risultati relativi alle domande poste dopo la scena focalizzata sull'emozione della **“Felicita'”** del secondo caso di studio. Per le domande specifiche, si rimanda all'Appendice A.2. I risultati statisticamente significativi, se presenti, sono marcati con un simbolo a forma di stella.

Per quanto concerne la scena relativa alla rabbia, i risultati sono mostrati in Figura 6.5. Come si può notare, in questo caso le due modalità sono molto vicine e pressochè equivalenti, seppur con un leggero margine a favore del Facial Tracker dell'HTC VIVE Pro rispetto che alla controparte SALSALip-Sync Suite. Tale vantaggio non è però significativo, pertanto non verrà ritenuto un risultato

importante dello studio presente. In questo caso, stando ad alcuni commenti dei partecipanti, questa scena era molto più “parlata” rispetto alle analoghe e pertanto gli effetti della modalità SALSA Lip-Sync Suite si sono sentiti maggiormente in positivo vista la sua efficacia dal lato “simulazione del parlato”. Ciò è in linea con i punti di forza definiti dallo stesso autore della Suite [105] e pertanto potrebbe essere interessante valutare in futuro se i dati cambino con una scena sempre focalizzata sulla rabbia, ma meno parlata. Altri commenti hanno inoltre notato come di base la *mesh* dell’*avatar* fosse un po’ corruciata e “imbronciata” già in posizione base, e quindi ha reso meno visibili i movimenti delle sopracciglia che il Facial Tracker provava a fare.

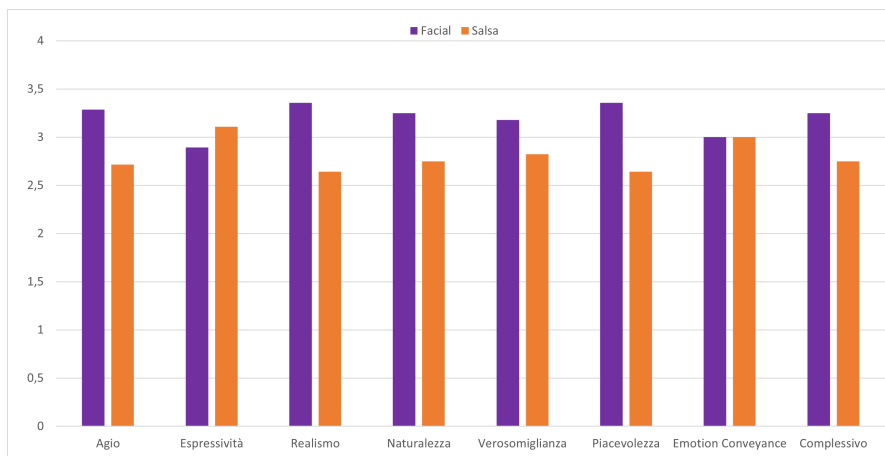


Figura 6.5: Risultati relativi alle domande poste dopo la scena focalizzata sull’emozione della “**Rabbia**” del secondo caso di studio. Per le domande specifiche, si rimanda all’Appendice A.2. I risultati statisticamente significativi, se presenti, sono marcati con un simbolo a forma di stella.

Infine, in Figura 6.6 sono mostrati i risultati relativi alla scena focalizzata sull’emozione della sorpresa. Qui, più di tutte le precedenti scene, le due modalità appaiono praticamente equivalenti e pertanto non vi è nessun risultato statisticamente rilevante, nonostante ci si aspettasse perlomeno un qualche margine a favore della modalità del Facial Tracker. Come detto precedentemente per quanto concerne il disgusto, la sorpresa è un’emozione molto effimera, che è molto espressiva in un particolare momento ma poi fa fatica ad essere mantenuta per lungo termine come ad esempio la rabbia o la tristezza. Si pensi anche al mondo reale: quando c’è una sorpresa, generalmente l’insieme delle caratteristiche facciali che la contraddistinguono si manifestano solo nei primi secondi, per poi sfociare in un’espressione più felice se la sorpresa è stata gradevole e triste o arrabbiata se è stata sgradevole. Studi futuri più approfonditi potrebbero focalizzarsi su queste teorie, ma per gli obiettivi di questa tesi i risultati derivati da questa scenetta non sono significativi e pertanto nulla di certo può essere detto al riguardo.

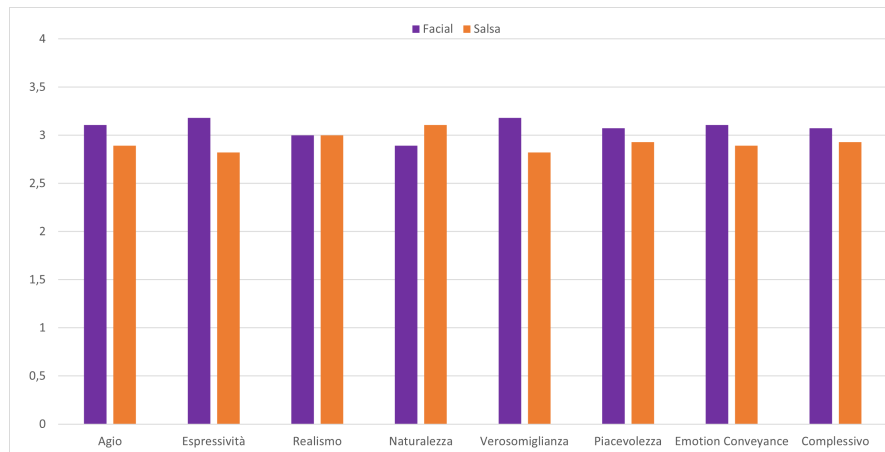


Figura 6.6: Risultati relativi alle domande poste dopo la scena focalizzata sull’emozione della **“Sorpresa”** del secondo caso di studio. Per le domande specifiche, si rimanda all’Appendice A.2. I risultati statisticamente significativi, se presenti, sono marcati con un simbolo a forma di stella.

6.3 Discussione

In questa sezione finale verrà fatta una piccola ricapitolazione dei risultati conseguiti con questo studio e verranno tratte alcune conclusioni sulla base di essi. In questo caso di studio sono state confrontate due tecniche di rappresentazione delle espressioni facciali in un contesto “multi-utente simulato”, ossia in un’esperienza singolo utente dove si ha davanti un *avatar*-attore NPC che fa le veci di un altro utente nello stesso spazio virtuale. Per questo esperimento sono stati usati due *tool* commerciali capaci di animare le espressioni del viso di un *avatar*: da una parte l’hardware HTC VIVE Pro Eye e Facial Tracker, che grazie a due telecamere ad infrarossi, una interna all’HMD per il tracciamento degli occhi e una applicabile esternamente per il tracciamento della parte bassa del volto riesce a ricostruire buona parte dei movimenti espressivi di un utente; dall’altra, la Suite Lip-Sync di SALSA, ossia un insieme di algoritmi e regole che sulla base delle informazioni contenute in un file audio animano la bocca dell’*avatar* e alcuni suoi visemi per cercare di renderla il più possibile fedele all’audio, aggiungendo anche dei movimenti generati casualmente di occhi e del resto del viso per rendere l’*avatar* meno statico.

Per questo studio sono state valutate le opinioni di 28 partecipanti sulla base del questionario riportato in Appendice A.2, domande ripetute dopo ognuna delle scene vista nelle due modalità. Anche qui, i risultati sono piuttosto vari: per alcune emozioni, forse quelle più emblematiche come tristezza, paura e felicità, si è avuta una significativa superiorità del Facial Tracker, in particolar modo nella preferenza complessiva e nella verosimiglianza. Tale superiorità è stata riscontrata anche per quanto concerne la scena riguardante l’emozione del disgusto, che sebbene più effimera delle precedenti è stata preferita rappresentata dal Facial Tracker in

termini di realismo, agio, naturalezza e piacevolezza complessiva. Lo stesso però non è avvenuto con l'altra emozione più effimera dell'insieme considerato, ossia la sorpresa, dove invece le due rappresentazioni si sono pressochè eguagliate. Lo stesso è successo con la rabbia, dove non ci sono state significatività statistiche.

