

# POLITECNICO DI TORINO

Laurea Magistrale in  
Ingegneria del Cinema e dei Mezzi di Comunicazione



**Politecnico  
di Torino**

Tesi di Laurea Magistrale

## Avatar espressivi per applicazioni di Realtà Virtuale

Relatori

Prof. Andrea BOTTINO

Dott. Francesco STRADA

Dott. Edoardo BATTEGAZZORRE

Candidato

**Stefano CALZOLARI**

Aprile 2023



# Sommario

Negli ultimi anni numerose tecnologie sono approdate sul mercato della Realtà Virtuale (VR), tema discusso e studiato in ambito accademico da decenni ma che ha ricevuto l'attenzione dei mass-media solo negli ultimi anni grazie anche al concetto di Metaverso. Questi strumenti hanno permesso di prototipare e sviluppare esperienze sempre più immersive che trovano applicazioni nei settori più vari, dal training in ambito sanitario all'intrattenimento, come il gaming o le piattaforme social VR.

L'obiettivo di questa tesi è di valutare come questi strumenti, possano portare un'innovazione in applicazioni di training o simulazione, per realizzare assistenti, istruttori o guide virtuali che accompagnano l'utente nelle task da svolgere. Queste entità assumono in letteratura vari nomi: Virtual Humans (VH), Virtual Agents, Digital Humans, Embodied Agents o semplicemente Avatar. Il loro scopo è quello di replicare il comportamento, entro certi limiti, di una persona reale ma in uno scenario virtuale.

Affinchè ciò accada è necessario che questi possano esprimere la loro "umanità" attraverso una qualche emulazione della sfera emotiva umana, instaurando così un legame empatico tra utente e VH. È dimostrato come questo legame possa determinare un maggior grado di coinvolgimento dell'utente nello svolgimento dei task e soprattutto un conseguente miglior apprendimento.

In letteratura è evidente come diversi aspetti concorrano nel simulare questo comportamento quali: interazione (bilaterale utente-VH, VH-ambiente virtuale), comportamento (movimento nell'ambiente) ed espressività facciale. Questa tesi di laurea si concentrerà proprio su quest'ultimo aspetto.

La tesi mira in primo luogo ad integrare il sistema FACS di animazione facciale per ottenere quello che è lo stato dell'arte della rappresentazione

delle espressioni. In secondo luogo la tesi valuta se questo sistema di animazione possa essere utilizzato per la rappresentazione dell'espressività in maniera univoca, e sostituendo così gli attuali metodi, molto più complessi, costosi e poco scalabili.



# Indice

<b>1</b>	<b>Introduzione</b>	<b>1</b>
1.1	Perchè i Virtual Humans? . . . . .	2
1.2	The <i>Uncanny Valley</i> . . . . .	4
1.3	Fotorealismo . . . . .	6
1.4	Emozioni ed Espressività Facciale . . . . .	7
1.5	Facial Action Coding System . . . . .	10
1.6	Obiettivi . . . . .	12
<b>2</b>	<b>Stato dell'arte</b>	<b>13</b>
2.1	Virtual Humans ed Empatia . . . . .	13
2.1.1	FACS . . . . .	15
2.2	Unreal Engine 5 . . . . .	16
2.3	MetaHuman . . . . .	17
2.3.1	Pipeline di produzione . . . . .	18
2.3.2	Level Of Details . . . . .	19
2.3.3	Animazione & Rig Logic . . . . .	20
<b>3</b>	<b>Metodologia</b>	<b>23</b>
3.1	FACS Action Units . . . . .	25
3.2	Generazione Weights Sequence . . . . .	26
3.3	Weights JSON file . . . . .	27
3.4	Mapping JSON file . . . . .	27
3.5	Animation Blueprint . . . . .	28
<b>4</b>	<b>Sperimentazione</b>	<b>31</b>
4.1	Protocollo . . . . .	32
4.1.1	Raccolta Dati . . . . .	32

4.1.2	Analisi Dati . . . . .	32
4.2	Applicazione Web . . . . .	34
4.2.1	Design . . . . .	34
4.2.2	Algoritmo <i>Round Robin</i> . . . . .	36
4.3	Sistema di Ranking . . . . .	38
<b>5</b>	<b>Risultati e Discussione</b>	<b>40</b>
5.1	Analisi dei risultati . . . . .	40
5.1.1	Risultati <i>Happiness</i> . . . . .	40
5.1.2	Risultati <i>Surprise</i> . . . . .	43
5.2	Limiti di ricerca . . . . .	45
<b>6</b>	<b>Conclusioni</b>	<b>48</b>
6.1	Ricerca futura . . . . .	48
<b>A</b>	<b>Disposizioni</b>	<b>52</b>
A.1	Disposizioni semplici . . . . .	52
A.2	Disposizioni con ripetizione . . . . .	52
<b>B</b>	<b>Kendall Tau Test</b>	<b>53</b>
B.1	Test A . . . . .	53
B.2	Test B . . . . .	54
	<b>Bibliografia e Sitografia</b>	<b>56</b>

# Capitolo 1

## Introduzione

La *Human Computer Interaction* è una branca di ricerca dell'informatica che sta ricevendo sempre più interesse negli ultimi anni grazie alla diffusione su larga scala delle tecnologie elettroniche ed informatiche, sia per quanto riguarda l'industria *business-to-business* che quella *business-to-consumer*. L'interazione tra l'utente ed un computer, smartphone, o totem che eroga ad esempio servizi bancari o sanitari è sempre più richiesta. In particolare negli ultimi anni questo campo di studi ha trovato una grande applicazione nelle moderne tecnologie di realtà virtuale (VR), anch'esse diventate disponibili per un pubblico più vasto e per diversi scopi: training (tra i più importanti industriale, medicale e militare), entertainment (videogames e social platforms) ed educational.

Applicazioni VR di questo tipo necessitano di ambienti virtuali realizzati *ad hoc* per il tipo di simulazione necessaria, in primis per quello che riguarda come l'utente può interagire con gli oggetti e le entità virtuali presenti. In particolare molti di questi necessitano di simulare il comportamento e l'interazione dell'utente con altre persone che siano ad esempio pazienti, medici o infermieri nel caso di simulazioni medicali. Questi ruoli possono essere svolti da un sistema di *Artificial Intelligence* (AI) che imita entro certi limiti stabiliti dall'applicativo, il comportamento di una persona umana, prendendo il nome di Virtual Humans (VH).

I Virtual Humans, anche chiamati in letteratura *Digital Humans*, *Embodied Conversational Agents*, *Virtual Agents* o semplicemente *Avatars* idealmente riproducono un'essere umano in carne ed ossa ma in virtuale,



**Figura 1.1:** Esempi di utilizzo dei Virtual Humans

governati direttamente dalle capacità computazionali di un PC. Chiaramente il loro stato dell'arte è ancora lontano da soluzioni di realismo estremo, ci si concentra solitamente ad implementare quelle specifiche funzionalità comportamentali richieste dalla simulazione. Grazie però alle moderne soluzioni di computer grafica come l'avvento di tecnologie quali il *ray-tracing* in *real-time*, i recenti sviluppi in materia di *Machine Learning* e *Affective Computing* come GPT-3 e GPT-4 (vedi ChatGPT) e il forte interesse per la VR (ricordiamo anche il concetto di Metaverso), si sono aperti innovativi percorsi di ricerca sui VH.

Con questo focus, la presente tesi si propone l'obiettivo di analizzare l'attuale stato dell'arte dei *Virtual Humans* (VH) e proporre un framework sperimentale per lo studio della loro capacità espressiva.

## 1.1 Perché i Virtual Humans?

I VH sono necessari a simulare delle persone reali che per motivi di monitoraggio o limiti tecnici non è possibile inserire all'interno della simulazione. Un VH è un'entità totalmente gestita dalla CPU e per questo ha un comportamento controllabile sia in termini di movimento che di misurazione dei parametri che ne caratterizzano l'interazione con

l'utente 1.2. In una simulazione di negoziazione militare ad esempio il comportamento del soldato virtuale è interamente controllabile da chi sviluppa la simulazione al fine di adattarlo all'operazione richiesta. Nel momento in cui l'utente, che si suppone essere l'ufficiale che sta eseguendo il training, interviene sul VH, questa interazione è quantificabile direttamente dalla simulazione.

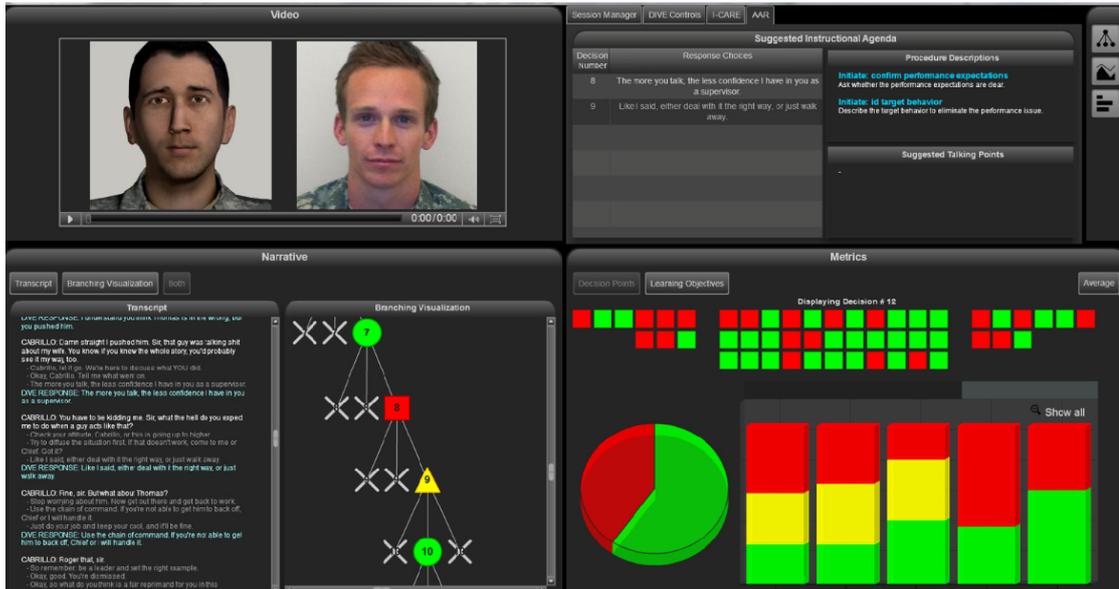


Figura 1.2: Esempio app VR di training militare con statistiche [1]

Inoltre in casi di simulazioni VR di scenari con alto rischio per l'incolumità dei partecipanti (ad esempio training dei vigili del fuoco, rischiose operazioni mediche, operazioni militari) i VH sono fondamentali per simulare il comportamento dei soggetti nello scenario reale.

Realizzare un VH non è un'operazione semplice, richiede l'applicazione di numerose tecniche e l'implementazione di software appositi per ogni singola applicazione. Considerando la caratterizzazione comportamentale dell'essere umano (quella osservabile dall'esterno ma anche la sua dimensione interiore ed emotiva) e la sua multimodalità dei sensi quali: udito, vista, tatto, olfatto e gusto; si può intuire quale sia la complessità che caratterizza un VH. Esaminando le esigenze delle applicazioni VR attuali ed i limiti hardware per l'interazione virtuale, ci si concentra principalmente su 3 sensi, in ordine di importanza: vista, udito e per ultimo tatto, le cui

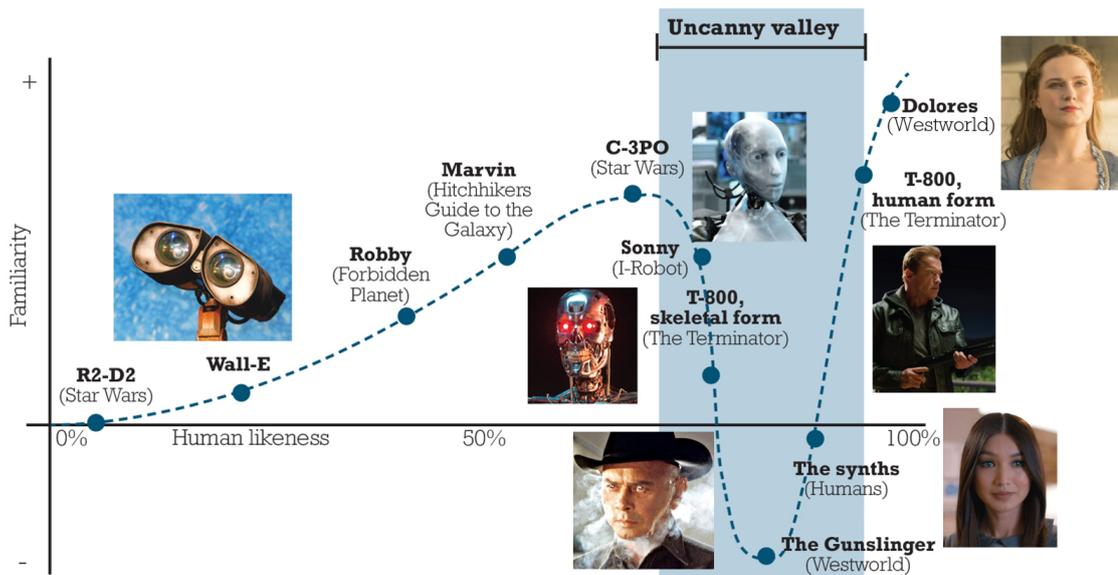
possibilità sono ancora piuttosto limitate.

Per quanto riguarda l'udito, gli algoritmi di *Speech To Text* e *Text To Speech* sono già implementabili con successo, anche a considerazione degli aspetti impliciti comunicati mediante il tono del testo analizzato. Lo sviluppo recente della ricerca in ambito *Machine Learning* riguardante algoritmi di *Natural Language Processing* sono promettenti nello sviluppo di una naturalezza dell'interfaccia uomo-macchina senza precedenti. Per ciò che riguarda la vista, intesa come aspetto grafico del VH e della percezione che l'utente ha di questo, negli ultimi anni si è visto un forte miglioramento grazie all'avvento delle GPU RTX e a software come Unreal Engine 5 che implementano algoritmi di *Sub Surface Scattering* o di simulazione fisica dei capelli. Oltre a queste tecnologie che fanno riferimento solo all'aspetto fotografico del realismo, ci sono anche moderni strumenti legati agli aspetti di animazione e dunque di movimento e comportamentali del VH.