Questi risultati suggeriscono che la performance delle due tecniche confrontate può variare molto a seconda della scena riprodotta o anche dell'emozione stessa riprodotta. Come già riportato nelle sezioni precedenti da parte degli stessi partecipanti, la performance del simulatore di espressioni facciali SALSA Lip-Sync Suite è stata maggiore nelle scene più discorsive e parlate velocemente, come quella della rabbia, suggerendo che SALSA possa essere percepito meglio di un tracciamento reale in tali situazioni. Altre situazioni dove la modalità di SALSA si è avvicinata di più alla performance del Facial Tracker sono quelle scene vittime delle limitazioni hardware esposte nel Capitolo 5: in particolare la scena della felicità, dove si faticava a vedere un sorriso nella modalità Facial Tracker e quindi ha perso un po' di espressività. Analogamente, per le scene di paura e sorpresa i movimenti limitati delle sopracciglia hanno impedito al Facial Tracker di raggiungere una vittoria più significativa.

In conclusione, come si poteva prevedere il Facial Tracker ha ottenuto nel complesso una performance migliore del simulatore Lip-Sync SALSA, tuttavia la dominanza non è stata assoluta. Questo suggerisce che a seconda delle applicazioni che usano le animazioni facciali che si voglia progettare, non si ha una risposta definitiva su quale tecnologia sia meglio usare, ma che la scelta dipende da quali emozioni verranno trattate.

Capitolo 7

Conclusioni e Sviluppi Futuri

7.1 Considerazioni finali

In conclusione, questo lavoro di tesi ha proposto due casi di studio volti a confrontare, rispettivamente per il corpo e per le espressioni del volto, due tecnologie di rappresentazione differenti e tra loro piuttosto diversificate. Se da un lato si sono comparate le due tecnologie più usate e popolari nello stato dell'arte più recente, dall'altro invece sono state confrontate due tecnologie dal funzionamento e dalla logica completamente diverse per vedere la loro performance espressiva.

I casi di studio della presente tesi hanno dimostrato come gli utenti nel complesso preferiscano avere quante più informazioni visive possibili riguardo le altre entità figuranti nel mondo virtuale. Se dal primo studio, l'aver un corpo visibile permette di capire meglio le intenzioni o i gesti delle altre persone, garantendo quindi un maggiore fattore educativo e di apprendimento specie in ambiti simili a quello scelto, dal secondo studio invece emerge come le tecnologie di tracciamento del volto rendano effettivamente gli avatar più espressivi, tuttavia non in modo così netto come si potrebbe pensare rispetto ad una tecnologia più basata sulla simulazione come SALSA Lip-Sync Suite.

Entrambi gli studi hanno mostrato, sulla base dei risultati più significativi, come la tecnologia con più potenziale "reale", nel senso di realismo, sia effettivamente preferita perchè in grado di collegarsi meglio a quell'insieme di nozioni psico-sociali a cui gli esseri umani si sono abituati nel corso della loro esistenza. D'altro canto, si è visto come per certi aspetti, in particolare per la rappresentazione di sé e per l'espressione di alcune emozioni, è totalmente praticabile anche l'uso di tecnologie meno accurate e computazionalmente più leggere, e ciò può risultare utile in contesti di ottimizzazione progettuale o simili.

7.2 Sviluppi futuri

Gli sviluppi futuri relativi alle tematiche trattate in questa tesi sono vari e ampi.

Come prima opzione, sicuramente possono essere prese in considerazione molte altre tecniche di rappresentazione di *avatar*, sia dal punto di vista del corpo che dal punto di vista del volto e delle espressioni, visto il campo continuamente in evoluzione negli ultimi tempi. Inoltre, possono essere definiti nuovi scenari o delle task *ad hoc* che possano spingere gli utenti a collaborare maggiormente tra loro, dando quindi l'occasione per poter ampliare ulteriormente l'insieme delle animazioni del corpo dell'*avatar* come ad esempio il movimento da prona, il saltare o l'arrampicarsi. Potrebbe essere più che valido anche un confronto con le più attuali tecniche di rappresentazione e animazione *avatar* basate sul *Machine Learning*, sia per quanto riguarda il corpo sia per il volto, magari valutando anche il loro impatto computazionale in relazione alle tecnologie esaminate negli studi della presente tesi.

Altre possibili strade, più orientate verso l'espressività dell'*avatar* e del suo corpo, possono essere percorse al fine di capire al meglio quale sia l'effettivo impatto delle singole componenti di corpo, espressioni del viso e tono della voce per veicolare un qualsivoglia contenuto emotivo. A tal riguardo, le emozioni possono essere esaminate individualmente, valutando quindi le performance delle tecnologie usate per una particolare emozione specifica.

Appendici

A.1 Questionario completo del primo caso di studio

A.1.1 Questionario “Embodiment”

Proprietà del corpo

(-3,+3 - -3: Non sono per nulla d'accordo; +3: Sono molto d'accordo)

1. Sentivo come se la mia rappresentazione virtuale fosse il mio corpo.
2. Sentivo come se la rappresentazione virtuale che vedevo fosse di qualcun altro.
3. Sembrava come se avessi più di un corpo.
4. Mi è sembrato che la rappresentazione virtuale che ho visto quando ho guardato me stesso specchiato fosse il mio corpo.
5. Mi è sembrato che la rappresentazione virtuale che ho visto quando ho guardato me stesso specchiato fosse un'altra persona.

Rappresentanza e Controllo del moto

(-3,+3 - -3: Non sono per nulla d'accordo; +3: Sono molto d'accordo)

1. Mi sembrava che potessi controllare la mia rappresentazione virtuale come se fosse il mio corpo reale.
2. I movimenti della mia rappresentazione virtuale erano causati dai miei movimenti.
3. Sentivo come se i movimenti della mia rappresentazione virtuale stessero influenzando i miei movimenti reali.
4. Sentivo come se la mia rappresentazione virtuale si stesse muovendo da sola.

Sensazione Tattile

(-3,+3 - -3: Non sono per nulla d'accordo; +3: Sono molto d'accordo)

1. Mi sembrava di sentire il tocco sulla mano nel punto in cui toccavo gli oggetti.
2. Mi sembrava come se il contatto che ho sentito fosse situato da qualche parte tra la mia mano fisica e quella virtuale.
3. Mi sembrava come se il tocco che ho sentito fosse causato dalla mia mano reale che toccava l'oggetto virtuale.
4. Mi sembrava che le mie mani stessero toccando gli oggetti.

Posizionamento del Corpo

(-3,+3 - -3: Non sono per nulla d'accordo; +3: Sono molto d'accordo)

1. Sentivo come se il mio corpo fosse posizionato dove vedevo la mia rappresentazione virtuale.
2. Mi sentivo fuori dal mio corpo.
3. Mi sembrava che il mio corpo reale stesse avvicinandosi alla rappresentazione virtuale o che la rappresentazione virtuale stesse allineandosi con il mio corpo reale.

Aspetto Esteriore

(-3,+3 - -3: Non sono per nulla d'accordo; +3: Sono molto d'accordo)

1. Mi sentivo come se il mio corpo 'reale' stesse diventando un corpo 'avatar'.
2. Sentivo come se il mio corpo reale stesse iniziando a posizionarsi allo stesso modo della rappresentazione virtuale che vedevo.
3. Sentivo come se la mia rappresentazione virtuale somigliasse al mio corpo vero, in termini di forma, colore della pelle o altre features visuali.
4. Sentivo come se stessi indossando abiti diversi rispetto a quelli che avevo quando sono entrato nel laboratorio per il test.

Risposta a Stimoli Esterni

(-3,+3 - -3: Non sono per nulla d'accordo; +3: Sono molto d'accordo)

1. Sentivo come se il mio corpo potesse essere affetto dagli elementi virtuali (fuoco, oggetti).

2. Ho sentito una sensazione di calore nel mio corpo quando ho visto il fuoco.
3. Quando mi sono avvicinato all'incidente, ho avuto l'istinto di correre via.
4. Sentivo come se il mio corpo si stesse scaldando. Ho avuto la sensazione che il fuoco potesse farmi male.

A.1.2 Questionario “Networked Minds and Social Presence”

Mutua Consapevolezza

(1,7 - 1: Non sono per nulla d'accordo; 7: Sono molto d'accordo)

1. Ho fatto fatica ad accorgermi di un'altra persona.
2. L'altra persona non mi ha notato nell'ambiente virtuale.
3. Ero spesso cosciente di altri nell'ambiente virtuale.
4. Gli altri erano spesso coscienti della mia presenza nell'ambiente virtuale.
5. Penso che l'altra persona si sia spesso sentita da sola.
6. Mi sono sentito spesso da solo.