## 1.2 The *Uncanny Valley*

Un importante fenomeno da considerare nella realizzazione estetica di un VH è il fenomeno della *Uncanny Valley*. Secondo questo fenomeno, personaggi virtuali (tra cui i VH) che presentano caratteristiche realistiche ma non totalmente fedeli al reale (come ad esempio occhi sproporzionati rispetto al resto del corpo o effetti indesiderati dovuti a limiti tecnici) vengono percepiti dall'utente come inquietanti o disgustevoli, in inglese *uncanny*. Questo concetto è mostrato dal grafico, all'aumentare dei livelli di realismo ad un certo punto si presenta un minimo locale della funzione di *familiarity* tra l'utente e il VH detto appunto *Uncanny Valley* [2]. Ciò non riguarda solo il realismo fotografico dei singoli frame ma anche di comportamento, considerando quindi l'estetica dell'avatar nella sua interezza e manifestazione multimodale agli occhi dell'utente.

È discusso in letteratura come questo effetto sia legato al realismo dei singoli canali sensoriali con cui l'utente percepisce l'avatar: vista, udito e tatto. Allo stesso modo si discute come la componente interattiva possa influire sulla percezione del VH e quindi su questo effetto di *uncanniness*. Uno studio condotto grazie ai maggiori esperti dell'animazione, il cui motore principale è l'industria dell'intrattenimento (VFX, Gaming, Cinema),



**Figura 1.3:** Relazione tra antropomorfismo/realismo e familiarità/affinità

ha messo infatti in luce come il personaggio con cui un giocatore interagisce all'interno di un videogame non venga necessariamente percepito come *uncanny* proprio perchè il suo realismo grafico viene compensato dal suo grado di interattività [3]. Lo stesso non si può dire per alcuni effetti speciali realizzati con tecniche ormai obsolete anche all'interno di film del cinema cult, in cui questa componente interattiva è totalmente assente. È evidente però come queste intuizioni debbano essere supportate da evidenze statistiche e quindi di ricerche mirate sull'argomento.



**Figura 1.4:** Esempi di personaggi percepiti come *uncanny* (a sinistra) e non (a destra)

Nella realizzazione di un VH, in particolare per ciò che riguarda un

personaggio con cui l'utente deve interagire e non essere volutamente *uncanny*, è importante evitare questo effetto. Ciò assume un ruolo di fondamentale importanza nel momento in cui il VH deve essere progettato per stabilire un legame empatico con l'utente (vedi 1.4).

Al fine di evitare questo fenomeno la soluzione più intuitiva da adottare è puntare al maggior realismo grafico possibile, tentando di superare la Uncanny Valley grazie proprio alle nuove tecnologie [4, 3].

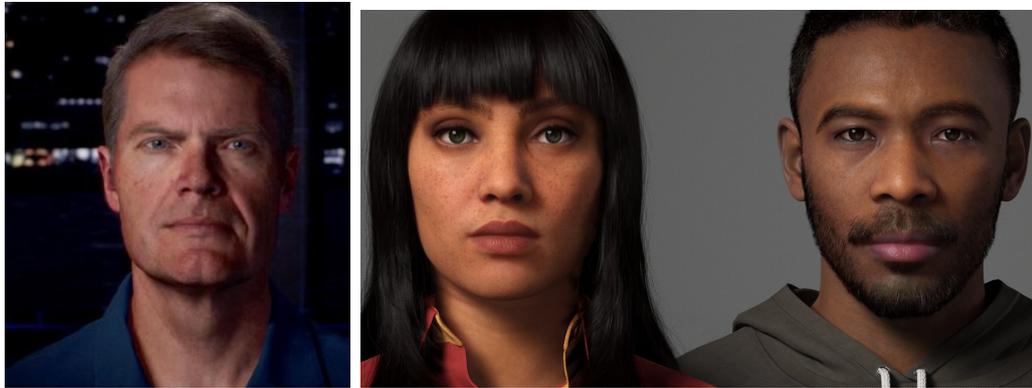
### 1.3 Fotorealismo

Al di là del fenomeno della *Uncanny Valley*, è discusso in letteratura quale sia l'effetto di un VH fotorealistico sull'apprendimento degli utenti in applicazione di training. Alcuni studi dimostrano come la resa più fotorealistica di un VH, in particolare di Virtual Patients (ovvero VH predisposti al ruolo di pazienti in simulazioni mediche), può creare una maggiore senso di immersione e di percezione umana del personaggio virtuale. L'utente infatti interpreta meglio la sua identità, il suo comportamento, la sua personalità, grazie ai maggiori dettagli espressivi forniti e ci interagisce maggiormente.[5] [6] Come di seguito spiegato questo è estremamente importante ai fini di un training, perché favorisce lo sviluppo di un legame emotivo tra l'utente ed il VH e dunque una migliore resa nell'apprendimento del task desiderato.

Altri studi invece dimostrano che il minor fotorealismo però permette all'utente di percepire il VH come maggiormente realistico nel suo movimento. Le animazioni infatti risultano più fedeli al contesto e al personaggio delineando una migliore risposta emotiva dell'utente rispetto agli stimoli (espressioni, voce, movimento del VH, ecc.). [7] Questo potrebbe indicare che il fotorealismo non sia un aspetto scollegato dal comportamento o dall'intrattività come è stato menzionato in precedenza.

È importante notare comunque come la maggior parte degli studi rilevanti utilizzino tecnologie non all'attuale stato dell'arte per il fotorealismo. Il rapido sviluppo di nuove tecnologie permette infatti netti miglioramenti rispetto ad 1 o 2 anni fa. In particolare plugin come MetaHumans, permettono la creazione e customizzazione di VH in maniera semplice ed intuitiva raggiungendo risultati di realismo senza precedenti [8]. MetaHuman, attualmente in modalità *early access*, da inizio 2022 ha aperto la

strada ad ulteriori ricerche e casi studio. Tra i più recenti si è dimostrato come questo livello di realismo possa superare il fenomeno della *Uncanny Valley* 1.2 uscendo dal minimo indicato dal grafico.



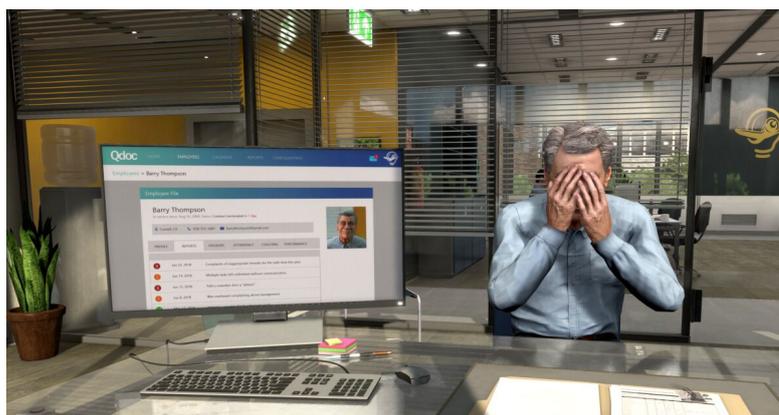
**Figura 1.5:** Esempi di VH allo Stato dell'Arte

## 1.4 Emozioni ed Espressività Facciale

Nella maggior parte dei training in cui è richiesta un'interazione con un VH non si può prescindere da tutti gli aspetti che riguardano il rapporto inter-personale ed empatico che si instaurerebbero con una persona reale. In alcuni casi inoltre la creazione di un legame empatico con il VH, come ad esempio il rapporto di un medico con il proprio paziente, è di fondamentale importanza per la buona riuscita del training. Diversi studi mostrano il beneficio che questo legame empatico tra utente e VH porta all'apprendimento del task proposto, di qualunque scenario VR esso sia.

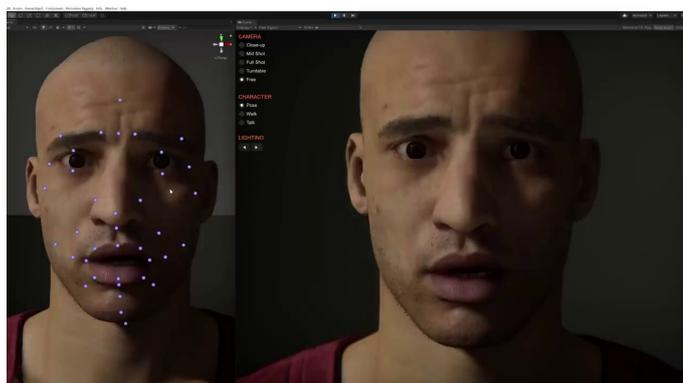
Per rendere l'interazione realistica ed instaurare questo legame, aspetti come l'uncanny valley, il fotorealismo, l'interattività ed in generale il realismo multimodale sono da considerare. Senza di essi infatti il VH non sarebbe in grado di esprimere alcun tipo di emotività il che porterebbe ad una percezione non umana dell'avatar e precluderebbe all'utente l'empatia con il VH.

É evidente come al fine di stabilire questa condizione è necessario che il VH sia capace di comunicare correttamente all'utente il suo stato emotivo. Il modo più immediato con cui le persone lo interpretano è attraverso



**Figura 1.6:** Reazione emotiva di un VH durante [9]

l'osservazione del comportamento ed in particolare delle espressioni facciali. Ciò comporta che, dal punto di vista della computer grafica e dei VH, sia necessario avere un sistema di animazione facciale capace di eseguire in real-time delle espressioni correttamente decodificabili dall'utente.



**Figura 1.7:** Esempio di animazione facciale in *Unity 3D*

L'approccio standard utilizzato per l'animazione facciale è mediante delle *skeletal mesh*, ovvero mesh dotate di un *rig* che tramite dei controlli *rig* e/o *blendshapes* interpolati nel tempo mediante *keyframes* definiscono una determinata deformazione della mesh originaria. Ciò che stabilisce il grado di realismo di queste animazioni è principalmente il tipo di controlli che si hanno a disposizione e l'effetto di deformazione che essi hanno sui singoli vertici della mesh.

Il volto umano è dotato di 36 muscoli facciali che variano però leggermente in forma, dimensioni e tonicità caratterizzando i diversi volti ed espressività della persona. È chiaro come, anche in questo caso, sia difficile realizzare un sistema di animazione facciale accurato e realistico. La soluzione solitamente adottata in ambito VFX per il Cinema ed il Gaming è una pipeline dedicata interamente alla ricostruzione dell'animazione facciale a partire dalla performance di un attore registrata mediante *mocap*.



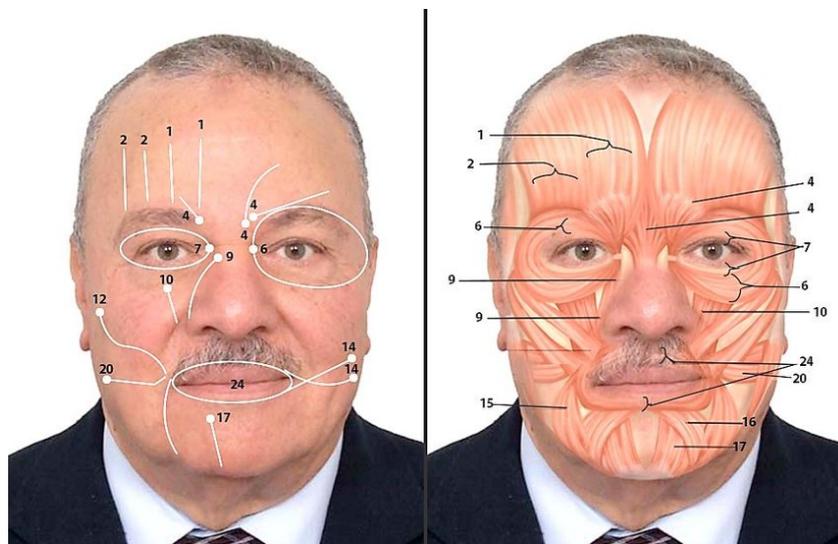
**Figura 1.8:** Esempio di *mocap* facciale

Essa risulta particolarmente dispendiosa in termini di tempo e denaro e richiede la partecipazione di ingegneri, tecnici ed artisti altamente specializzati per ogni step della catena. Approcci del genere sono quindi difficilmente adattabili alle esigenze di prototipazione e produzione VR per quanto MetaHuman per esempio implementa uno sperimentale sistema di *mocap* tramite il sensore *LIDAR* di iPhone. Il problema principale è la difficoltà di generalizzazione di questo metodo, per ogni VH avente fisionomia diversa infatti è necessario non solo rieseguire l'intera pipeline di produzione ma anche introdurre, per i VH, fattori creativi ed artistici non derivanti da attori in carne ed ossa.

Per risolvere questo problema la maggior parte degli studi sull'argomento utilizzano un sistema di codifica dell'espressività facciale che identifica quali sono i singoli movimenti facciali che l'essere umano può attuare. Lo stesso sistema può così essere impiegato nella realizzazione di controlli meno arbitrari e più generalizzabili rispetto all'approccio sopracitato.

## 1.5 Facial Action Coding System

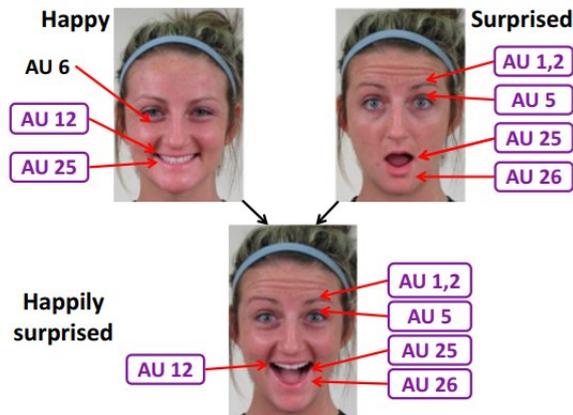
Il Facial Action Coding System è uno standard di descrizione e tassonomia delle espressioni e dei movimenti di faccia e testa in generale [10]. Vede la sua utilità nel campo della computer grafica, dall'animazione in VR al cinema per descrivere e generare le espressioni dei VH. Esso prevede l'utilizzo di Action Units (AUs) le quali singolarmente vanno a definire dei movimenti atomici localizzati sul volto umano, cercando di riprodurre fedelmente quelle che sono le contrazioni e le distensioni dei muscoli facciali (vedi figura 1.9). La composizione ed il bilanciamento di più AUs può così definire una certa espressione 1.10.



**Figura 1.9:** Mappatura delle AUs sul volto umano

Vengono definite 6 espressioni [11] che sintetizzano le 6 corrispondenti emozioni universali, comprensibili dalla specie umana: *Happiness*, *Sadness*, *Surprise*, *Fear*, *Anger*, *Disgust*. Esse sono definite *basic emotions*. Ognuna di queste può corrispondere a più sequenze di AUs, e dunque non hanno una codifica e rappresentazione univoca [12]. In letteratura, a seconda degli studi si individuano infatti sequenze di AUs diverse considerata una stessa emozione [13].

Lo standard definisce originariamente una scala di intensità a 3 livelli che successivamente in letteratura aumentano a 5. Non è chiaro come questi livelli vadano assegnati alle singole AUs che compongono le varie



**Figura 1.10:** Esempi di FACS per la descrizione di emozioni di base

basic emotions, limitandosi a definire le sequenze di AUs di cui esse sono composte. Diversi studi mostrano come sia difficile stabilire per un movimento facciale corrispondente ad una AU un valore su questa scala. Molti preferiscono quantificare la misura di intensità con valori compresi tra 0, totale assenza della AU (idealmente muscoli totalmente rilassati), ad 1 presenza massima della AU (idealmente muscoli contratti al massimo). In questo modo si possono direttamente mappare l'intensità delle AUs ai blendshape realizzati per il modello da animare, con casi limite 0, *shapekey* totalmente disattivata ed 1, *shapekey* totalmente attivata (vedi figura 1.11).



**Figura 1.11:** Esempio di interpolazione *shapekey* da 0 ad 1

## 1.6 Obiettivi

Alla luce delle problematiche attuali nella realizzazione di VH per la VR la seguente tesi propone un implementazione dello standard FACS nel *game engine* Unreal Engine 5. Inoltre definisce un framework sperimentale per:

- Generare un dataset di espressioni facciali di VH
- Sperimentare su larga scala questo dataset
- Studiare l'espressività mediante i risultati ottenuti

Lo scopo della sperimentazione è capire se è possibile stabilire una correlazione tra avatar diversi nell'espressione delle *basic emotions* ed individuare mediante il sistema FACS il modo di codificarle al meglio. In letteratura infatti non sono presenti studi che identifichino una univocità nella rappresentazione delle emozioni di base tra avatar appartenenti a diverse etnie e dataset diversi. Ad esempio, dati due avatar qualsiasi e considerata l'emozione *happiness* non è ancora stata identificata una combinazione di AUs univoca per esprimerla su entrambi.

Considerati inoltre gli avanzamenti della tecnologia nella pipeline di rendering grafico, si vuole valutare a che punto il realismo (in questo caso solamente statico essendo il dataset composto da screenshot e non animazioni video) è determinante nella percezione dell'emozione da parte dell'utente.

Le ipotesi prefissate dal lavoro di tesi sono dunque:

1. Le basic emotions come definite col sistema FACS hanno una valenza univoca. Al variare delle caratteristiche estetiche (sesso, età, etnia, ecc.) degli avatar la sequenza di AU e weights, definiti per rappresentare una data emozione ad una certa intensità, è costante.
2. Esistono AUs definite per ogni emozione più significative di altre. Queste sono determinanti nell'identificare l'emozione data anche in assenza delle altre AUs. (o con weights significativamente inferiori).
3. Un maggior fotorealismo garantisce una migliore identificazione dell'espressività facciale dell'avatar e di conseguenza una migliore interpretazione dell'emozione.

# Capitolo 2

## Stato dell'arte

Nel presente stato dell'arte ci si concentra sul ruolo svolto dall'empatia nel rapporto tra utente e VH considerando, tra gli altri, gli aspetti legati al fotorealismo e di animazione facciale real-time con l'integrazione di sistemi per l'espressività dei VH, quali tra i più usati il FACS.

Come sopra citato le attuali tecnologie di rendering tramite ray-tracing e solving per animazioni facciali hanno aperto la strada per la realizzazione di Virtual Humans all'avanguardia sia per applicazioni VR che non. La *game industry*, prima fra tutte le industrie a contribuire alla ricerca e sviluppo di nuove soluzioni hardware e software in ambito grafico, ha messo a disposizione innovativi strumenti quali *game engine* e plugin dedicati proprio allo sviluppo di Virtual Humans sempre più fedeli al reale [8].

### 2.1 Virtual Humans ed Empatia

In letteratura sono numerosi gli studi che analizzano il ruolo dell'empatia tra un VH e l'utente, in particolare per trarne un beneficio in termini di velocità e qualità nell'apprendimento dei task proposti in ambito e-learning [14, 15, 16]. Ad esempio per quanto riguarda training in ambito medico è provato come l'emotività espressa da un paziente virtuale possa portare l'utente ad un maggior senso di presenza nella simulazione, e ad interagire maggiormente con il VH [17, 18, 19, 20].

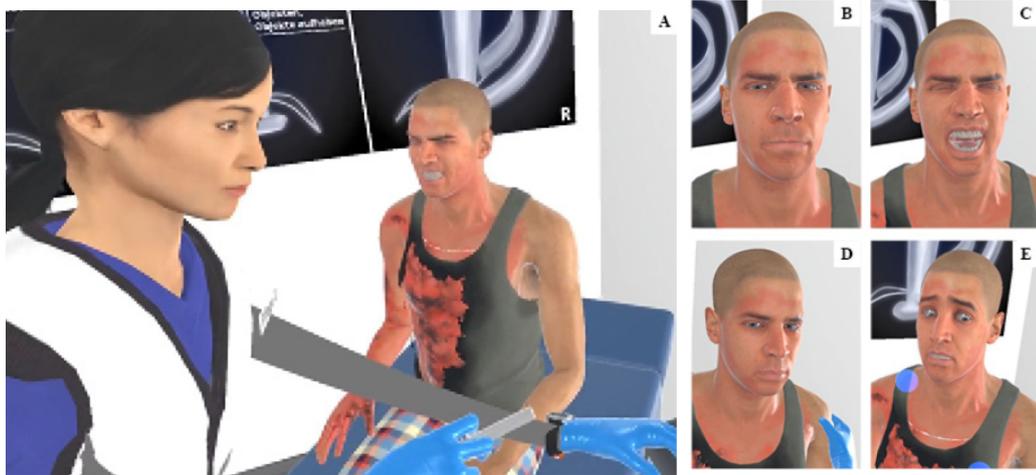
Allo stesso modo anche il rapporto empatico che si stabilisce tra l'utente ed un istruttore virtuale può garantire un miglior risultato della simulazione, qualunque sia il campo di applicazione [21, 22].



**Figura 2.1:** Esempio di istruttore virtuale [21]

Come precedentemente spiegato (vedi 1.4), è evidente come sia estremamente dispendioso e difficile raggiungere questo obiettivo a causa dei limiti tecnici con cui ci si scontra nella realizzazione di un VH. In particolare gli aspetti comportamentali ed espressivi dei VH sono difficili da simulare ed in letteratura non è stabilito uno standard di misurazione riguardante come questi parametri influenzino la creazione del legame empatico [23, 24, 25]. Diversi studi si concentrano sull'effetto dell'*uncanny valley* supponendo che questo possa far percepire all'utente il VH come disgustevole e dunque non empatizzare con esso [26, 27, 21]. È bene però notare come tutti questi articoli non utilizzino tecniche all'avanguardia come quelle descritte in questa tesi (vedi 2.3), introdotte nell'ultimo anno ed ancora poco utilizzate nella sperimentazione dei VH. Alcuni recenti studi mostrano come grazie a questi avanzamenti sia possibile scostarsi dal minimo del grafico di *uncanniness* e risalire verso un livello di familiarità consono ai VH aumentandone proprio il realismo [4, 3, 28].

Oltre al fotorealismo è bene notare come l'effetto descritto dalla *uncanny valley* si estende a come l'utente percepisce l'avatar nella sua multimodalità. Alcune studi menzionano come il ruolo di un livello di realismo maggiore nell'interazione con l'utente [7, 17] e nel comportamento del VH [14, 3, 17] possa contribuire al miglioramento della *uncanniness* ed è dimostrato come questi aspetti siano comunque determinanti nello stabilire un legame empatico con l'utente.



**Figura 2.2:** Esempi di interazione (sinistra) ed espressività dei VH (destra) [17]

### 2.1.1 FACS

Per ciò che riguarda il comportamento del VH è di focale importanza la sua capacità di comunicare le proprie emozioni mediante le espressioni facciali [6, 3, 17]. Tra i metodi proposti in letteratura il più importante è sicuramente il FACS (vedi 1.5). Lo standard è l'unico a gettare le fondamenta per una possibile codifica dell'espressività umana in maniera univoca [29]. Inoltre il FACS renderebbe parametrizzabile la sintesi di espressioni su un qualunque avatar che implementi lo standard, abbattendo significativamente i costi in una pipeline di produzione VR. La maggior parte degli studi sull'argomento applica questo sistema per la realizzazione di volti espressivi al fine di valutarne l'efficacia [30, 14, 17]. Gli avatar utilizzati risultano però, anche in questi casi, non all'avanguardia in senso fotorealistico. Alcuni utilizzano avatar creati ad hoc [31], che rendono il sistema poco adattabile ad altri, oppure usano set appositi come RocketBox, soluzione sviluppata da Microsoft contenente 115 VH, costituiti da un'unica mesh per ogni avatar e stesso rig sull'intero set.

Esistono inoltre dei framework appositi per la creazione di VH, alcuni che direttamente implementano il sistema FACS come ad esempio FACSHuman [32] (plugin di Make Human [33], a sua volta estensione del software Blender [34]) o Virtual Human Toolkit [35], mentre altri come



**Figura 2.3:** Alcuni avatar del set Rocketbox

MetaHuman (vedi 2.3) non hanno direttamente un'implementazione.

Questa tesi si concentra in primis sull'implementazione di questo sistema FACS per MetaHuman e RocketBox all'interno del motore Unreal Engine 5, per poi proseguire nella sperimentazione di questo(vedi 1.6).

## 2.2 Unreal Engine 5

Lo standard attuale *de facto* per lo sviluppo di soluzioni 3D real-time fotorealistiche è Unreal Engine la cui versione 5 (UE5) [36] lanciata nell'aprile 2022 ha introdotto novità rivoluzionarie [36]. In particolare, a parità di prestazioni con le versioni precedenti, sono stati introdotti:

- Lumen: sistema che permette il calcolo dell'illuminazione globale dinamica rendendo luci, ombre e riflessioni sempre più accurate
- Nanite: sistema di rendering virtuale della geometria, capace di gestire modelli da milioni di vertici in real-time
- Virtual Texturing: sistema di gestione delle texture che riduce la RAM necessaria al loro caricamento e permette di conseguenza di aumentarne la qualità (4K, 8K o 16K)

UE5, come altri engine, per esempio il suo rivale commerciale Unity 3D, offre all'interno del proprio framework dei sottosistemi di rendering, fisica,



**Figura 2.4:** Esempio delle capacità del ray-tracing real-time Lumen

audio, animazione e networking che consentono lo sviluppo di applicazioni 3D sia real-time che non. Tra queste, di interesse per la presente tesi, UE5 mette a disposizione SDK proprietari ed open source per lo sviluppo di soluzioni VR adattabili a qualunque HMD sul mercato. Questi si basano su OpenXR, che è uno standard open per la realizzazione di applicazioni di realtà estesa, aumentata o virtuale multiplatforma.

Queste caratteristiche rendono l'ambiente di UE5 un ottimo candidato per lo sviluppo VR con VH che necessitano di un adeguato livello di fotorealismo. La stessa Epic Games Inc., casa di sviluppo di UE, ha rilasciato un plugin apposito per la creazione ed animazione di Virtual Humans chiamato MetaHuman. Per motivi commerciali esso è limitato solo all'utilizzo dentro UE garantendone il primato.

## 2.3 MetaHuman

MetaHuman mette a disposizione un intero sistema di creazione, modellazione, rig, animazione, mocap ed integrazione nell'ambiente 3D di VH caratterizzati da un elevato realismo grafico (sia di animazione che per render statici) [37].

Grazie alle nuove capacità di ray tracing e modelli di simulazione come il subsurface scattering è possibile ad esempio effettuare delle simulazioni degli hair, peli, ciglia, barba e capelli tali da avere movimento coerente con il corpo dell'avatar ed una maggiore fedeltà del colore usando texture per rughe e flusso sanguigno sotto pelle per citare alcuni esempi.



Figura 2.5: Esempi di MetaHuman

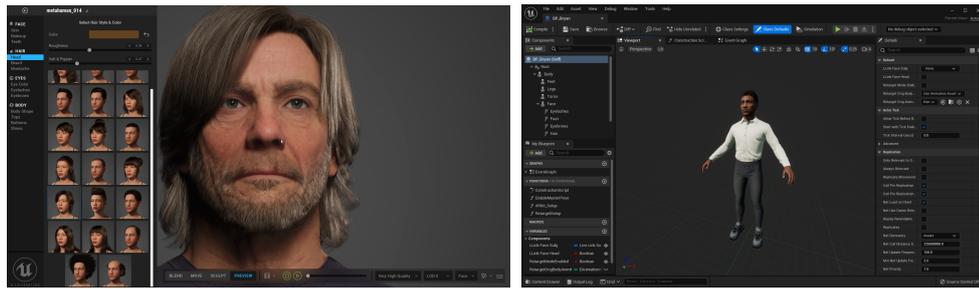
### 2.3.1 Pipeline di produzione

Il tool per la creazione di questi VH è MetaHuman Creator. A partire da un set di avatar aventi sesso, etnia, età, corporatura e fisionomia diversa già utilizzabili è possibile crearne di nuovi mediante la combinazioni delle caratteristiche di questi. Al momento non è possibile crearne a partire da zero se non con un sistema di conversione da uno scan facciale 3D a MetaHuman, processo che richiede comunque uno scanning di ottima qualità ad aggiustamenti *ad hoc* per il singolo volto.

Il MetaHuman (MH) può essere dunque importato in un progetto UE tramite Quixel Bridge, plugin direttamente integrato in UE5 che, oltre ai MH, offre modelli ottenuti tramite 3D scanning ad elevata qualità geometrica e di texturing da utilizzare mediante Nanite. Gli asset importati sono un insieme di materiali, *skeletal* e *physics meshes*, *animation blueprint* che vengono raggruppati in un unico *blueprint* pronto all'uso.

Gli asset principali di cui è composto un MH sono:

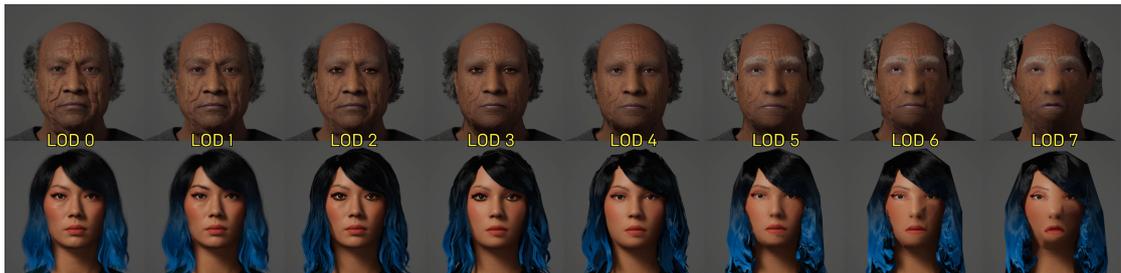
- *Skeletal Mesh components* per corpo, gambe, piedi, faccia e torso
- *Grooms* per capelli, sopracciglia, ciglia, peluria, baffi e barba
- *LodSync* componente che gestisce i Level Of Details dei due componenti di cui sopra
- *Control Rigs* per l'animazione dello scheletro del corpo (*metahuman\_base\_skel*) e per il volto (*Face\_Archetype\_Skeleton*)



**Figura 2.6:** Viewport di un MH su MetaHuman Creator (destra) e UE5 (sinistra)

### 2.3.2 Level Of Details

Gli asset grafici di ogni MH sono dotati di 8 LOD che garantiscono una buona scalabilità su qualunque piattaforma, in particolare per applicazioni VR richiedenti ottime prestazioni in termini di FPS. La moltitudine di questi LOD è raggruppata in LOD del MH, gestiti dal componente *LodSync* che ne permette una completa customizzazione.