Allocazione dell'Attenzione

(1,7 - 1: Non sono per nulla d'accordo; 7: Sono molto d'accordo)

1. A volte ho fatto finta di prestare attenzione all'altra persona.
2. L'altra persona a volte ha fatto finta di prestare attenzione a me.
3. L'altra persona mi ha prestato molta attenzione.
4. Ho prestato molta attenzione all'altra persona.
5. L'altra persona veniva facilmente distratta quando altre cose succedevano intorno a noi.
6. Sono stato distratto facilmente quando altre cose succedevano intorno a me.
7. L'altra persona tendeva ad ignorarmi.
8. Tendevo a ignorare l'altra persona.

Mutua Comprensione

(1,7 - 1: Non sono per nulla d'accordo; 7: Sono molto d'accordo)

1. Le mie intenzioni erano chiare all'altra persona.
2. Le intenzioni dell'altra persona mi erano chiare.
3. I miei pensieri erano chiari all'altra persona.
4. I pensieri dell'altra persona mi erano chiari.
5. L'altra persona capiva cosa io volessi dirle.
6. Capivo cosa l'altra persona mi stesse dicendo.

Interdipendenza Comportamentale

(1,7 - 1: Non sono per nulla d'accordo; 7: Sono molto d'accordo)

1. Le mie azioni dipendevano dalle azioni dell'altra persona.
2. Le azioni dell'altra persona dipendevano dalle mie azioni.
3. Il mio comportamento era in risposta diretta al comportamento dell'altra persona.
4. Il comportamento dell'altra persona era in risposta diretta al mio comportamento.
5. Ciò che faceva l'altra persona condizionava ciò che facevo io.
6. Ciò che facevo io condizionava ciò che faceva l'altra persona.

Mutua Assistenza

(1,7 - 1: Non sono per nulla d'accordo; 7: Sono molto d'accordo)

1. L'altra persona non mi ha aiutato molto.
2. Non ho aiutato l'altra persona molto.
3. L'altra persona ha lavorato insieme a me per completare il task.
4. Ho lavorato insieme all'altra persona per completare il task.

Azioni Dipendenti

(1,7 - 1: Non sono per nulla d'accordo; 7: Sono molto d'accordo)

1. L'altra persona non poteva agire senza di me.
2. Io non potevo agire senza l'altra persona.

A.1.3 Questionario “VR-Use”

Senso di Immersione e Presenza

(1,5 - 1: Non sono per nulla d'accordo; 5: Sono molto d'accordo)

1. Ho sentito la sensazione di essere immerso nell'ambiente virtuale.
2. Non ho avuto bisogno di sentirmi immerso nell'ambiente virtuale per completare il mio task.
3. Ho avuto un senso di presenza (di essere lì)
4. La qualità dell'immagine ha ridotto il mio senso di presenza.
5. Penso che il campo di vista del visore abbia migliorato il mio senso di presenza.
6. La risoluzione del visore ha ridotto il mio senso di immersione.
7. Mi sono sentito isolato e non parte dell'ambiente virtuale.
8. Avevo un buon senso delle proporzioni nell'ambiente virtuale.
9. Spesso non sapevo dove fossi nell'ambiente virtuale.
10. In generale, il senso di presenza è stato soddisfacente.

A.1.4 Domande finali di confronto tra modalità

(1,2 - 1: Full Body; 2: VR kit (Visore e Controller))

1. Per quanto riguarda l'usabilità (avatar proprio), ho preferito la versione...
2. Per quanto riguarda l'estetica (avatar proprio), ho preferito la versione...
3. Per quanto riguarda l'estetica (altro avatar), ho preferito la versione...
4. Per quanto riguarda le interazioni multiplayer, ho preferito la versione...
5. Per quanto riguarda il mio avatar, ho preferito la versione...
6. Per quanto riguarda l'altro avatar, ho preferito la versione...
7. Considerando l'esperienza complessiva, ho preferito la versione...

A.2 Questionario completo del secondo caso di studio

Le domande elencate qui sotto sono state poste dopo la visualizzazione di ogni scena dopo aver visto le due modalità. Il partecipante doveva rispondere con un numero da 1 a 5, dove 1 esprimeva totale preferenza per la prima modalità vista, mentre 5 per la seconda, lasciando al resto dei valori (2,3,4) la scala opportuna tra le due.

1. Con quale delle due modalità mostrate ti sei sentito più a tuo agio?
2. Quale delle due modalità ritieni sia più espressiva?
3. Quale delle due modalità ritieni sia più realistica?
4. Quale delle due modalità ritieni sia più naturale?
5. Quale delle due modalità ritieni sia più verosimile?
6. Quale delle due modalità ritieni sia più piacevole?
7. Quale delle due modalità può veicolare meglio il contenuto emotivo?
8. Quale delle due modalità ritieni sia più adatta per l'interazione tra utenti in ambienti virtuali sociali?

A.3 Copioni delle scene mostrate nel secondo caso di studio

A.3.1 Scena “Happy”

(L'attore sta di fronte allo spettatore, sorridente)

-Allora, ora che siamo qui cosa ti va di fare? Abbiamo la ruota panoramica, le montagne russe oppure... Oh! La casa infestata! Qualche preferenza?? Altrimenti scelgo io!

(Dice, parlando piuttosto velocemente come in preda all'eccitazione)

-Vediamo...

(L'attore fa una posa pensierosa mentre si guarda intorno per qualche secondo)

-...Non mi va molto la casa infestata...

(Trascorrono alcuni secondi dove l'attore si guarda in giro)

-...Però quelle montagne russe sembrano divertenti! Senti quanto strillano!

(Dice, guardando e indicando un punto lontano, dopo un po' di tempo)

-Uh guarda laggiù, la mascotte del parco sta venendo verso di noi!

(Dice, ora riposizionato di fronte allo spettatore, per poi girarsi verso la platea quando menziona la mascotte)

(Si aspetta qualche secondo, magari salutando la mascotte quando arriva)
-Ahahahahahah! Guarda com'è simpatica! E soprattutto ha il pelo super morbido!
(Con entusiasmo mentre ridacchia e la accarezza)
-Facciamoci una foto insieme!

A.3.2 Scena “Sad”

(L'attore parte in piedi, con una mano sul fianco e una sul viso, come esasperato)
-... Come siamo finiti così...
(L'attore sospira, scuotendo poi la testa)
-... Tu eri... Eri... Tutto per me. Letteralmente tutto...
(Inizia, levandosi la mano dal volto e guardando in basso, ha un atteggiamento vagamente esasperato)
-... Non c'era giorno dove non ridevamo insieme... dove la nostra complicità era all'ordine del giorno...
(Sospira, prendendosi qualche secondo di pausa)
-... Giorni dove il mio unico desiderio era annegare nel dolce profumo che lasciavi sul cuscino, e perdermi tra quelle onde dorate dei tuoi capelli... Queste piccole cose... erano tutta la mia felicità. Non avevo bisogno di altro, perché tu eri tutto ciò di cui avevo davvero bisogno.
(Un'altra pausa, dove l'attore scuote la testa)
-... Come dovrei andare avanti ora, senza averti al mio fianco? Come ti aspetti che io... possa continuare senza di te...?
-... Non posso farlo... Non ne sono in grado...
(Dice, per poi portarsi nuovamente le mani sul volto)

A.3.3 Scena “Angry”

(L'attore è fermo in piedi, con le braccia conserte e sguardo accigliato)
-Davvero?! Questo è tutto ciò che hai da dire dopo quello che è successo?!? Sono davvero senza parole.
(Dice, molto distaccatamente e con media enfasi e un pizzico di sarcasmo)
-Prima di tutto la colpa è tua se siamo finiti in questo diavolo di casino. Come ti viene in mente di difenderlo dopo ciò che ha fatto?! Siamo chiusi qui dentro per colpa TUA! TUA AAAA! Non so se ti è chiara questa cosa.
(Dice, alzando progressivamente la voce)
(Seguono alcuni silenzi dove passeggia lì in giro sul posto mentre usa gesti esasperati)
-... E poi... Come diavolo puoi pensare di fare una cosa simile? Pensi che la passerai liscia una volta che saremo usciti da qui?! Io non credo proprio. Oh no che non lo credo.
(Dice, incrociando di nuovo le braccia al petto; seguono alcuni attimi di silenzio)
-No, lasciami. Non provare a toccarmi. Non starmi nemmeno vicino, guarda.

(Dice poi, rivolto allo spettatore, infastidito e accigliato; con gesti mirati ad allontanarsi)

-Stai lì fermo e non darmi fastidio.

(Con gesto indicatore e sguardo accigliato verso lo spettatore)

A.3.4 Scena “Fear”

(Movimenti lenti e cauti, incerti; l'attore si guarda ripetutamente intorno)

-...N-non riesco a vedere molto qui...

(come a parlare tra sé e sé, pensare a voce alta, continuando a guardarsi intorno, con respiro pesante; aspetta qualche secondo in silenzio continuando le azioni precedenti)

-...Cos'è stato...? Cos'era quel rumore...?

(Si volta improvvisamente come a cercare qualcosa; dopo qualche secondo di silenzio fa un respiro profondo per calmarsi)

(L'attore continua quindi a camminare, guardandosi intorno)

(All'improvviso si ferma, osservando un punto preciso alle spalle dello spettatore)

-Cosa... Cosa diavolo è quell-

(Al posto dello spettatore è come se vedesse un fantasma o un mostro)

-AAAAAAAAAAAAAH!!

(Sposta le braccia in posizione di difesa, per coprirsi, mentre fa una faccia terrorizzata)

-Stammi lontano! Stammi lontano!

(Urla, indietreggiando piano piano sempre più)

(Dopo un po' di tensione e respiro affannato l'attore prova a guardarsi di nuovo in giro ancora tremante)

-E'... sparito...

(Commenta mettendosi una mano sul petto)

A.3.5 Scena “Surprise”

(L'attore sta passeggiando tranquillo, quando dopo qualche secondo all'improvviso si ferma e guarda lo spettatore)

-...Mh? Come dici? Mi devo coprire gli occhi? Perché?

(Chiede l'attore, confuso, per poi sospirare)

-Va bene, va bene...

(Dice, enfatizzando con gesti come alzare le mani in segno di resa ecc, sospirando; si copre poi il viso con le mani, aspettando qualche secondo)

-Ora? Dove hai intenzione di portarmi?

(Chiede quindi, mantenendo il viso coperto; inizia poi a camminare nelle vicinanze, per poi fermarsi più o meno di fronte allo spettatore, o un po' lateralmente)

-Posso aprire? Di già? Non mi pare ci siamo spostati così tanto...

(Commenta, con leggerezza, per poi dopo un po' levare le mani e guardarsi intorno; il suo viso diventa quindi sorpreso, guardando un'ipotetica cosa piacevole oltre lo spettatore, come un tramonto ad esempio)

(Vari secondi di silenzio e meraviglia)

-Woah...! non ci posso credere... è stupendo. È davvero... meraviglioso.

(Commenta poi, con tono un po' esitante ma piacevolmente sorpreso)

A.3.6 Scena “Disgust”

(L'attore è come se si fosse appena alzato dal letto, e va verso il frigorifero per fare colazione)

-Yaaaawn.

(L'attore sbadiglia e si stiracchia)

-Vediamo un po' cosa c'è per colazione...

(L'attore usa un tono piuttosto assonnato)

(Alcuni secondi passano mentre osserva dall'alto verso il basso gli ipotetici alimenti presenti nel frigorifero e ne estrae qualcuno; all'improvviso annusa l'aria, come se avesse percepito un odore strano)

-... Cos'è questo odore?

(Si chiede, continuando ad annusare l'aria e a spostare poi lo sguardo sul frigorifero; l'attore inizia quindi a smaneggiare virtualmente come se spostasse delle cose, cercando di osservare e trovare la fonte del cattivo odore)

-... Che sia questo...?

(Dopo vari secondi di ricerca estrae un ipotetico oggetto virtuale e lo guarda per qualche secondo per poi portarlo vicino al viso ed annusarlo)

-... Bleah, che schifo! E' completamente marcio...

(Dice l'attore, molto inorridito e disgustato, allontanandolo subito dal proprio viso)

Bibliografia

- [1] Kristine L. Nowak e Jesse Fox. «Avatars and computer-mediated communication: a review of the definitions, uses, and effects of digital representations». In: *Review of Communication Research* 6 (2018), pp. 30–53. ISSN: 2255-4165. DOI: <https://doi.org/10.12840/issn.2255-4165.2018.06.01.015> (cit. a p. 2).
- [2] Eun-Ju Lee. «Character-based team identification and referent informational influence in computer-mediated communication». In: *Media Psychology* 9.1 (2007), pp. 135–155 (cit. a p. 2).
- [3] L NAKAMURA. *Cybertypes: race, ethnicity, and identity on the Internet/-Lisa Nakamura*. 2002 (cit. a p. 2).
- [4] Igor Kotlyar e Dan Ariely. «The effect of nonverbal cues on relationship formation». In: *Computers in Human Behavior* 29.3 (2013), pp. 544–551. ISSN: 0747-5632. DOI: <https://doi.org/10.1016/j.chb.2012.11.020>. URL: <https://www.sciencedirect.com/science/article/pii/S0747563212003251> (cit. a p. 2).
- [5] Gary Bente e Nicole C Krämer. «Virtual gestures: embodiment and nonverbal behavior in computer-mediated communication». In: *Face-to-face communication over the internet: issues, research, challenges* (2011), pp. 176–209 (cit. a p. 2).
- [6] Sun Joo Ahn e Jeremy N Bailenson. «201 Self-Endorsing Versus Other-Endorsing in Virtual Environments». In: *Journal of Advertising* 40.2 () (cit. a p. 2).
- [7] Davide Calandra, Filippo Gabriele Praticò, Massimo Migliorini, Vittorio Verda e Fabrizio Lamberti. «A Multi-role, Multi-user, Multi-technology Virtual Reality-based Road Tunnel Fire Simulator for Training Purposes». In: *Proc. of 16th International Conference on Computer Graphics Theory and Applications (GRAPP 2021)*. 2021, pp. 96–105. ISBN: 978-989-758-488-6. DOI: 10.5220/0010319400960105 (cit. alle pp. 4, 24, 31).
- [8] Amy L. Baylor. «Promoting motivation with virtual agents and avatars: role of visual presence and appearance». In: *Phil. Trans. R. Soc.* B3643559–3565 (2009). DOI: 10.1098/rstb.2009.0148 (cit. a p. 7).

- [9] Alexander Schäfer, Gerd Reis e Didier Stricker. «A Survey on Synchronous Augmented, Virtual and Mixed Reality Remote Collaboration Systems.» In: *arXiv preprint arXiv:2102.05998* (2021) (cit. a p. 8).
- [10] Federico De Lorenzis, Filippo Gabriele Praticò e Fabrizio Lamberti. «HCP-VR: Training First Responders through a Virtual Reality Application for Hydrogeological Risk Management.» In: *VISIGRAPP (2: HUCAPP)*. 2022, pp. 273–280 (cit. a p. 9).
- [11] Rajinder S. Sodhi, Brett R. Jones, David Forsyth, Brian P. Bailey e Giuliano Maciocci. «BeThere: 3D Mobile Collaboration with Spatial Input». In: New York, NY, USA: Association for Computing Machinery, 2013. ISBN: 9781450318990. DOI: 10.1145/2470654.2470679. URL: <https://doi.org/10.1145/2470654.2470679> (cit. a p. 9).
- [12] Martijn J. Schuemie, Peter van der Straaten e Merel Krijn. «Research on Presence in Virtual Reality: A Survey.» In: *CyberPsychology & Behavior*, 4(2), 183–201 (2001). DOI: 10.1089/109493101300117884 (cit. a p. 8).
- [13] Doug A. Bowman e Ryan P. McMahan. «Virtual Reality: How Much Immersion Is Enough?» In: *Computer* 40.7 (2007), pp. 36–43. DOI: 10.1109/MC.2007.257 (cit. a p. 8).
- [14] Di Tore S., Todino M. D. e Sibilio M. «La realtà virtuale come strumento didattico per favorire lo sviluppo della presa di prospettiva.» In: *In: (a cura di): Panciroli C., Animazione digitale per la didattica. p. 155-164, MILANO:FrancoAngeli, ISBN: 978-88-351-0728-6* (2020) (cit. a p. 9).
- [15] Anthony Steed, Ye Pan, Fiona Zisch e William Steptoe. «The impact of a self-avatar on cognitive load in immersive virtual reality». In: *2016 IEEE Virtual Reality (VR)*. 2016, pp. 67–76. DOI: 10.1109/VR.2016.7504689 (cit. a p. 9).
- [16] Guo Freeman e Divine Maloney. «Body, Avatar, and Me: The Presentation and Perception of Self in Social Virtual Reality». In: *4.CSCW3* (2021). DOI: 10.1145/3432938. URL: <https://doi.org/10.1145/3432938> (cit. a p. 9).
- [17] *Tencent ISUX - 2019–2020 Design Trend Avatar*. URL: <https://blog.prototypr.io/2019-2020-design-trend-avatar-70a449cc5f21> (cit. a p. 10).
- [18] Mar Gonzalez-Franco, Brian Cohn, Eyal Ofek, Dalila Burin e Antonella Maselli. «The Self-Avatar Follower Effect in Virtual Reality». In: *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 2020, pp. 18–25. DOI: 10.1109/VR46266.2020.00019 (cit. a p. 10).
- [19] Riku Otono, Naoya Isoyama, Hideaki Uchiyama e Kiyoshi Kiyokawa. «Third-Person Perspective Avatar Embodiment in Augmented Reality: Examining the Proteus Effect on Physical Performance». In: *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 2022, pp. 730–731. DOI: 10.1109/VRW55335.2022.00216 (cit. a p. 10).