**Figura 2.7:** Esempi di LOD sue due MetaHumans

### Grooms

Particolare attenzione va fatta sui LOD dei *grooms*. A differenza delle *skeletal mesh* al cui diminuire del LOD corrisponde una riduzione del numero di vertici, i *grooms* presentano 3 metodologie diverse di rendering ordinate per qualità decrescente: *Strands*, *Cards*, *Mesh*. Tralasciando le *mesh*, classiche rappresentazione mediante un unico modello texturizzato, *strands* e *cards* introducono la capacità di simulare il movimento dei capelli ed effetti di illuminazione nettamente più realistici.

Gli *strands* sono un particolare tipo di asset che definisce, mediante dei parametri standardizzati per l'interscambio tra software come il formato alembic *.abc*, un insieme di curve modellabili con appositi strumenti (simili a quelli reali a disposizione di un parrucchiere) che definiscono la forma del singolo pelo. Le *cards* sono una semplificazione di queste curve mediante dei *quads*, che composti insieme generano i singoli peli.

Nel primo caso è possibile ottenere il miglior fotorealismo grazie all'applicazione di appositi *shader* per definire le proprietà di colore, trasparenza, assorbimento e riflessione del capello. Nell'altro caso queste proprietà sono definite dalle texture applicate, meno puntuali e dettagliate ma che garantiscono comunque un buon risultato (figura 2.8).



**Figura 2.8:** *Groom* renderizzato mediante *strands* (sinistra) e mediante *cards* (destra)

### 2.3.3 Animazione & Rig Logic

Ogni metahuman è dotato di un rig standardizzato che è possibile animare tramite gli appositi controller sopraccitati. In particolare il rig facciale permette un accurato setup delle singole microespressioni facciali tramite controlli (figura 2.9). Questo livello di dettaglio è garantito anche in animazioni real time, dove vi è la necessità di modificare in maniera

dinamica le espressioni, grazie a Rig Logic, il sistema di *evaluation* e *solving* del rigging facciale. [38]

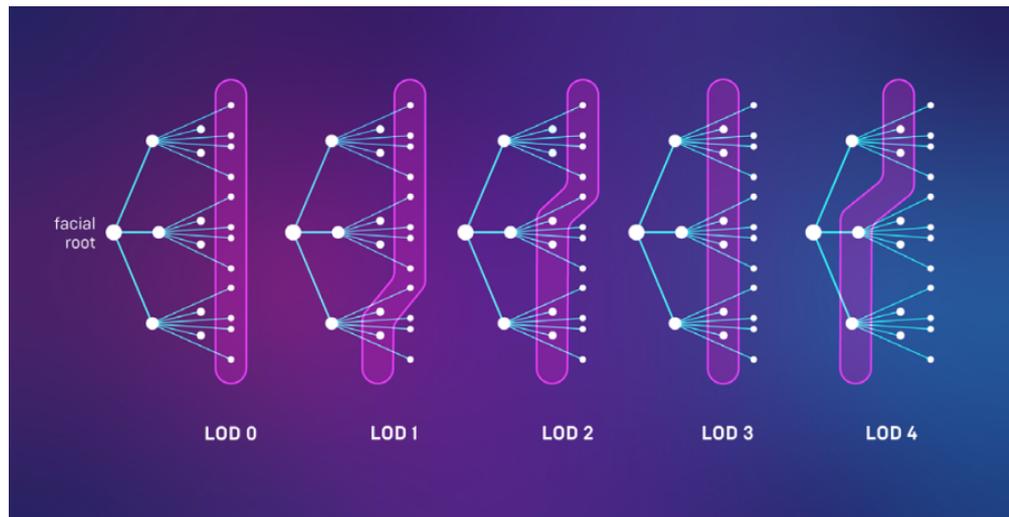


**Figura 2.9:** UI del rig facciale di MetaHuman

Al fine di garantire una soglia minima di operazione di 30 fps, indispensabile per realizzare un movimento percepito dall'utente come naturale, il sistema utilizza un formato di file proprietario chiamato MetaHuman DNA, contenente tutti metadati del MH (tra gli altri nome, età, corporatura) e le informazioni riguardanti la geometria del volto (mesh e mapping dei controlli di *joints*, *blendshapes* e *animated maps*). A partire da uno di questi file è possibile ricostruire interamente il volto ed il rig di animazione di un MH.

Attraverso questi dati, per come sono strutturati all'interno della codifica, è possibile semplificare il processo di *solving* dell'animazione grazie ad un sistema di "sfolgimento" mediante LOD. I *joints* dello scheletro facciale sono organizzati secondo uno schema ad albero. Essi sono raggruppati secondo la loro posizione facciale ed in maniera graduale rispetto ai livelli dell'albero. In questo modo si possono individuare dei LOD e selezionare i *joints* corrispondenti tralasciando gli altri come accade nel caso dei LOD geometrici. Questo stesso concetto viene inoltre applicato per la semplificazione dei *blendshapes* e delle *animated maps*.

Il sistema, secondo gli sviluppatori, è realizzato idealmente per un



**Figura 2.10:** Struttura gerarchica dei *joints*

mapping 1:1 con le AUs del sistema FACS, per quanto i nomi dei singoli controlli non corrispondano a quelli della codifica [38]. Attualmente non è stato rilasciato un mapping ufficiale, cosa che assieme alle altre sarà oggetto di questa tesi.

# Capitolo 3

## Metodologia

La sperimentazione proposta consiste nel far valutare a degli utenti quali screenshot, associati ad una determinata espressione seguendo il metodo FACS, fossero più consoni a rappresentare una data emozione di base. Al fine di testare la veridicità dell'ipotesi 2 sono stati utilizzati due set di avatar diversi, uno più fotorealistico dell'altro: MetaHuman e Rocketbox.

In figura 3.1 è rappresentato lo schema generale del sistema di generazione. A partire dalle AUs delle varie emozioni e dai nomi dei controlli del rig descritti mediante file JSON, viene creata una struttura dati interna ad Unreal. Questa mappa le informazioni sullo specifico avatar e controlla la posa finale del rig. Viene infine generato lo screenshot risultante, salvato nel dataset.

Il primo passo nella generazione degli screenshot è la generazione delle sequenze di pesi corrispondenti alle sequenze di AUs che descrivono una data espressione.

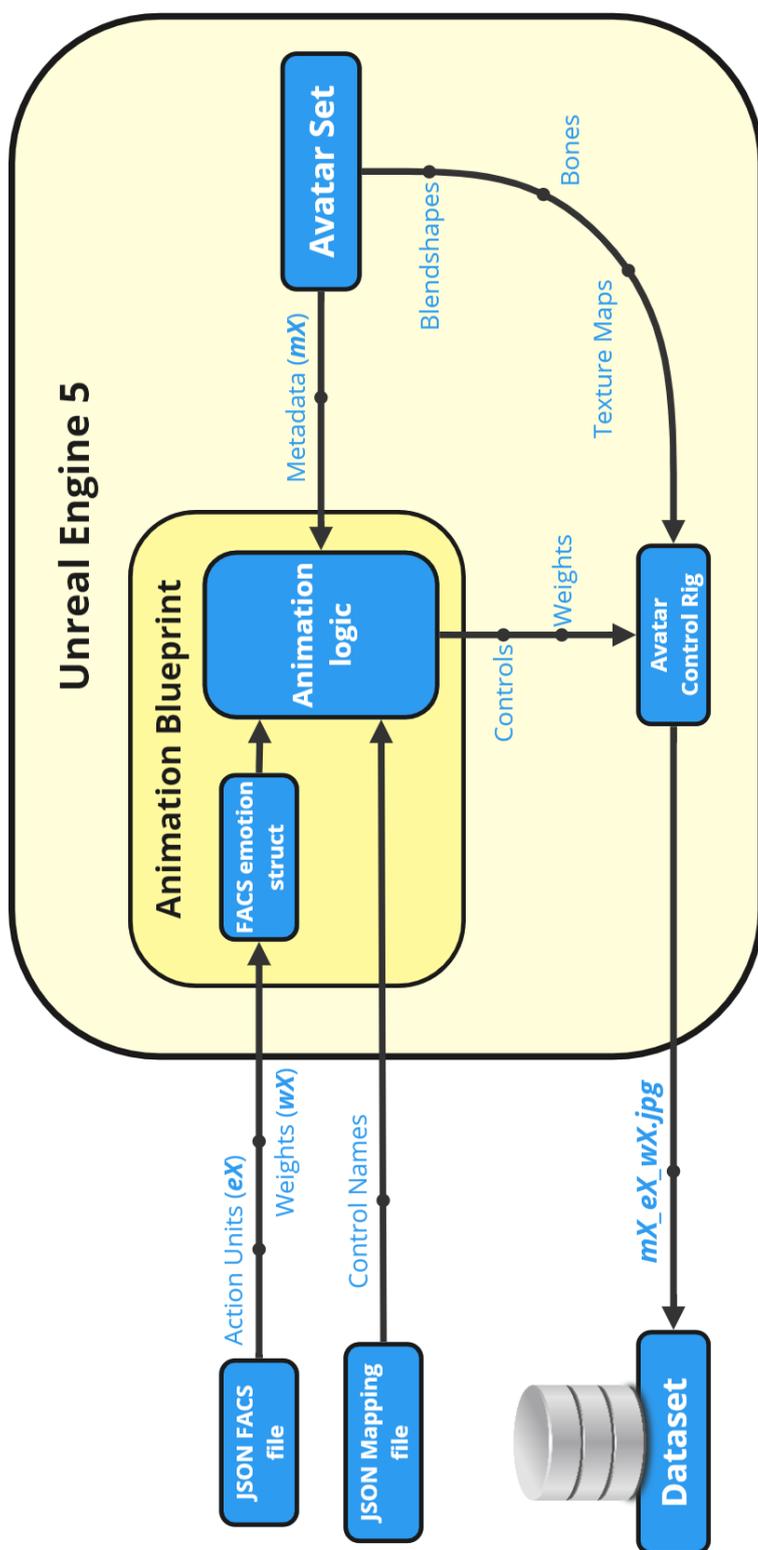


Figura 3.1: Schema del sistema di generazione del dataset

### 3.1 FACS Action Units

In letteratura non sono presenti univoche sequenze di AUs che rappresentino le varie emozioni di base, ma una moltitudine di esse per ogni emozione. Per ridurre la complessità dello studio è stato scelto quindi di ridurre il dataset unificando le sequenze di AUs afferenti ad una stessa emozione in un'unica sequenza. In questo modo si ottiene una sola sequenza per ogni emozione (vedi tabella 3.1).

Emotions	AUs Sequence
Anger	[4, 5, 7, 10, 17, 22, 23, 24, 25, 26]
Disgust	[9, 10, 16, 17, 25, 26]
Fear	[1, 2, 4, 5, 20, 25, 26, 27]
Happiness	[6, 12]
Sadness	[1, 4, 6, 11, 15, 17]
Surprise	[1, 2, 5, 25, 26, 27]

Ognuna di queste sequenze indica quali sono le AUs che combinate tra loro generano un'espressione felice piuttosto che triste o arrabbiata, ma ancora non vengono definiti i pesi tra 0 ed 1 da assegnargli. Considerando quindi una emozione per volta e la sua sequenza di AUs vengono generate tante sequenze di pesi (weights). Si considera la disposizione come una sequenza ordinata di  $k$  elementi formata a partire da  $n$  elementi distinti. Le sequenze di pesi di una data emozione sono risultanti da tutte le disposizioni con ripetizione di questi  $k$  ed  $n$  numero di AUs.

Si è considerato di utilizzare  $k = 3$ , ovvero 3 pesi diversi per ognuna AU, aventi valori compresi tra 0 (AU totalmente "off") e 1 (AU totalmente "on"). Si è considerato di scartare pesi troppo bassi che non sarebbero stati rappresentativi di una data espressione ottenendo, a parità di grandezza del dataset, una maggiore granularità sulla fascia più alta. I valori scelti sono 0.4, 0.675, 0.95. Il valore 1 è stato scartato per evitare problemi riguardanti possibili distorsioni esagerate nelle mesh degli avatar. Le sequenze di pesi sono state generate mediante uno specifico algoritmo.

## 3.2 Generazione Weights Sequence

L'algoritmo per generare le sequenze di pesi (weights array) è costituito da un counter ed una formula di divisione con resto che ne mappa il risultato con i valori dei pesi sopracitati.

Considerato  $i$  ID dell'array di pesi e  $j$  indice dell'array, sapendo il numero totale di disposizioni possibili  $dispositionsNumber$  la lunghezza della sequenza di AUs  $sequenceLength$  e, dati valore minimo dei pesi  $minValue$  e lo step di incremento  $step$ , si può calcolare tutte le sequenze di AUs possibili.

---

```

1      //itero per tutte le possibili disposizioni
2      for(int i=0; i < dispositionsNumber; i++){
3          //itero su tutti i valori della sequenza
4          for(int j=0; j < sequenceLength; j++){
5              //genero il valore corrispondente alla sequenza i posizione j
6              digitWeight = pow(numberOfValues, j+1);
7              value = (i % digitWeight) * numberOfValues / digitWeight;
8              weightsMatrix[i][j] = minValue + value * step;
9          }
10     }

```

---

**Figura 3.2:** Snippet riguardante la formula di generazione delle sequenze di weights.

L'operazione che l'algoritmo implementa è una conversione da base decimale del contatore  $i$  ad una base  $k = numberOfValues$  i cui valori non sono interi ma mappati sui pesi discreti scelti: 0.4, 0.675 e 0.95. Il valore  $i$  viene dunque rappresentato utilizzato ogni cella dell'array  $weightsMatrix[i]$  come cifra del numero.

Al fine di ridurre ulteriormente la grandezza del dataset per problemi di sperimentazione è stata introdotta in questa fase la possibilità di fare un merging delle AUs, replicando, in tutte le sequenze, il valore di una AU con una o più altre.