- [20] J.M. Ferreira. «What If We Look at the Body? An Embodied Perspective of Collaborative Learning». In: *Educ Psychol Rev* 33, 1455–1473 (2021). DOI: 10.1007/s10648-021-09607-8. URL: <https://doi.org/10.1007/s10648-021-09607-8> (cit. a p. 10).
- [21] P. Bjørn, M. Wulff, M.S. Peträus e al. «Immersive Cooperative Work Environments (CWE): Designing Human-Building Interaction in Virtual Reality». In: *Comput Supported Coop Work* 30, 351–391 (2021). DOI: 10.1007/s10606-021-09395-3. URL: <https://doi.org/10.1007/s10606-021-09395-3> (cit. a p. 10).
- [22] Alexander Schäfer, Gerd Reis e Didier Stricker. «A Survey on Synchronous Augmented, Virtual and Mixed Reality Remote Collaboration Systems». In: *arXiv* (2021). DOI: 10.48550/ARXIV.2102.05998. URL: <https://arxiv.org/abs/2102.05998> (cit. a p. 10).
- [23] R.A. Ruddle e S. Lessels. «For efficient navigational search, humans require full physical movement, but not a rich visual scene». In: *Psychol Sci* 17(6):460–5 (2006). DOI: 10.1111/j.1467-9280.2006.01728.x. (cit. a p. 11).
- [24] D. Waller e E. Hodgson. «Sensory Contributions to Spatial Knowledge of Real and Virtual Environments». In: *Human Walking in Virtual Environments: Perception, Technology, and Applications*. A cura di Frank Steinicke, Yon Visell, Jennifer Campos e Anatole Lécuyer. New York, NY: Springer New York, 2013, pp. 3–26. ISBN: 978-1-4419-8432-6. DOI: 10.1007/978-1-4419-8432-6_1. URL: https://doi.org/10.1007/978-1-4419-8432-6_1 (cit. a p. 11).
- [25] Evan A. Suma, Sabarish Babu e Larry F. Hodges. «Comparison of Travel Techniques in a Complex, Multi-Level 3D Environment». In: *2007 IEEE Symposium on 3D User Interfaces*. 2007. DOI: 10.1109/3DUI.2007.340788 (cit. a p. 11).
- [26] Alberto Cannavò, Davide Calandra, Filippo Gabriele Praticò, Valentina Gatteschi e F. Lamberti. «An Evaluation Testbed for Locomotion in Virtual Reality». In: *IEEE Transactions on Visualization and Computer Graphics* 27 (ott. 2020), pp. 1871–1889. DOI: 10.1109/TVCG.2020.3032440 (cit. alle pp. 11, 25).
- [27] Tuncay Cakmak e Holger Hager. «Cyberith virtualizer: a locomotion device for virtual reality». In: (lug. 2014). DOI: 10.1145/2614066.2614105 (cit. a p. 12).
- [28] Alexandra Kitson, Abraham M. Hashemian, Ekaterina R. Stepanova, Ernst Kruijff e Bernhard E. Riecke. «Comparing leaning-based motion cueing interfaces for virtual reality locomotion». In: *2017 IEEE Symposium on 3D User Interfaces (3DUI)*. 2017, pp. 73–82. DOI: 10.1109/3DUI.2017.7893320 (cit. a p. 12).

- [29] Jeff Feasel, Mary C. Whitton e Jeremy D. Wendt. «LLCM-WIP: Low-Latency, Continuous-Motion Walking-in-Place». In: *2008 IEEE Symposium on 3D User Interfaces*. 2008, pp. 97–104. DOI: 10.1109/3DUI.2008.4476598 (cit. a p. 12).
- [30] Yun Suen Pai e Kai Kunze. «Armswing: Using Arm Swings for Accessible and Immersive Navigation in AR/VR Spaces». In: *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia*. MUM '17. Stuttgart, Germany: Association for Computing Machinery, 2017, pp. 189–198. ISBN: 9781450353786. DOI: 10.1145/3152832.3152864. URL: <https://doi.org/10.1145/3152832.3152864> (cit. a p. 12).
- [31] Evan A. Suma, Zachary Lipps, Samantha Finkelstein, David M. Krum e Mark Bolas. «Impossible Spaces: Maximizing Natural Walking in Virtual Environments with Self-Overlapping Architecture». In: *IEEE Transactions on Visualization and Computer Graphics* 18.4 (2012), pp. 555–564. DOI: 10.1109/TVCG.2012.47 (cit. a p. 12).
- [32] Shunichi Kasahara, Keina Konno, Richi Owaki, Tsubasa Nishi, Akiko Takeshita, Takayuki Ito, Shoko Kasuga e Junichi Ushiba. «Malleable Embodiment: Changing Sense of Embodiment by Spatial-Temporal Deformation of Virtual Human Body». In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. CHI '17. Denver, Colorado, USA: Association for Computing Machinery, 2017, pp. 6438–6448. ISBN: 9781450346559. DOI: 10.1145/3025453.3025962. URL: <https://doi.org/10.1145/3025453.3025962> (cit. a p. 12).
- [33] Fan Jiang, Xubo Yang e Lele Feng. «Real-Time Full-Body Motion Reconstruction and Recognition for off-the-Shelf VR Devices». In: *Proceedings of the 15th ACM SIGGRAPH Conference on Virtual-Reality Continuum and Its Applications in Industry - Volume 1*. VRCAI '16. Zhuhai, China: Association for Computing Machinery, 2016, pp. 309–318. ISBN: 9781450346924. DOI: 10.1145/3013971.3013987. URL: <https://doi.org/10.1145/3013971.3013987> (cit. alle pp. 12, 13).
- [34] Mel Slater e Sylvia Wilbur. «A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments». In: *Presence: Teleoperators and Virtual Environments* 6.6 (dic. 1997), pp. 603–616. DOI: 10.1162/pres.1997.6.6.603. eprint: <https://direct.mit.edu/pvar/article-pdf/6/6/603/1623151/pres.1997.6.6.603.pdf>. URL: <https://doi.org/10.1162/pres.1997.6.6.603> (cit. a p. 12).
- [35] Henrique G. Debarba, Sami Perrin, Bruno Herbelin e Ronan Boulic. «Embodied Interaction Using Non-Planar Projections in Immersive Virtual Reality». In: *Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology*. VRST '15. Beijing, China: Association for Computing Machinery, 2015, pp. 125–128. ISBN: 9781450339902. DOI: 10.1145/2821592.2821603. URL: <https://doi.org/10.1145/2821592.2821603> (cit. a p. 12).

- [36] Polona Caserman, Augusto Garcia-Agundez e Stefan Göbel. «A Survey of Full-Body Motion Reconstruction in Immersive Virtual Reality Applications». In: *IEEE Transactions on Visualization and Computer Graphics* 26.10 (2020), pp. 3089–3108. DOI: 10.1109/TVCG.2019.2912607 (cit. a p. 12).
- [37] Paolo Leoncini, Bogdan Sikorski, Vincenzo Baraniello, Francesco Martone, Carlo Luongo e Mariano Guida. «Multiple NUI Device Approach to Full Body Tracking for Collaborative Virtual Environments». In: *Augmented Reality, Virtual Reality, and Computer Graphics*. A cura di Lucio Tommaso De Paolis, Patrick Bourdot e Antonio Mongelli. Cham: Springer International Publishing, 2017, pp. 131–147. ISBN: 978-3-319-60922-5 (cit. a p. 12).
- [38] A. Aristidou, J. Lasenby, Y. Chrysanthou e A. Shamir. «Inverse Kinematics Techniques in Computer Graphics: A Survey». In: *Computer Graphics Forum* 37.6 (2018), pp. 35–58. DOI: <https://doi.org/10.1111/cgf.13310>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.13310> (cit. a p. 13).
- [39] Polona Caserman, Philipp Achenbach e Stefan Göbel. «Analysis of Inverse Kinematics Solutions for Full-Body Reconstruction in Virtual Reality». In: *2019 IEEE 7th International Conference on Serious Games and Applications for Health (SeGAH)*. 2019, pp. 1–8. DOI: 10.1109/SeGAH.2019.8882429 (cit. alle pp. 13, 14).
- [40] Michael Meredith e Steve Maddock. «Real-time inverse kinematics: The return of the jacobian». In: (2004) (cit. a p. 13).
- [41] Andreas Aristidou e Joan Lasenby. «FABRIK: A fast, iterative solver for the Inverse Kinematics problem». In: *Graphical Models* 73.5 (2011), pp. 243–260. ISSN: 1524-0703. DOI: <https://doi.org/10.1016/j.gmod.2011.05.003>. URL: <https://www.sciencedirect.com/science/article/pii/S1524070311000178> (cit. a p. 13).
- [42] Michael N Louka e Claudio Balducelli. «Virtual reality tools for emergency operation support and training». In: *Proceedings of TIEMS (The International Emergency Management Society), Oslo* (2001) (cit. a p. 14).
- [43] Ruggiero Lovreglio. «Virtual and Augmented reality for human behaviour in disasters: a review». In: *Fire and Evacuation Modeling Technical Conference (FEMTC) 2020 Online Conference*. 2020, pp. 9–11 (cit. a p. 14).
- [44] Filippo Gabriele Praticò, Federico De Lorenzis, Davide Calandra, Alberto Cannavò e Fabrizio Lamberti. «Exploring Simulation-Based Virtual Reality as a Mock-Up Tool to Support the Design of First Responders Training». In: *Applied Sciences* 11.16 (2021). ISSN: 2076-3417. DOI: 10.3390/app11167527. URL: <https://www.mdpi.com/2076-3417/11/16/7527> (cit. a p. 14).

- [45] Huyen Nguyen e Tomasz Bednarz. «User Experience in Collaborative Extended Reality: Overview Study». In: *Virtual Reality and Augmented Reality*. A cura di Patrick Bourdot, Victoria Interrante, Regis Kopper, Anne-Hélène Olivier, Hideo Saito e Gabriel Zachmann. Cham: Springer International Publishing, 2020, pp. 41–70. ISBN: 978-3-030-62655-6 (cit. a p. 14).
- [46] Jeremy N Bailenson, Nick Yee, Dan Merget e Ralph Schroeder. «The Effect of Behavioral Realism and Form Realism of Real-Time Avatar Faces on Verbal Disclosure, Nonverbal Disclosure, Emotion Recognition, and Copresence in Dyadic Interaction». In: *Presence: Teleoperators and Virtual Environments* 15.4 (ago. 2006), pp. 359–372. DOI: 10.1162/pres.15.4.359. eprint: <https://direct.mit.edu/pvar/article-pdf/15/4/359/1624480/pres.15.4.359.pdf>. URL: <https://doi.org/10.1162/pres.15.4.359> (cit. a p. 14).
- [47] Daniel Roth, Jean-Luc Lugin, Dmitri Galakhov, Arvid Hofmann, Gary Bente, Marc Erich Latoschik e Arnulph Fuhrmann. «Avatar realism and social interaction quality in virtual reality». In: *2016 IEEE Virtual Reality (VR)*. 2016, pp. 277–278. DOI: 10.1109/VR.2016.7504761 (cit. a p. 14).
- [48] Vlasios Kasapakis e Elena Dzardanova. «Using High Fidelity Avatars to Enhance Learning Experience in Virtual Learning Environments». In: *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 2021, pp. 645–646. DOI: 10.1109/VRW52623.2021.00205 (cit. a p. 15).
- [49] Djamel Eddine Benrachou, Mostefa Masmoudi, Oualid Djekoune, Nadia Zenati e Mehdi Ousmer. «Avatar-Facilitated Therapy and Virtual Reality: Next-Generation of Functional Rehabilitation Methods». In: *2020 1st International Conference on Communications, Control Systems and Signal Processing (CCSSP)*. 2020, pp. 298–304. DOI: 10.1109/CCSSP49278.2020.9151528 (cit. a p. 15).
- [50] Mar Gonzalez-Franco, Anthony Steed, Steve Hoogendyk e Eyal Ofek. «Using Facial Animation to Increase the Enfacement Illusion and Avatar Self-Identification». In: *IEEE Transactions on Visualization and Computer Graphics* 26.5 (2020), pp. 2023–2029. DOI: 10.1109/TVCG.2020.2973075 (cit. a p. 16).
- [51] Charles Darwin. «The Expression of Emotion in Man and Animals». In: Penguin Books Ltd (2009) (cit. alle pp. 16, 17).
- [52] Jianwen Lou, Yiming Wang, Charles Nduka, Mahyar Hamedi, Ifigeneia Mavridou, Fei-Yue Wang e Hui Yu. «Realistic Facial Expression Reconstruction for VR HMD Users». In: *IEEE Transactions on Multimedia* 22.3 (2020), pp. 730–743. DOI: 10.1109/TMM.2019.2933338 (cit. a p. 16).

- [53] T. Ishikawa, H. Sera, S. Morishima e D. Terzopoulos. «Facial image reconstruction by estimated muscle parameter». In: *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition*. 1998, pp. 342–347. DOI: 10.1109/AFGR.1998.670972 (cit. a p. 16).
- [54] N. Eveno, A. Caplier e P.-Y. Coulon. «Accurate and quasi-automatic lip tracking». In: *IEEE Transactions on Circuits and Systems for Video Technology* 14.5 (2004), pp. 706–715. DOI: 10.1109/TCSVT.2004.826754 (cit. a p. 16).
- [55] Fumihiko Nakamura, Masaaki Murakami, Katsuhiro Suzuki, Masaaki Fukuoka, Katsutoshi Masai e Maki Sugimoto. «Analyzing the Effect of Diverse Gaze and Head Direction on Facial Expression Recognition with Photo-Reflective Sensors Embedded in a Head-Mounted Display». In: *IEEE Transactions on Visualization and Computer Graphics* (2022), pp. 1–1. DOI: 10.1109/TVCG.2022.3179766 (cit. a p. 16).
- [56] Niklas Stein, Diederick C. Niehorster, Tamara Watson, Frank Steinicke, Katharina Rifai, Siegfried Wahl e Markus Lappe. «A Comparison of Eye Tracking Latencies Among Several Commercial Head-Mounted Displays». In: *i-Perception* 12.1 (2021). DOI: 10.1177/2041669520983338. URL: <https://doi.org/10.1177/2041669520983338> (cit. a p. 16).
- [57] Masahiro Mori, Karl F. MacDorman e Norri Kageki. «The Uncanny Valley [From the Field]». In: *IEEE Robotics & Automation Magazine* 19.2 (2012), pp. 98–100. DOI: 10.1109/MRA.2012.2192811 (cit. a p. 16).
- [58] Catrin Misselhorn. «Empathy with Inanimate Objects and the Uncanny Valley». In: *Minds and Machines* 19.3 (2009), pp. 1572–8641. DOI: 10.1007/s11023-009-9158-2 (cit. a p. 17).
- [59] Tom Geller. «Overcoming the Uncanny Valley». In: *IEEE Computer Graphics and Applications* 28.4 (2008), pp. 11–17. DOI: 10.1109/MCG.2008.79 (cit. a p. 17).
- [60] David Hanson, Andrew Olney, Steve Prilliman, Eric Mathews, Marge Zielke, Derek Hammons, Raul Fernandez e Harry Stephanou. «Upending the uncanny valley». In: *AAAI*. Vol. 5. 2005, pp. 1728–1729 (cit. a p. 17).
- [61] Shensheng Wang, Scott O. Lilienfeld e Philippe Rochat. «The Uncanny Valley: Existence and Explanations». In: *Review of General Psychology* 19.4 (2015), pp. 393–407. DOI: 10.1037/gpr0000056 (cit. a p. 17).
- [62] Dawid Ratajczyk. «Uncanny Valley in Video Games: An Overview». In: *Homo Ludens* (dic. 2019), pp. 135–148. DOI: 10.14746/hl.2019.12.7 (cit. a p. 17).
- [63] D. Grandjean e K. R. Scherer. «Unpacking the cognitive architecture of emotion processes.» In: *Emotion*, 8(3) (2008), pp. 341–351. DOI: 10.1037/1528-3542.8.3.341. URL: <https://doi.org/10.1037/1528-3542.8.3.341> (cit. alle pp. 17, 18).