Il risultato prodotto da questo algoritmo viene scritto in output su un file JSON, che rende di facile lettura (sia per lo sviluppatore che per UE5

che implementa specifiche funzioni di parsing) ed eventuale modifica il suo contenuto.

### 3.3 Weights JSON file

Il file che descrive le sequenze di AUs e gli array di weights corrispondenti è organizzato come un unico macro oggetto *EmotionType* contenente un array di oggetti aventi 3 campi: *name* nome dell'emozione, *AU* array contenente gli id delle AU di cui l'emozione è composta, e *Weights*. Quest'ultimo è a sua volta un array contenente oggetti con chiave *Values* e come valore un array contenente i pesi delle AUs corrispondenti.

Il file viene dato in input all'algoritmo sopra descritto con solo i primi tre campi compilati *Name*, *AU* e *Merged*. In output si ottiene lo stesso file con aggiunta dei campi *Weights* generati dall'algoritmo stesso.

### 3.4 Mapping JSON file

Viene utilizzato un secondo file JSON che descrive il mapping tra le AU definite dal FACS e l'implementazione dei controlli sul rig degli avatar. Per ogni set di avatar, avendo controlli diversi, viene usato un file dedicato ma avente struttura uguale, così da standardizzare il metodo di mapping. In questo modo nel caso si volesse implementare lo stesso sistema su un altro set di avatar basterebbe creare un nuovo file di mapping con medesima struttura ma valori custom.

Il file è organizzato come un un unico oggetto JSON contenente un array di sotto-oggetti il cui indice definisce l'AU corrispondente. Questi hanno chiave *Controls* e valore contenente la sequenza di stringhe corrispondenti ai nomi dei controlli del rig.

É da notare come alcune AUs non abbiano controlli corrispondenti (e dunque array vuoto) poichè sono state considerate ai fini di questa sperimentazione solo le AUs che descrivono espressioni facciali. Lo standard FACS non si limita però solo a queste, ma definisce anche AUs corrispondenti ad esempio a movimenti del collo o degli occhi, non oggetto di questa tesi.

## 3.5 Animation Blueprint

A partire da questi due file JSON, uno che descrive gli weights e l'altro che ne definisce il mapping sul rig desiderato, per realizzare l'effettiva animazione in UE5 si utilizza un Animation Blueprint. Questo visual script legge entrambi i file e, per ogni weights array mette in posa un avatar desiderato per poi renderizzare uno screenshot. Si ottiene un numero di screenshot pari al numero di weights array totali.

I file JSON vengono caricati a runtime mediante una libreria di funzioni custom che operano su struct gestite da UE ed aventi stessa nidificazione dei file (figura 3.5). Questo permette alla libreria di IO di gestire in maniera semplice ed efficace il parsing da i file JSON alle variabili, liste e mappe interne all'engine.

Per ogni weights array si utilizza un insieme di funzioni di posa che controllano i valori di una mappa stringa-float. Questa mappa associa il nome dei controlli del rig ai pesi da assegnargli, descrivendo così la posa da applicare all'avatar (figura 3.6). A partire da *emotionIndex* indice dell'emozione all'interno dell'oggetto JSON *EmotionType* ed *weightsIndex* indice corrispondente ad un dato weights array, si possono ottenere tutti i pesi come descritti nel weights JSON file.

Ogni set di avatar possiede il suo blueprint con funzioni adattate al diverso rig su cui i controlli devono essere applicati. UE associa ad ogni Animation Blueprint un solo rig.

Il file di ogni immagine è nominato secondo una codifica standardizzata su tutto il dataset e così composta: 3 sezioni separate da underscores, *mX* (avatar ID) che identifica l'avatar, *eX* (emotion ID) l'emozione espressa e *wX* (weights sequence ID) l'array di weights delle AUs corrispondente al volto rappresentato.

```
//Esempi di nomi delle immagini  
m11_e2_w53.jpg  
m7_e4_w35.jpg  
m1_e1_w18.jpg  
m3_e4_w26.jpg
```

```

1  {
2      "EmotionType": [
3          {
4              "Name": "Happiness",
5              "AU": [6, 12]
6              "Merged": [0, 1],
7              "Weights": [
8                  { "Values": [0.4000000059604645, 0.4000000059604645] },
9                  { "Values": [0.699999988079071, 0.4000000059604645] },
10                 { "Values": [1.0, 0.4000000059604645] },
11                 ...]
12             }
13         {
14             "Name": "Surprise",
15             "AU": [1, 2, 5, 26, 27],
16             "Merged": [0, 1, 2, 3, 3, 3],
17             "Weights": [
18                 ...
19                 {"Values": [
20                     0.949999988079071,
21                     0.949999988079071,
22                     0.949999988079071,
23                     0.4000000059604645,
24                     0.4000000059604645,
25                     0.4000000059604645
26                 ]},
27                 {"Values": [
28                     0.4000000059604645,
29                     0.4000000059604645,
30                     0.4000000059604645,
31                     0.675000011920929,
32                     0.675000011920929,
33                     0.675000011920929
34                 ]},
35                 ...]
36             }
37         ]
38     }

```

Figura 3.3: Esempio di Weights file.

```

1  {
2    "AUMapping":
3    [
4      {"Controls": []},
5      {"Controls":
6        ["CTRL_expressions_browRaiseInL", "CTRL_expressions_browRaiseInR"]},
7      {"Controls":
8        ["CTRL_expressions_browRaiseOuterL", "CTRL_expressions_browRaiseOuterR"]},
9      {"Controls": []},
10     {"Controls":
11       ["CTRL_expressions_browDownL", "CTRL_expressions_browDownR"]},
12     ...
13   ]
14 }

```

Figura 3.4: Esempio di Mapping file.

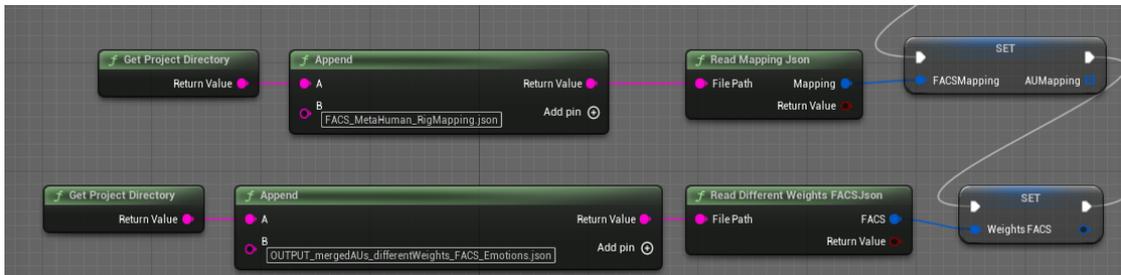


Figura 3.5: Read dei file JSON

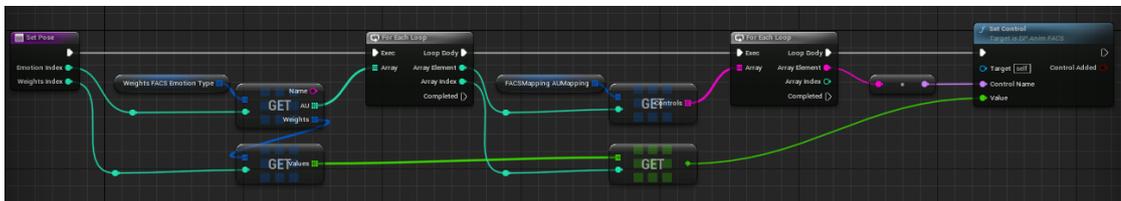


Figura 3.6: Funzione *SetPose(int emotionIndex, int weightsIndex)*

# Capitolo 4

## Sperimentazione

Per l'elevato numero di record necessari al test e per la semplicità della task da svolgere, non richiedente assistenza o controllo da parte dell'esaminatore, si è scelto di sviluppare una web app *ad hoc*. Questo ha permesso la somministrazione del test in maniera asincrona, facilmente diffondibile mediante il link e su larga scala.

Per testare il funzionamento del framework la sperimentazione è stata condotta su due emozioni: *happiness* e *surprise*, per ognuna delle quali è stata anche aggiunta l'espressione neutra (tutte le AUs a 0). Esse sono state scelte in relazione al loro numero di AUs, più basso rispetto alle altre, ulteriormente ridotto dalla fusione di alcune AUs tra di loro (vedere 3). I parametri di testing sono indicati alla tabella 4.1.

Avatar	<b>24</b>
Basic Emotions	<b>2</b>
Valori AU	<b>3</b>
<i>Happiness</i> (2 AUs)	<b>9 + 1</b>
<i>Surprise</i> (4 AUs)	<b>81 + 1</b>

**Tabella 4.1:** Parametri utilizzati per la generazione del dataset

## 4.1 Protocollo

### 4.1.1 Raccolta Dati

La web app per lo svolgimento del test è stata fornita tramite link. È stato pubblicato in vari forum e social network online per raggiungere un campione di persone rilevante in termini di numero e varietà degli esaminati. All'istante dell'accesso alla web page viene direttamente presentata una sessione di test il cui risultato sarà un record. L'esaminato può quindi scegliere l'immagine che più rappresenta l'emozione menzionata e cliccare su *NEXT*.

Per prevenire eventuali disconnessioni, page refresh o problematiche legate alla distribuzione delle immagini corretta ma seguita da una non ricezione del risultato da parte del server, è stato implementato un tempo di timeout server-side. Allo scadere di questo lasso di tempo il server ignora qualunque futura risposta dal client e ricarica le immagini nella queue in attesa di distribuirle ad una nuova richiesta.

Nel momento in cui l'utente preme il pulsante viene inviata al server una richiesta di inserimento record con la scelta effettuata, poi il server si occupa di registrare il risultato sul database.

Si rimanda alla sezione 4.2 per una descrizione approfondita.

### 4.1.2 Analisi Dati

A partire dal database si estraono per ogni immagine numero di vittorie e numero di sconfitte, ovvero il dataset di partenza per procedere con l'analisi. Essa è stata strutturata secondo uno schema ad albero (vedi 4.1) che va a generalizzare per ogni step, ovvero ogni livello di ramificazione, il confronto tra gruppi di avatar aventi internamente stesse caratteristiche estetiche.

Questo metodo consente di validare *step-by-step*, al risalire verso la radice dell'albero (graficamente procedendo più in basso), se l'ipotesi di universalità nell'espressione delle *basic emotions* mediante FACS è confermata. Nel momento in cui un nodo di uno step infatti dovesse invalidare l'ipotesi di associatività (vedi 1.6) si può considerare non attendibili i risultati agli step successivi.

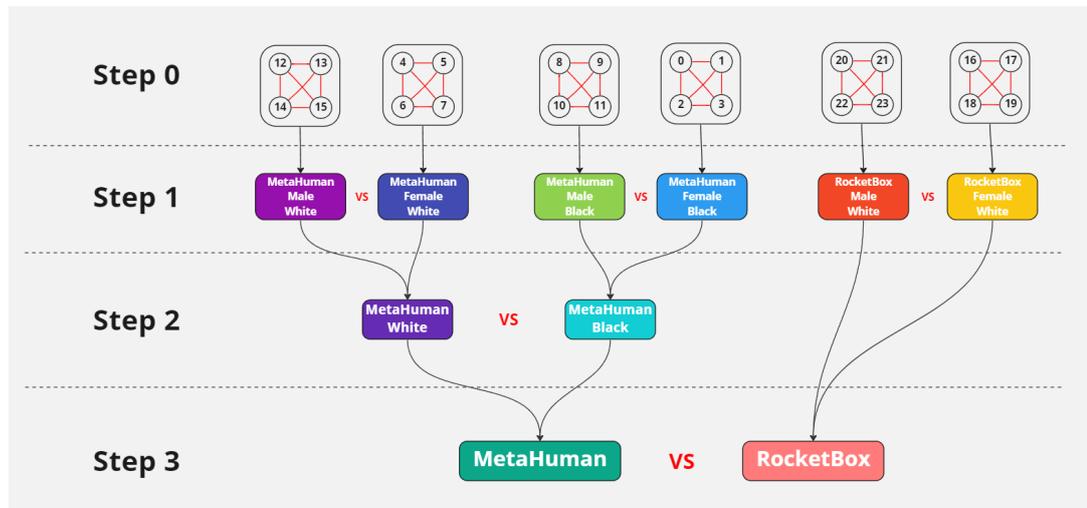


Figura 4.1: Schema di analisi *ad albero*

Ogni step prevede un dato numero di confronti tra avatar o gruppi di avatar aventi almeno una caratteristica comune. Ogni confronto è indicato con *vs* ad eccezione dello step 0 dove è indicato con una linea rossa.

Allo step 0 vengono confrontati tutti gli avatar appartenenti ai gruppi più piccoli (ad esempio MetaHuman Black Female, con avatar ID 0,1,2 e 3) tra loro. Considerata la quantità di confronti è utile verificare quanti di questi in percentuale superano il test di confronto per ogni gruppo. Agli step successivi i confronti non vengono più effettuati su singoli avatar ma su gruppi di questi via via crescenti.

L'elaborazione di ogni confronto è suddivisa in due fasi:

## 1. Group Ranking

Vengono calcolati i ranking di ogni *weights sequence* rispetto alle altre (questo per ogni emozione).

Considerati i singoli avatar allo step 0 si calcola il rank assoluto di ogni immagine corrispondente ad una *weights sequence* applicata ad un singolo avatar, come mostrato in 4.3. Si ottiene così per ogni avatar una sequenza di ranks, ordinata secondo gli ID delle *weights sequence* corrispondenti, da poter confrontare con gli altri avatar.

Per quanto riguarda gli step successivi è necessario effettuare un'operazione di *merging* tra gli avatar del gruppo. Si calcola così un rank per ogni *weights sequence ID* considerando le vittorie e sconfitte totali di quell'ID in tutto il gruppo di avatar. In questo modo si ottengono le sequenze ordinate di ranks dei vari ID da poter confrontare tra gruppi diversi.

## 2. Kendall Tau Test

Vengono calcolati  $\tau$  e  $p$  mediante il test Tau fornendo in input le sequenze ordinate dei rank di *weights sequence ID*. Queste sequenze, come in recedenza menzionato, non vengono ordinate mediante il ranking ma il loro ordinamento rimane quello dei *weights sequence ID* in maniera crescente. Così facendo si confronta qual'è il grado di associatività tra le varie  $wX$  considerando i voti collezionati durante il test. Per una spiegazione dettagliata del test statistico che calcola il grado di associatività vedere l'appendice B.