- [64] Ruud Hortensius, Felix Hekele e Emily S. Cross. «The Perception of Emotion in Artificial Agents». In: *IEEE Transactions on Cognitive and Developmental Systems* 10.4 (2018), pp. 852–864. DOI: 10.1109/TCDS.2018.2826921 (cit. a p. 18).
- [65] Rosalind W. Picard. *Affective computing*. MIT press, 2000 (cit. a p. 18).
- [66] Klaus R. Scherer. «Measuring the meaning of emotion words: A domain-specific componential approach». In: *Components of Emotional Meaning: A sourcebook*. Oxford University Press, ago. 2013. ISBN: 9780199592746. DOI: 10.1093/acprof:oso/9780199592746.003.0002. URL: <https://doi.org/10.1093/acprof:oso/9780199592746.003.0002> (cit. a p. 18).
- [67] Carroll E. Izard. «Emotion Theory and Research: Highlights, Unanswered Questions, and Emerging Issues». In: *Annual Review of Psychology* 60.1 (2009). PMID: 18729725, pp. 1–25. DOI: 10.1146/annurev.psych.60.110707.163539. URL: <https://doi.org/10.1146/annurev.psych.60.110707.163539> (cit. a p. 18).
- [68] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt e Ioannis Patras. «DEAP: A Database for Emotion Analysis ;Using Physiological Signals». In: *IEEE Transactions on Affective Computing* 3.1 (2012), pp. 18–31. DOI: 10.1109/T-AFFC.2011.15 (cit. a p. 18).
- [69] S. Ojha, J. Vitale e M.A. Williams. «Computational Emotion Models: A Thematic Review». In: *Int J of Soc Robotics*.13 (2021), pp. 1253–1279. DOI: 10.1007/s12369-020-00713-1 (cit. a p. 18).
- [70] Gelareh Mohammadi, Dimitri Van De Ville e Patrik Vuilleumier. «Brain networks subserving functional core processes of emotions identified with componential modelling». In: *bioRxiv* (2020). DOI: 10.1101/2020.06.10.145201. URL: <https://www.biorxiv.org/content/early/2020/06/12/2020.06.10.145201> (cit. a p. 18).
- [71] Paul Ekman. «Are there basic emotions?» In: (1992) (cit. alle pp. 18, 55, 61).
- [72] Jan Mizgajski e Mikolaj Morzy. «Affective recommender systems in online news industry: how emotions influence reading choices». In: *User Modeling and User-Adapted Interaction* 29 (apr. 2019). DOI: 10.1007/s11257-018-9213-x (cit. a p. 18).
- [73] Rukshani Somarathna, Tomasz Bednarz e Gelareh Mohammadi. «Virtual Reality for Emotion Elicitation - A Review». In: *IEEE Transactions on Affective Computing* (2022), pp. 1–21. DOI: 10.1109/TAFFC.2022.3181053 (cit. a p. 19).
- [74] Ana Paiva, Iolanda Leite e Tiago Ramos Ribeiro. «Emotion Modelling for Social Robots». In: 2014 (cit. a p. 19).

- [75] E.S. Dan-Glauser e K.R. Scherer. «The Geneva affective picture database (GAPED): a new 730-picture database focusing on valence and normative significance». In: *Behav Res.*43 (2011), p. 468. DOI: 10.3758/s13428-011-0064-1. URL: <https://doi.org/10.3758/s13428-011-0064-1> (cit. a p. 19).
- [76] Jacqueline M Kory e Sidney K D’Mello. «27 Affect Elicitation for Affective Computing». In: *The Oxford handbook of affective computing* (2014), p. 371 (cit. a p. 19).
- [77] Simo Järvelä, Benjamin Cowley, Mikko Salminen, Giulio Jacucci, Juho Hamari e Niklas Ravaja. «Augmented Virtual Reality Meditation: Shared Dyadic Biofeedback Increases Social Presence Via Respiratory Synchrony». In: *Trans. Soc. Comput.* 4.2 (mag. 2021). ISSN: 2469-7818. DOI: 10.1145/3449358. URL: <https://doi.org/10.1145/3449358> (cit. a p. 20).
- [78] Xiaolan Peng, Jin Huang, Alena Denisova, Hui Chen, Feng Tian e Hongan Wang. «A Palette of Deepened Emotions: Exploring Emotional Challenge in Virtual Reality Games». In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI ’20. Honolulu, HI, USA: Association for Computing Machinery, 2020, pp. 1–13. ISBN: 9781450367080. DOI: 10.1145/3313831.3376221. URL: <https://doi.org/10.1145/3313831.3376221> (cit. a p. 20).
- [79] Federica Pallavicini, Alessandro Pepe et al. «Virtual reality games and the role of body involvement in enhancing positive emotions and decreasing anxiety: within-subjects pilot study». In: *JMIR serious games* 8.2 (2020), e15635 (cit. a p. 20).
- [80] Meriem El-Yamri, Alejandro Romero-Hernandez, Manuel Gonzalez-Riojo e Borja Manero. «Emotions-Responsive Audiences for VR Public Speaking Simulators Based on the Speakers’ Voice». In: *2019 IEEE 19th International Conference on Advanced Learning Technologies (ICALT)*. Vol. 2161-377X. 2019, pp. 349–353. DOI: 10.1109/ICALT.2019.00108 (cit. a p. 20).
- [81] Camilo Rojas, Malena Corral, Niels Poulsen e Pattie Maes. «Project Us: A Wearable for Enhancing Empathy». In: *Companion Publication of the 2020 ACM Designing Interactive Systems Conference*. DIS’ 20 Companion. Eindhoven, Netherlands: Association for Computing Machinery, 2020, pp. 139–144. ISBN: 9781450379878. DOI: 10.1145/3393914.3395882. URL: <https://doi.org/10.1145/3393914.3395882> (cit. a p. 20).
- [82] Iulia Stanica, Maria-Iuliana Dascalu, Constanta Nicoleta Bodea e Alin Dragos Bogdan Moldoveanu. «VR Job Interview Simulator: Where Virtual Reality Meets Artificial Intelligence for Education». In: *2018 Zooming Innovation in Consumer Technologies Conference (ZINC)*. 2018, pp. 9–12. DOI: 10.1109/ZINC.2018.8448645 (cit. a p. 20).

- [83] J. Gutiérrez-Maldonado, M. Rus-Calafell e J. González-Conde. «Creation of a new set of dynamic virtual reality faces for the assessment and training of facial emotion recognition ability». In: *Virtual Reality* 18 (2014), pp. 61–71. DOI: 10.1007/s10055-013-0236-7. URL: <https://doi.org/10.1007/s10055-013-0236-7> (cit. a p. 20).
- [84] Esperanza Johnson, Ramón Hervás, Carlos Gutiérrez López de la Franca, Tania Mondéjar, Sergio F Ochoa e Jesús Favela. «Assessing empathy and managing emotions through interactions with an affective avatar». In: *Health Informatics Journal* 24.2 (2018), pp. 182–193. DOI: 10.1177/1460458216661864. URL: <https://doi.org/10.1177/1460458216661864> (cit. a p. 20).
- [85] E. Bekele, D. Bian, J. Peterman, S. Park e N. Sarkar. «Design of a Virtual Reality System for Affect Analysis in Facial Expressions (VR-SAAFE); Application to Schizophrenia». In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 25.6 (2017), pp. 739–749. DOI: 10.1109/TNSRE.2016.2591556 (cit. a p. 20).
- [86] Natalie Hube, Kresimir Vidackovic e Michael Sedlmair. «Using Expressive Avatars to Increase Emotion Recognition: A Pilot Study». In: *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*. CHI EA '22. New Orleans, LA, USA: Association for Computing Machinery, 2022. ISBN: 9781450391566. DOI: 10.1145/3491101.3519822. URL: <https://doi.org/10.1145/3491101.3519822> (cit. a p. 20).
- [87] Pamela Perniss. «Why We Should Study Multimodal Language». In: *Frontiers in Psychology* 9 (2018). ISSN: 1664-1078. DOI: 10.3389/fpsyg.2018.01109. URL: <https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01109> (cit. a p. 20).
- [88] D. Peeters. «Virtual reality: A game-changing method for the language sciences». In: *Psychon Bull Rev*.26 (2019), pp. 894–900. DOI: 10.3758/s13423-019-01571-3 (cit. a p. 20).
- [89] N. Eichert, D. Peeters e P. Hagoort. «Language-driven anticipatory eye movements in virtual reality». In: *Behav Res*.50 (2018), pp. 1102–1115. DOI: 10.3758/s13428-017-0929-z (cit. a p. 21).
- [90] DAVID PEETERS e TON DIJKSTRA. «Sustained inhibition of the native language in bilingual language production: A virtual reality approach». In: *Bilingualism: Language and Cognition* 21.5 (2018), pp. 1035–1061. DOI: 10.1017/S1366728917000396 (cit. a p. 21).
- [91] Tom Gijssels, Laura Staum Casasanto, Kyle Jasmin, Peter Hagoort e Daniel Casasanto. «Speech Accommodation Without Priming: The Case of Pitch». In: *Discourse Processes* 53.4 (2016), pp. 233–251. DOI: 10.1080/0163853X.2015.1023965. URL: <https://doi.org/10.1080/0163853X.2015.1023965> (cit. a p. 21).