## 4.2 Applicazione Web

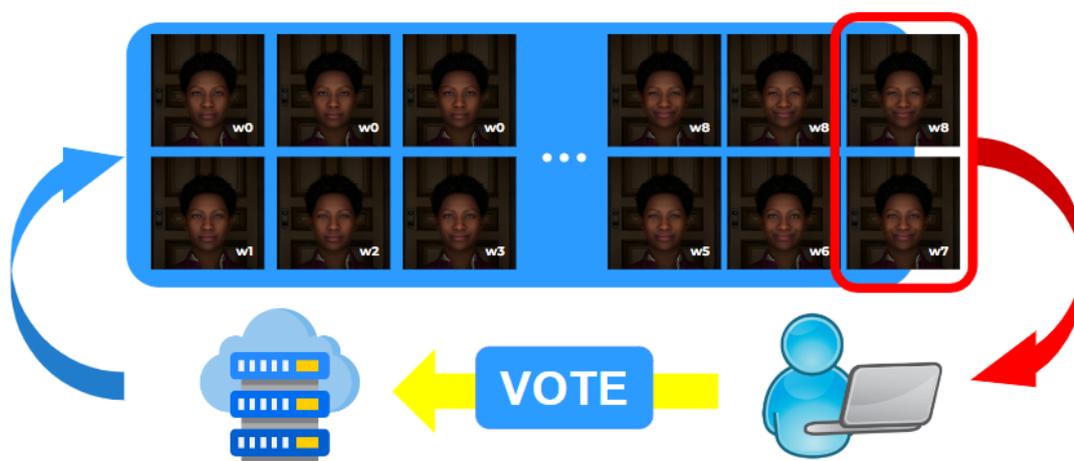
L'applicazione è stata sviluppata con JavaScript, HTML e CSS per quanto riguarda il front-end e NodeJS per il back-end. Il server è stato hostato su una macchina virtuale EC2 di Amazon AWS. Esso gestisce le richieste dei client, la distribuzione delle immagini mediante algoritmo *Round Robin* e l'accesso al database per la registrazione dei record risultanti dal test. Il database utilizzato è Firebase Real-Time, soluzione NoSQL che struttura i dati ad albero secondo uno schema chiave-valore come in un file JSON. Le chiavi o i valori possono essere *int*, *float*, *string*, altri oggetti o array di questi.

### 4.2.1 Design

L'applicazione è composta da una singola pagina avente un box di testo, due immagini ed un pulsante. Le due immagini vengono selezionate ed inviate dal server secondo la logica del 4.2.

L'utente può selezionare una delle due immagini ed inviare la selezione dell'immagine cliccando sul tasto *NEXT*. A questo punto viene inviata una richiesta al server che provvede ad aggiungere un nuovo record sul database.

Nella figura 4.2 è rappresentato lo schema del funzionamento della web app. Il server genera delle queue secondo l'algoritmo di Round Robin che provvede, ad ogni richiesta da parte di un client a prelevare una coppia di immagini



**Figura 4.2:** Schema web app: dalla queue generata dal server viene estratta una coppia per richiesta

Considerato il possibile elevato numero di richieste tra client e server è stato scelto NodeJS lato server per soddisfare questa esigenza. NodeJs infatti gestisce tutte le richieste che riceve dai client in maniera asincrona, eseguendo ognuno di queste su un thread separato che opera parallelamente agli altri.

Sul database i record vengono strutturati avendo *key* identificativo delle due immagini presentate all'utente separate dal carattere *-*, ed ordinate in maniera da avere prima la vincente e poi la perdente. *value* memorizza il numero di vittorie della prima immagine rispetto alla seconda.

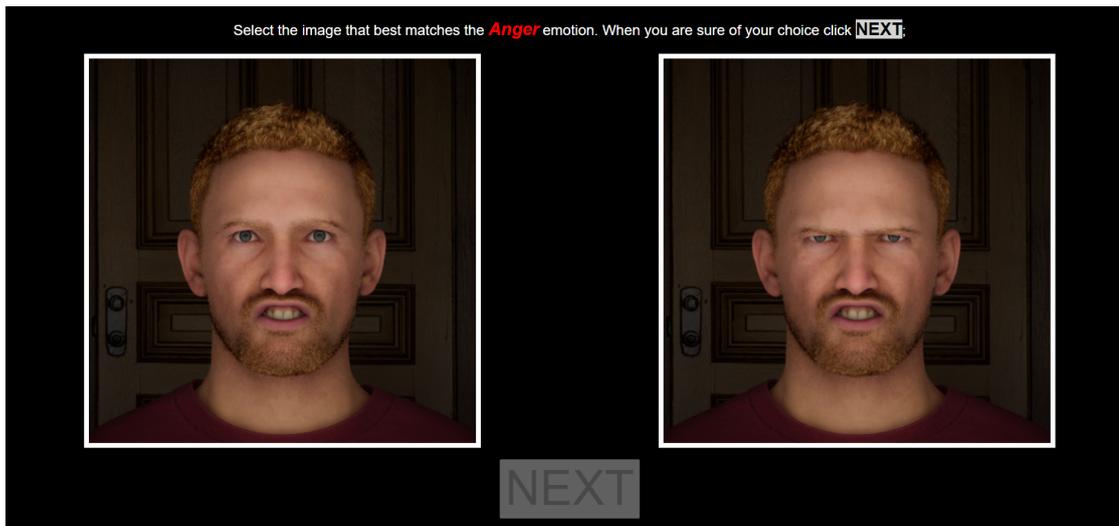


Figura 4.3: Layout della web page

- m7\_e0\_w22-m7\_e0\_w35: 1
- m7\_e0\_w22-m7\_e0\_w36: 2
- m7\_e0\_w22-m7\_e0\_w39: 3

Figura 4.4: Esempio di records sul database.

### 4.2.2 Algoritmo *Round Robin*

Le immagini vengono distribuite dal server seguendo un algoritmo di Round Robin. Esso si assicura che dato un avatar, ed una emozione, gli screenshot corrispondenti a tutti i pesi verranno confrontati tra di loro uno ad uno. Si evita quindi, per esempio, che l'immagine A venga confrontata con l'immagine B due volte prima che tutte le altre non siano state confrontate almeno una volta.

Le immagini da confrontare vengono immagazzinate in delle queue da cui ad ogni richiesta vengono prelevate le due immagini da confrontare. Per ogni avatar ed ogni emozione esiste una queue dedicata che ordina gli idici  $wX$  corrispondenti. L'algoritmo agisce direttamente nel momento in cui le queue vengono generate, inserendo le coppie di  $wX$  in maniera

contigua tra loro 4.5.

```

m22: {
  e0: [
    9, 15, 36, 56, 60, 24, 52, 20, 8, 55, 57, 25,
    14, 81, 7, 73, 42, 76, 31, 54, 21, 43, 27, 22,
    6, 0, 58, 51, 30, 10, 32, 33, 80, 46, 41, 75,
    26, 18, 16, 79, 48, 50, 68, 19, 62, 39, 28, 70,
    4, 38, 77, 11, 5, 29, 44, 12, 74, 47, 1, 40,
    67, 64, 71, 2, 37, 78, 66, 72, 59, 13, 23, 45,
    53, 61, 63, 3, 34, 69, 17, 65, 35, 49, 15, 36,
    56, 60, 24, 52, 20, 8, 55, 57, 25, 14, 81, 7,
    73, 42, 76, 31,
    ... 13184 more items
  ]
}

```

**Figura 4.5:** Esempio di queue relativa ad  $m22\_e0$  contenente gli weights array ID  $wX$  da inviare

Il round robin, considerati  $n$  screenshot (dato un avatar ed un'emozione) prevede la generazione di  $n$  round aventi  $n - 1$  confronti ciascuno.

$\times$	w1	w2	w3	w4	w5	w6	w7	w8	w9	w10	w11	w12	w13
w1	1	2	3	4	5	6	7	8	9	10	11	12	13
w2	2	3	4	5	6	7	8	9	10	11	12	13	1
w3	3	4	5	6	7	8	9	10	11	12	13	1	2
w4	4	5	6	7	8	9	10	11	12	13	1	2	3
w5	5	6	7	8	9	10	11	12	13	1	2	3	4
w6	6	7	8	9	10	11	12	13	1	2	3	4	5
w7	7	8	9	10	11	12	13	1	2	3	4	5	6
w8	8	9	10	11	12	13	1	2	3	4	5	6	7
w9	9	10	11	12	13	1	2	3	4	5	6	7	8
w10	10	11	12	13	1	2	3	4	5	6	7	8	9
w11	11	12	13	1	2	3	4	5	6	7	8	9	10
w12	12	13	1	2	3	4	5	6	7	8	9	10	11
w13	13	1	2	3	4	5	6	7	8	9	10	11	12

**Tabella 4.2:** Esempio di scheduling Round Robin con 14 immagini. Esempio di selezione coppie al round 5. Tutte le corrispondenze  $wX : wX$  vanno intese come  $wX : w0$

Nella tabella 4.2 viene mostrato un esempio di scheduling dell'algoritmo

di round robin considerando 14 immagini. Nella prima riga e nella prima colonna sono rappresentati gli id corrispondenti ai weights id. Ogni valore all'interno della tabella è il numero di round a cui le due immagini vengono confrontate. Ogni diagonale della tabella rappresenta così un round.

### 4.3 Sistema di Ranking

Considerata la modalità di confronto delle immagini esplicita nel dettaglio nella sezione 4.2, si è scelto di utilizzare un sistema di ranking simile a quello di una classifica di un torneo. I dati contenuti nel database, sono i conteggi del numero di vittorie di una immagine A contro l'immagine B. Ogni volta che immagine A vince contro B, ovvero l'utente esaminato seleziona A quando confrontata con B, viene incrementato di 1 il contatore. Si calcola dunque il numero totale di vittorie  $wins_x$  di una immagine  $x$  ed il numero di sconfitte  $losses_x$  e si procede a realizzare un ranking  $rank_x$  assoluto mediante la formula:

$$rank_x = \frac{wins_x - losses_x}{wins_x + losses_x} \quad (4.1)$$

In questo modo si ottiene un rank per ogni immagine confrontabile con le altre immagini (di uno stesso avatar, con stessa emozione) e normalizzato su un range di valori da 0 ad 1. Più è alto il numero di volte in cui l'immagine è stata valutata, più alta è l'accuratezza del ranking.

È importante notare che il numero di volte in cui un'immagine viene confrontata determina il  $\Delta rank$  ovvero la minima variazione che può esistere tra un rank di un'immagine e una qualsiasi altra. Per avere quindi una classifica attendibile, considerando  $n_x = wins_x + losses_x$  si ha che servono almeno  $n_x - 1$  confronti tali che nel caso migliore si ottengano ranking tutti diversi tra loro e quindi perfettamente ordinabili. Il ranking che si ottiene è idealmente definito senza *tied ranks*, per approfondire si rimanda all'appendice B.

Il sistema così descritto, valido per  $x$  corrispondente ad una singola immagine, è generalizzabile con  $x$  corrispondente ad una *weights sequence* ma che considera più avatar, appartenenti ad uno stesso gruppo. Così  $rank_x$  viene calcolato sul totale di vittorie  $wins_x$  e sconfitte  $losses_x$  di tutti gli avatar del gruppo. Questa modalità di calcolo è necessaria al

fine di eseguire gli step successivi allo 0 in cui si vuole confrontare gruppi diversi di avatar come mostrato in 4.1.

# Capitolo 5

## Risultati e Discussione

In questo capitolo vengono presentati i primi risultati derivanti dall'applicazione del framework sperimentale per le emozioni *Happiness* e *Surprise* ad AUs ridotte (si rimanda a 4). A seguito, sono discussi i limiti dello studio e come essi possano aver influenzato la sperimentazione.

### 5.1 Analisi dei risultati

L'analisi viene effettuata emozione per emozione su tutti e 24 gli avatar secondo lo schema ad albero precedentemente descritto (vedi 4.1).

#### 5.1.1 Risultati *Happiness*

Tabella 5.1: Risultati step 0

Gruppo Avatar	Coppia Avatar	$\tau$	$P_{value}$
Metahuman Black F	m0 : m1	0.584	0.02
Metahuman Black F	m0 : m2	0.809	0.001
Metahuman Black F	m0 : m3	0.614	0.015
Metahuman Black F	m1 : m2	0.556	0.029
Metahuman Black F	m1 : m3	0.539	0.031
Metahuman Black F	m2 : m3	0.539	0.031
Metahuman Black F	Number above threshold tau=0.3	100.0%	

Tabella 5.1: Risultati step 0

Gruppo Avatar	Coppia Avatar	$\tau$	$P_{value}$
Metahuman Black M	m8 : m9	0.886	0.0
Metahuman Black M	m8 : m10	0.584	0.02
Metahuman Black M	m8 : m11	0.92	0.0
Metahuman Black M	m9 : m10	0.584	0.02
Metahuman Black M	m9 : m11	0.874	0.001
Metahuman Black M	m10 : m11	0.477	0.058
Metahuman Black M	Number above threshold tau=0.3	100.0%	
Metahuman White F	m4 : m5	0.523	0.038
Metahuman White F	m4 : m6	0.795	0.002
Metahuman White F	m4 : m7	0.899	0.0
Metahuman White F	m5 : m6	0.432	0.087
Metahuman White F	m5 : m7	0.539	0.031
Metahuman White F	m6 : m7	0.764	0.002
Metahuman White F	Number above threshold tau=0.3	100.0%	
Metahuman White M	m12 : m13	0.511	0.047
Metahuman White M	m12 : m14	0.629	0.012
Metahuman White M	m12 : m15	0.719	0.004
Metahuman White M	m13 : m14	0.584	0.02
Metahuman White M	m13 : m15	0.629	0.012
Metahuman White M	m14 : m15	0.568	0.024
Metahuman White M	Number above threshold tau=0.3	100.0%	
RocketBox White F	m16 : m17	0.315	0.209
RocketBox White F	m16 : m18	0.022	1.0
RocketBox White F	m16 : m19	0.764	0.002
RocketBox White F	m17 : m18	0.405	0.106
RocketBox White F	m17 : m19	0.568	0.024
RocketBox White F	m18 : m19	0.225	0.369
RocketBox White F	Number above threshold tau=0.3	66.667%	
RocketBox White M	m20 : m21	0.6	0.017
RocketBox White M	m20 : m22	0.719	0.004
RocketBox White M	m20 : m23	0.556	0.029
RocketBox White M	m21 : m22	0.494	0.048

**Tabella 5.1:** Risultati step 0

Gruppo Avatar	Coppia Avatar	$\tau$	$P_{value}$
RocketBox White M	m21 : m23	0.689	0.005
RocketBox White M	m22 : m23	0.674	0.007
RocketBox White M	Number above threshold tau=0.3	100.0%	

I risultati ottenuti dai confronti allo step 0 ?? mostrano un alto grado di associatività. In tutti questi gruppi ogni confronto supera il valore soglia di  $\tau = 0.3$ , che indica un grado di associatività moderato ( $0.3 < \tau < 0.5$ ), superiore se confrontato al pmente random  $\tau = 0$ . Risulta un'unica eccezione nel caso di *RocketBox White F* dove i  $P_{value}$  assumono valori non accettabili ( $P_{value} > 0.05$ ) e dunque si verifica nella maggior parte di questi confronti associatività nulla. È probabile sia dovuto ad alcune caratteristiche del sesso femminile di RocketBox, poichè confrontato con quelli maschili dello stesso set il problema non si presenta.

**Tabella 5.2:** Risultati step 1, 2, 3

Gruppi Avatar	$\tau$	$P_{value}$
MetaHuman Black (F : M)	0.854	0.001
MetaHuman White (F : M)	0.867	0.0
RocketBox White (F : M)	0.822	0.0
MetaHuman (Black : White)	0.733	0.002
All (MetaHuman : RocketBox)	0.867	0.0

Si osserva come negli step 1, 2, 3  $\tau$  aumenta raggiungendo un grado di associatività molto forte ( $0.7 < \tau < 1.0$ ). Questo denota la conferma dell'ipotesi di universalità 1.6, mostrando come sequenze di weights array siano state ordinate con una stessa tendenza. Ciò dimostra che le espressività FACS dell'emozione *Happiness* vengono percepite allo stesso modo dall'utente anche su avatar aventi caratteristiche di sesso, età, etnia e qualità grafiche diverse.

La migliore espressione risultante dall'analisi è quella definita dal *weights array w8*, corrispondente ad AUs [6,12] aventi peso [0.95,0.95].

weights_id	wins	losses	rank
0	145	355	-0.420000
1	163	339	-0.350598
2	121	378	-0.515030
3	254	245	0.018036
4	304	190	0.230769
5	307	191	0.232932
6	388	108	0.564516
7	393	106	0.575150
8	406	92	0.630522
9	12	489	-0.952096

Tabella 5.3: Risultati globali di *happiness*

### 5.1.2 Risultati *Surprise*

Tabella 5.4: Risultati step 0

Gruppo Avatar	Coppia Avatar	$\tau$	$P_{value}$
Metahuman Black F	m0 : m1	0.326	0.0
Metahuman Black F	m0 : m2	0.325	0.0
Metahuman Black F	m0 : m3	0.451	0.0
Metahuman Black F	m1 : m2	0.292	0.0
Metahuman Black F	m1 : m3	0.242	0.003
Metahuman Black F	m2 : m3	0.278	0.001
Metahuman Black F	Number above threshold tau=0.3	50.0	
Metahuman Black M	m8 : m9	0.357	0.0
Metahuman Black M	m8 : m10	0.181	0.025
Metahuman Black M	m8 : m11	0.29	0.0
Metahuman Black M	m9 : m10	0.286	0.0
Metahuman Black M	m9 : m11	0.332	0.0
Metahuman Black M	m10 : m11	0.182	0.023
Metahuman Black M	Number above threshold tau=0.3	33.333	
Metahuman White F	m4 : m5	0.288	0.0

**Tabella 5.4:** Risultati step 0

<b>Gruppo Avatar</b>	<b>Coppia Avatar</b>	$\tau$	$P_{value}$
Metahuman White F	m4 : m6	0.31	0.0
Metahuman White F	m4 : m7	0.276	0.001
Metahuman White F	m5 : m6	0.279	0.001
Metahuman White F	m5 : m7	0.25	0.002
Metahuman White F	m6 : m7	0.254	0.001
Metahuman White F	Number above threshold tau=0.3	16.667	
Metahuman White M	m12 : m13	0.305	0.0
Metahuman White M	m12 : m14	0.232	0.004
Metahuman White M	m12 : m15	0.249	0.002
Metahuman White M	m13 : m14	0.359	0.0
Metahuman White M	m13 : m15	0.308	0.0
Metahuman White M	m14 : m15	0.367	0.0
Metahuman White M	Number above threshold tau=0.3	66.667	
RocketBox White F	m16 : m17	0.361	0.0
RocketBox White F	m16 : m18	0.366	0.0
RocketBox White F	m16 : m19	0.436	0.0
RocketBox White F	m17 : m18	0.322	0.0
RocketBox White F	m17 : m19	0.242	0.003
RocketBox White F	m18 : m19	0.442	0.0
RocketBox White F	Number above threshold tau=0.3	83.333	
RocketBox White M	m20 : m21	0.297	0.0
RocketBox White M	m20 : m22	0.316	0.0
RocketBox White M	m20 : m23	0.225	0.006
RocketBox White M	m21 : m22	0.239	0.003
RocketBox White M	m21 : m23	0.31	0.0
RocketBox White M	m22 : m23	0.213	0.009
RocketBox White M	Number above threshold tau=0.3	33.333	

I risultati dello step zero mostrano valori di  $\tau$  misti molto diversi tra i vari gruppi. La maggior parte di questi non supera la soglia di  $\tau = 0.3$  denotando una debole associatività ( $0 < \tau < 0.3$ ). Il risultato in sè indicherebbe l'impossibilità di conferma dell'ipotesi 1.6, se confrontato

però con i risultati di *Happiness* è incoerente. Questo si ipotizza possa essere dovuto alla mancanza di abbastanza voti per poter calcolarne il rank. I risultati infatti sono preliminari e parziali in quanto necessiterebbero di ulteriore raccolta. Per una spiegazione più approfondita si rimanda alla sezione 5.2. Se questa debole associatività non dovesse venire confermata non si potrebbe proseguire nell'analisi, che considera, per ogni step, di unire gruppi via via più affini. Avendo beneficio di dubbio procediamo nell'analisi.

**Tabella 5.5:** Risultati step 1, 2, 3

Gruppi Avatar	$\tau$	$P_{value}$
MetaHuman Black (F vs M)	0.508	0.0
MetaHuman White (F vs M)	0.615	0.0
RocketBox White (F vs M)	0.536	0.0
MetaHuman (Black vs White)	0.636	0.0
All (MetaHuman vs RocketBox)	0.604	0.0

Si può osservare una forte associatività ( $0.5 < \tau < 0.7$ ) in tutti gli step 1, 2, 3 confermando nuovamente l'ipotesi 1.6 di universalità (considerando il caso in cui in step 0 sia confermata).

La migliore espressione risultante dall'analisi è quella definita dal *weights array w25*, corrispondente ad AUs [1,2,5,25,26,27] aventi pesi [0.675,0.95,0.95,0.4,0.4,0.4].

## 5.2 Limiti di ricerca

Lo studio condotto presenta numerosi limiti legati principalmente ai campioni raccolti e considerando il framework sperimentale utilizzato. In particolare si citano gli aspetti più rilevanti:

- Per quanto riguarda *Happiness* sono stati raccolti un numero sufficiente di voti per far sì che ogni immagine sia stata confrontata almeno una volta con tutte le altre (circa 2000 confronti). Nel caso invece di *Surprise*, avente una maggior complessità (4 AUs invece di 2 AUs) i dati raccolti non coprono tutti i confronti necessari. Le immagini

weights_id	wins	losses	rank
20	234	153	0.209302
21	258	130	0.329897
22	260	128	0.340206
23	237	147	0.234375
24	261	125	0.352332
<b>25</b>	<b>282</b>	<b>103</b>	<b>0.464935</b>
26	269	116	0.397403
27	190	195	-0.012987
28	169	216	-0.122078
29	176	208	-0.083333

**Tabella 5.6:** Estratto dei risultati globali di *surprise*

infatti sono state confrontate con circa il 20% di tutte le immagini necessarie. I confronti ottenuti qui sono stati solamente 12.000 su un totale di 70.000 circa, mancando il confronto con 80% del dataset. Pur con questo limite comunque i risultati più generici sono promettenti.

- I set utilizzati sono stati solamente 2, RocketBox e MetaHuman. Sarebbe necessario avere un dataset maggiormente vario con altri avatar set tra i più utilizzati nel settore VR dei VH. Inoltre la mancanza di un gruppo di avatar RocketBox Black non permette di valutare l'influenza del fotorealismo sul confronto tra avatar aventi altre caratteristiche diverse.
- Sono state testate solamente 2 emozioni di base su 6 (una come mostrato solo in maniera parziale), per validare l'ipotesi di universalità nell'applicazione del FACS per le *basic emotions* sarebbe necessario effettuare uno studio anche sulle rimanenti: *Anger*, *Disgust*, *Fear*, *Sadness*.
- Di 425 partecipanti alla sperimentazione il 70% ha votato le coppie di immagini tramite mobile. La web app infatti è stata distribuita per essere accessibile da qualunque dispositivo dotato di un browser mediante la stessa UI. Questo aspetto può aver influito in particolare modo per quei confronti tra immagini aventi pesi molto simili che

richiedevano una maggior attenzione al dettaglio grafico. La grandezza del display potrebbe giocare un ruolo fondamentale in questi casi. Sembra comunque, almeno per quanto riguarda *Happiness*, che ciò non abbia influito più di tanto sul risultato finale.

# Capitolo 6

## Conclusioni

Alla luce dei risultati ottenuti si può confermare, almeno in maniera preliminare, sia l'efficacia del framework di sperimentazione che l'applicazione del sistema FACS al fine di implementare l'espressività facciale dei VH.

È evidente come lo studio sia piuttosto limitato da fattori quali, il campione testato, il numero di emozioni e la grandezza del dataset di avatar. Si potrebbe inoltre considerare che il numero di pesi scelti (0.4, 0.675, 0.95) è ristretto a 3, cosa che non garantisce una grande granularità alle AUs. Nonostante ciò i risultati sono promettenti nel confermare le ipotesi proposte dalla tesi.

Con ulteriori studi si potrebbe attestare l'utilità del sistema FACS definendo un set di emozioni pronte all'uso per lo sviluppo di VH. Questo sarebbe un primo passo verso la standardizzazione dell'emotività facciale, aspetto assente nella letteratura di riferimento.

Considerando l'emozione *Happiness* ad esempio, risulta che, l'espressività avente entrambe le AUs [6, 12] con pesi di valore 0.95, sia quella preferibile. Ipotizzando di dover implementare la stessa emozione ed espressione su un qualunque altro avatar dotato di FACS si potrebbe applicare gli stessi pesi 6.1.

### 6.1 Ricerca futura

Considerati i limiti dello studio ed i promettenti risultati è necessario proseguire in questa direzione di ricerca. A partire proprio dai limiti



**Figura 6.1:** Migliore combinazione di pesi ( $w_8$  di *Happiness*) applicata a vari avatar

sarebbe utile migliorare il framework al fine da invogliare maggiormente l'utente a votare ed ottenere così un maggior numero di voti.

Questo ad esempio sarebbe possibile implementando nella web app un qualche sistema di reward o aspetto di gamification. Ciò consentirebbe di estendere il dataset di test e sperimentare più emozioni su un maggior numero di avatar e con una scala di pesi aventi una granularità maggiore

di 3 livelli.

Lo stesso framework si potrebbe inoltre sfruttare anche con video degli avatar interpolando pesi diversi, ed introducendo così la variabile tempo nella percezione dell'emozione. Ciò verificherebbe l'utilità del sistema FACS nel caso in cui sia necessario implementare una transizione tra stati d'animo espressi dai VH, di fondamentale importanza nella maggior parte delle simulazioni di interesse.

Considerando nuovamente il concetto di empatia si potrebbe successivamente proseguire nella ricerca di come l'espressività mediante FACS possa essere integrata da altri aspetti della sfera sociale e dell'interazione inter-personale. Menzionando i principali, in ordine sparso si ricordano:

- L' **interazione** tra l'utente ed il VH, cioè tutti quegli aspetti che riguardano un inter-scambio di stimoli tra i due. Come nella realtà il modo migliore di conoscere e comprendere una persona è passarci del tempo assieme, interagendo con lui in una qualunque forma che sia parlando, giocando, ballando o, nei casi di interesse medico ad esempio, operando sul paziente o verificandone il suo stato di salute.



**Figura 6.2:** Esempio di interazione con VH

- Il **comportamento** del VH, ovvero come si comporta ed interagisce con l'ambiente. Le animazioni dell'intero corpo del VH ricoprono un

importante ruolo nella decodifica del suo stato emotivo. Basti ad esempio pensare allo storytelling di un personaggio in un film. Nelle pellicole più acclamate della storia del cinema le persone vengono conosciute dallo spettatore proprio tramite l'osservazione delle loro reazioni agli stimoli dell'ambiente e delle loro azioni. Si può ipotizzare che lo stesso avvenga in una simulazione VR tanto quanto accade nella realtà.



**Figura 6.3:** Comportamento del paziente al dolore causato dalla ferita

- La **personalità**. Quello che rende umane le persone in fondo è la loro unicità, il loro modo di essere, i loro pregi e difetti. Questo aspetto, forse il più avanguardistico rispetto al settore di studi delinea degli obiettivi di studio di lungo periodo.

Come descritto alla sezione 1.4 è evidente come in letteratura questi aspetti siano importanti nella creazione di un legame empatico ed interessino anche la risoluzione di problemi come l'*Uncanny Valley*. È bene notare come di fatto come tutte queste componenti siano strettamente correlate le une alle altre. È pertanto necessario che, in studi futuri, ci sia una valutazione di questi aspetti in maniera parallela e complementare, od osservandone gli effetti nella loro interezza ed interdipendenza.

# Appendice A

## Disposizioni

Si definisce disposizione una sequenza ordinata di  $n$  elementi selezionati da un insieme di  $k$  elementi. A seconda che questi  $k$  elementi possano ripetersi nella sequenza si definiscono le disposizioni semplici e quelle con ripetizione.

### A.1 Disposizioni semplici

Una disposizione semplice è una disposizione in cui considerata una sequenza ordinata di  $n$  elementi questa non presenta elementi uguali, e dunque gli elementi di  $k$  compaiono una sola volta.

La condizione di esistenza è tale per cui  $n \geq k$ . Il numero di disposizioni totali dati  $n$  e  $k$  è dato da:

$$D_{n,k} = \frac{n!}{(n-k)!} \quad (\text{A.1})$$

### A.2 Disposizioni con ripetizione

Una disposizione con ripetizione è una disposizione in cui considerata una sequenza ordinata di  $n$  elementi questa può presentare elementi uguali, e dunque gli elementi di  $k$  possono comparire più volte.

Il numero di disposizioni totali dati  $n$  e  $k$  è dato da:

$$D_{n,k} = n^k \quad (\text{A.2})$$

# Appendice B

## Kendall Tau Test

Il test Tau è una misura della correlazione tra due quantità. In particolare viene utilizzato per misurare la correlazione tra due sequenze ordinate, della stessa lunghezza, secondo un dato sistema di ranking. I valori risultanti dal test sono due:

- $\tau \in [-1,1] \in \mathbf{R}$  coefficiente di correlazione
- $p \in [0,1] \in \mathbf{R}$  probabilità di *null hypothesis* (ovvero che le sequenze siano così ordinate per pura randomicità)

Intuitivamente,  $\tau$  è alto quando il grado di correlazione è alto (idealmente 1 avendo sequenze identiche), mentre è basso quando si ha correlazione inversa (idealmente -1 avendo sequenze inversamente correlate). Il valore 0 indica che non è possibile stabilire alcun tipo di correlazione.