- [92] E. Heyselaar, P. Hagoort e K. Segaert. «In dialogue with an avatar, language behavior is identical to dialogue with a human partner». In: *Behav Res.* 49 (2017), pp. 46–60. DOI: 10.3758/s13428-015-0688-7 (cit. a p. 21).
- [93] UV Ramesh, A Harini, Ch Sri Divya Gowri, K Vyshnavi Durga, P Druvitha e K Siva Kumar. «Metaverse: future of the internet». In: *Journal homepage: www.ijrpr.com ISSN 2582* (2022), p. 7421 (cit. a p. 21).
- [94] Stylianos Mystakidis. «Metaverse». In: *Encyclopedia 2.1* (2022), pp. 486–497 (cit. a p. 21).
- [95] Yuntao Wang, Zhou Su, Ning Zhang, Rui Xing, Dongxiao Liu, Tom H Luan e Xuemin Shen. «A survey on metaverse: Fundamentals, security, and privacy». In: *IEEE Communications Surveys & Tutorials* (2022) (cit. alle pp. 21, 22).
- [96] Li Jiaxin e Gao Gongjing. «Socializing in the Metaverse: The Innovation and Challenge of Interpersonal Communication». In: *Proceedings of the 2022 8th International Conference on Humanities and Social Science Research (ICHSSR 2022)*. Atlantis Press, 2022, pp. 2128–2131. ISBN: 978-94-6239-580-0. DOI: <https://doi.org/10.2991/assehr.k.220504.385>. URL: <https://doi.org/10.2991/assehr.k.220504.385> (cit. a p. 21).
- [97] Davide Calandra, Michele Billi, Fabrizio Lamberti, Andrea Sanna e Romano Borchellini. «Arm Swinging vs Treadmill: A Comparison Between Two Techniques for Locomotion in Virtual Reality». In: *EG 2018 - Short Papers*. A cura di Olga Diamanti e Amir Vaxman. The Eurographics Association, 2018. DOI: 10.2312/egs.20181043 (cit. a p. 25).
- [98] Davide Calandra, Fabrizio Lamberti e Massimo Migliorini. «On the Usability of Consumer Locomotion Techniques in Serious Games: Comparing Arm Swinging, Treadmills and Walk-in-Place». In: *2019 IEEE 9th International Conference on Consumer Electronics (ICCE-Berlin)*. 2019, pp. 348–352. DOI: 10.1109/ICCE-Berlin47944.2019.8966165 (cit. a p. 25).
- [99] Mar Gonzalez-Franco e Tabitha C. Peck. «Avatar Embodiment. Towards a Standardized Questionnaire». In: *Frontiers in Robotics and AI* 5 (2018). ISSN: 2296-9144. DOI: 10.3389/frobt.2018.00074. URL: <https://www.frontiersin.org/articles/10.3389/frobt.2018.00074> (cit. alle pp. 34, 38).
- [100] Frank Biocca, Chad Harms e Jenn Gregg. «The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity». In: *4th annual international workshop on presence, Philadelphia, PA*. 2001, pp. 1–9 (cit. alle pp. 34, 39).
- [101] Roy S. Kalawsky. «VRUSE—a computerised diagnostic tool: for usability evaluation of virtual/synthetic environment systems». In: *Applied Ergonomics* 30.1 (1999), pp. 11–25. ISSN: 0003-6870. DOI: [https://doi.org/10.1016/S0003-6870\(98\)00047-7](https://doi.org/10.1016/S0003-6870(98)00047-7). URL: <https://www.sciencedirect.com/science/article/pii/S0003687098000477> (cit. alle pp. 35, 40).

- [102] Jean-Luc Lugin et al. «Any “Body” There? Avatar Visibility Effects in a Virtual Reality Game». In: *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 2018, pp. 17–24. DOI: 10.1109/VR.2018.8446229 (cit. a p. 38).
- [103] Elena Kokkinara e Mel Slater. «Measuring the Effects through Time of the Influence of Visuomotor and Visuotactile Synchronous Stimulation on a Virtual Body Ownership Illusion». In: *Perception* 43.1 (2014), pp. 43–58. DOI: 10.1068/p7545 (cit. a p. 38).
- [104] James A Russell, Anna Weiss e Gerald A Mendelsohn. «Affect grid: a single-item scale of pleasure and arousal.» In: *Journal of personality and social psychology* 57.3 (1989), p. 493 (cit. a p. 55).
- [105] *SALSA Lip-Sync Online Documentation by Crazy Minnow Studio*. URL: <https://crazyminnowstudio.com/docs/salsa-lip-sync/> (cit. a p. 64).

Ringraziamenti

In questa ultima pagina utile della presente tesi, vorrei riservare qualche parola di ringraziamento e dedica a tutti coloro che mi hanno accompagnato in questo lungo e tortuoso percorso universitario magistrale.

E' stato un periodo duro, che mi ha messo duramente e continuamente alla prova e che col tempo ha portato la fiamma della mia motivazione a spegnersi sempre di più. L'arrivo poi della pandemia di Covid-19 e la solitudine che poi è seguita nella mia vita qui a Torino sembravano aver dato quell'ultimo soffio fatale alla ormai piccola fiammella rimasta in me, ma fortunatamente così non è stato, specialmente grazie alle persone qui di seguito.

Ringrazio il Prof. Lamberti, per la sua guida da docente e successivamente nel ruolo di relatore, ma soprattutto per il non avermi, nonostante tutto ciò che si è susseguito negli ultimi anni, mai tolto l'opportunità di portare a compimento tale tesi e per essere stato comprensivo e disponibile nei miei periodi più duri.

Ringrazio anche Davide, dottorando e co-relatore della presente tesi, per tutte le peripezie vissute in questi due anni di percorso insieme. Farò tesoro di tutti gli insegnamenti che mi hai dato, ma soprattutto di tutti i ricordi creati insieme in questa esperienza. Grazie anche ad Alessandro, Federico, Alberto e anche a Gabriele, per l'aiuto quotidiano, le risate, i ricordi e l'avermi sopportato nel Lab 2 per fin troppo tempo.

Vorrei poi rivolgere un mio più grande e sentito grazie a Vittoria, Alessia, Valentina, Tommaso e Mario, amici online che mi hanno spesso tenuto compagnia nelle serate solitarie di Torino. Altrettanto sentiti "grazie" vanno anche a Marl, Veronica, Erica, Giada, Flavia, Jack, Ava e Manu, da cui ho ricevuto sempre enorme supporto seppur tantissimi fossero i chilometri che ci separavano nella realtà.

E infine, i ringraziamenti più importanti li rivolgo al mio amico di una vita, Matteo, e alla mia famiglia, che non hanno mai smesso, neppure per un secondo, di farmi sentire amato, sostenuto e, soprattutto, mai da solo in questo lungo e tortuoso percorso di tesi.

A voi voglio dedicare questo lavoro. Vi voglio bene, e vi sarò per sempre riconoscente per tutto ciò che avete portato e che porterete nella mia vita. Grazie.

Gianmario Lupini