$p$  allo stesso modo quando è alto indica la probabilità che il risultato di  $\tau$  è randomico (idealmente 1 quando è puramente random), e basso quando il risultato non è randomico (idealmente 0 quando non è presente alcuna randomicità).

### B.1 Test A

Il test  $\tau_a$  verifica il grado di associazione tra due sequenze che non presentano *tied ranks* ovvero hanno tutti gli elementi aventi rank diverso tra loro. Data  $n$  lunghezza delle due sequenze si definisce ed  $n_c$  numero di coppie concordanti, mentre  $n_d$  numero di coppie discordanti si definisce:

$$n_0 = n(n - 1)/2 \quad (\text{B.1})$$

$$\tau_A = \frac{n_c - n_d}{n_0} \quad (\text{B.2})$$

Per quanto riguarda il calcolo della probabilità di *null hypothesis*, considerando le due sequenze come variabili statisticamente indipendenti, si ha che:

$$p_A = \frac{3(n_c - n_d)}{\sqrt{n(n - 1)(2n + 5)/2}} \quad (\text{B.3})$$

## B.2 Test B

Il test  $\tau_b$  a differenza di  $\tau_a$  considera anche i *tied ranks*. Si definisce quindi:

$$\begin{aligned} n_1 &= \sum_i t_i(t_i - 1)/2 \\ n_2 &= \sum_j u_j(u_j - 1)/2 \\ t_i &= \text{numero di tied values nel } i\text{-esimo gruppo di ties della prima sequenza} \\ u_j &= \text{numero di tied values nel } j\text{-esimo gruppo di ties della seconda sequenza} \end{aligned} \quad (\text{B.4})$$

$$\tau_B = \frac{n_c - n_d}{\sqrt{(n_0 - n_1)(n_0 - n_2)}} \quad (\text{B.5})$$

Il calcolo della probabilità di *null hypothesis* sempre considerando le due sequenze come statisticamente indipendenti è definito da:

$$\begin{aligned} v &= (v_0 - v_t - v_u)/18 + v_1 + v_2 \\ v_0 &= n(n - 1)(2n + 5) \\ v_t &= \sum_i t_i(t_i - 1)(2t_i + 5) \\ v_u &= \sum_j u_j(u_j - 1)(2u_j + 5) \\ v_1 &= \sum_i t_i(t_i - 1) \sum_j u_j(u_j - 1)/(2n(n - 1)) \\ v_2 &= \sum_i t_i(t_i - 1)(t_i - 2) \sum_j u_j(u_j - 1)(u_j - 2)/(9n(n - 1)(n - 2)) \end{aligned} \quad (\text{B.6})$$

$$p_B = \frac{n_c - n_d}{\sqrt{v}} \quad (\text{B.7})$$

# Ringraziamenti

Dopo anni di studio, passione e costante dedizione questo lungo viaggio volge al termine.

Tra tutte le persone che sul mio cammino di vita ho incontrato vorrei ringraziare prima di tutto i miei genitori Filippo, Valeria e mio fratello, Marco, che mi hanno ascoltato pur non capendomi, che mi hanno assecondato pur non vedendomi.

Per secondi vorrei ringraziare i miei amici e parenti, chi qui presente e chi non, per avermi accompagnato tra gioie e dolori della vita, e da una vita, sia nei miei studi che al di fuori di questi.

Per terza ma non per ordine di importanza, vorrei ringraziare la mia ragazza, Margherita, che mi ha sostenuto anche nei momenti più bui di questo ultimo anno di studi e con la quale ho riacquisito la mia sicurezza.

Vorrei inoltre ringraziare le mie coinquiline, per avermi sopportato durante questo periodo difficile ed aver condiviso momenti indimenticabili durante il percorso.

Un grazie speciale vorrei dedicare al professor Bottino e ai dottorandi e ricercatori del laboratorio: Francesco, Leonardo ed Edoardo che mi hanno aiutato e con cui ho collaborato tra consigli e pause caffè.

Un ultimo grazie vorrei dedicare al Politecnico di Torino ed ai miei compagni di corso, per avermi permesso di studiare ciò che altrove non era possibile. In questo corso di laurea ho trovato sfogo alle mie passioni, e forse, trovato anche la mia vocazione, nella speranza che un giorno scienza ed arte possano fondersi nuovamente.

*Stefano Calzolari*

# Bibliografia e Sitografia

- [1] Julia C Campbell, Matthew Jensen Hays, Mark Core, Mike Birch, Matt Bosack e Richard E Clark. «Interpersonal and Leadership Skills: Using Virtual Humans to Teach New Officers». In: (2011) (cit. a p. 3).
- [2] Valentin Schwind, Katrin Wolf e Niels Henze. «Avoiding the uncanny valley in virtual character design». In: *interactions* 25.5 (2018), pp. 45–49 (cit. a p. 4).
- [3] Mike Seymour, Kai Riemer e Judy Kay. «Mapping Beyond the Uncanny Valley: A Delphi Study on Aiding Adoption of Realistic Digital Faces». In: () (cit. alle pp. 5, 6, 14, 15).
- [4] Mike Seymour, Lingyao Yuan, Kai Riemer e Alan R Dennis. «Crossing the Uncanny Valley? Understanding Affinity, Trustworthiness, and Preference for More Realistic Virtual Humans in Immersive Environments». In: () (cit. alle pp. 6, 14).
- [5] Matias Volonte, Reza Ghaiumy Anaraky, Rohith Venkatakrisnan, Roshan Venkatakrisnan, Bart P Knijnenburg, Andrew T Duchowski e Sabarish V Babu. «Empirical evaluation and pathway modeling of visual attention to virtual humans in an appearance fidelity continuum». In: *Journal on Multimodal User Interfaces* 15.2 (2021), pp. 109–119 (cit. a p. 6).
- [6] Katja Zibrek, Sean Martin e Rachel McDonnell. «Is photorealism important for perception of expressive virtual humans in virtual reality?» In: *ACM Transactions on Applied Perception (TAP)* 16.3 (2019), pp. 1–19 (cit. alle pp. 6, 15).

- [7] Matias Volante, Sabarish V Babu, Himanshu Chaturvedi, Nathan Newsome, Elham Ebrahimi, Tania Roy, Shaundra B Daily e Tracy Fasolino. «Effects of virtual human appearance fidelity on emotion contagion in affective inter-personal simulations». In: *IEEE transactions on visualization and computer graphics* 22.4 (2016), pp. 1326–1335 (cit. alle pp. 6, 14).
- [8] *The Matrix Awakens. An Unreal Engine 5 Experience*. Dic. 2021. URL: <https://www.youtube.com/watch?v=WU0gvPcc3jQ> (cit. alle pp. 6, 13).
- [9] Will Knight. *You can now practice firing someone in virtual reality*. URL: <https://www.technologyreview.com/2019/08/09/133799/you-can-now-practice-firing-someone-in-virtual-reality/> (visitato il 31/03/2023) (cit. a p. 8).
- [10] Paul Ekman e Wallace V Friesen. «Facial action coding system». In: *Environmental Psychology & Nonverbal Behavior* (1978) (cit. a p. 10).
- [11] Paul Ekman, E Richard Sorenson e Wallace V Friesen. «Pan-cultural elements in facial displays of emotion». In: *Science* 164.3875 (1969), pp. 86–88 (cit. a p. 10).
- [12] Paul Ekman, Wallace V Friesen e Joseph C Hager. *Facial Action Coding System: Facial action coding system: the manual: on CD-ROM*. Research Nexus, 2002 (cit. a p. 10).
- [13] Mehdi Ghayoumi e Arvind K Bansal. «Unifying geometric features and facial action units for improved performance of facial expression analysis». In: *arXiv preprint arXiv:1606.00822* (2016) (cit. a p. 10).
- [14] V Vinayagamoorthy, M Gillies, A Steed, E Tanguy, X Pan, C Loscos e M Slater. «Building Expression into Virtual Characters». In: () (cit. alle pp. 13–15).
- [15] Saverio Cinieri, Bill Kapralos, Alvaro Uribe-Quevedo e Fabrizio Lamberti. «Eye Tracking and Speech Driven Human-Avatar Emotion-Based Communication». In: *2020 IEEE 8th International Conference on Serious Games and Applications for Health (SeGAH)*. 2020 IEEE 8th International Conference on Serious Games and Applications for Health(SeGAH). Vancouver, BC, Canada: IEEE, ago. 2020, pp. 1–5.

- ISBN: 978-1-72819-042-6. DOI: 10.1109/SeGAH49190.2020.9201874. URL: <https://ieeexplore.ieee.org/document/9201874/> (visitato il 29/03/2023) (cit. a p. 13).
- [16] Nesreen Mejbri, Fathi Essalmi, Mohamed Jemni e Bader A. Alyoubi. «Trends in the use of affective computing in e-learning environments». In: *Education and Information Technologies* 27.3 (apr. 2022), pp. 3867–3889. ISSN: 1360-2357, 1573-7608. DOI: 10.1007/s10639-021-10769-9. URL: <https://link.springer.com/10.1007/s10639-021-10769-9> (visitato il 29/03/2023) (cit. a p. 13).
- [17] Guillermo Carbonell e Jonas Schild. «Conceptual Design of Emotional and Pain Expressions of a Virtual Patient in a Virtual Reality Training for Paramedics». In: *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). Christchurch, New Zealand: IEEE, mar. 2022, pp. 222–226. ISBN: 978-1-66548-402-2. DOI: 10.1109/VRW55335.2022.00053. URL: <https://ieeexplore.ieee.org/document/9757401/> (visitato il 15/03/2023) (cit. alle pp. 13–15).
- [18] Matias Volonte, Sabarish V. Babu, Himanshu Chaturvedi, Nathan Newsome, Elham Ebrahimi, Tania Roy, Shaundra B. Daily e Tracy Fasolino. «Effects of Virtual Human Appearance Fidelity on Emotion Contagion in Affective Inter-Personal Simulations». In: *IEEE Transactions on Visualization and Computer Graphics* 22.4 (21 apr. 2016), pp. 1326–1335. ISSN: 1077-2626, 1941-0506, 2160-9306. DOI: 10.1109/TVCG.2016.2518158. URL: <https://ieeexplore.ieee.org/document/7383334/> (visitato il 25/03/2023) (cit. a p. 13).
- [19] Matias Volonte, Reza Ghaiumy Anaraky, Rohith Venkatakrisnan, Roshan Venkatakrisnan, Bart P. Knijnenburg, Andrew T. Duchowski e Sabarish V. Babu. «Empirical evaluation and pathway modeling of visual attention to virtual humans in an appearance fidelity continuum». In: *Journal on Multimodal User Interfaces* 15.2 (giu. 2021), pp. 109–119. ISSN: 1783-7677, 1783-8738. DOI: 10.1007/s12193-020-00341-z. URL: <https://link.springer.com/10.1007/s12193-020-00341-z> (visitato il 25/03/2023) (cit. a p. 13).

- [20] Andrew Robb, Regis Kopper, Ravi Ambani, Farda Qayyum, David Lind, Li-Ming Su e Benjamin Lok. «Leveraging Virtual Humans to Effectively Prepare Learners for Stressful Interpersonal Experiences». In: *IEEE Transactions on Visualization and Computer Graphics* 19.4 (apr. 2013), pp. 662–670. ISSN: 1077-2626. DOI: 10.1109/TVCG.2013.35. URL: <http://ieeexplore.ieee.org/document/6479207/> (visitato il 29/03/2023) (cit. a p. 13).
- [21] Ilona Buchem, Oskar Stamm, Susan Vorweg, Kai Kruschel e Kristain Hildebrand. «Evaluation of Rapport in Human-Agent Interactions with a VR Trainer after a 6-week Exergame Training for Senior Users with Hypertension.» In: 13th International Conference on Applied Human Factors and Ergonomics (AHFE 2022). 2022. DOI: 10.54941/ahfe1002071. URL: [https://openaccess.cms-conferences.org/publications/book/978-1-958651-26-1/article/978-1-958651-26-1\\_17](https://openaccess.cms-conferences.org/publications/book/978-1-958651-26-1/article/978-1-958651-26-1_17) (visitato il 25/03/2023) (cit. a p. 14).
- [22] Kogilathah Segaran, Ahmad Zamzuri Mohamad Ali e Wee Hoe Tan. «Mediating effects of avatar realism on motivation in game based learning-review». In: *Online Journal for TVET Practitioners* 4.2 (2019), pp. 83–92 (cit. a p. 14).
- [23] Jacqueline D. Bailey e Karen L. Blackmore. «A Novel Method of Exploring the Uncanny Valley in Avatar Gender(Sex) and Realism Using Electromyography». In: *Big Data and Cognitive Computing* 6.2 (30 mag. 2022), p. 61. ISSN: 2504-2289. DOI: 10.3390/bdcc6020061. URL: <https://www.mdpi.com/2504-2289/6/2/61> (visitato il 15/03/2023) (cit. a p. 14).
- [24] Victoria Groom, Clifford Nass, Tina Chen, Alexia Nielsen, James K. Scarborough e Erica Robles. «Evaluating the effects of behavioral realism in embodied agents». In: *International Journal of Human-Computer Studies* 67.10 (ott. 2009), pp. 842–849. ISSN: 10715819. DOI: 10.1016/j.ijhcs.2009.07.001. URL: <https://linkinghub.elsevier.com/retrieve/pii/S1071581909000767> (visitato il 25/03/2023) (cit. a p. 14).
- [25] Yukiko Nakano, Michael Neff, Ana Paiva e Marilyn Walker, cur. *Intelligent Virtual Agents: 12th International Conference, IVA 2012, Santa Cruz, CA, USA, September, 12-14, 2012. Proceedings*. Red. da

- David Hutchison et al. Vol. 7502. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. ISBN: 978-3-642-33196-1 978-3-642-33197-8. DOI: 10.1007/978-3-642-33197-8. URL: <http://link.springer.com/10.1007/978-3-642-33197-8> (visitato il 31/03/2023) (cit. a p. 14).
- [26] Ding Li e Xu Yi. «Analysis of the Uncanny Valley Theory and the Effect of Character Realism». In: *Journal of Humanities* () (cit. a p. 14).
- [27] Valentin Schwind, Katrin Wolf e Niels Henze. «Avoiding the uncanny valley in virtual character design». In: *Interactions* 25.5 (22 ago. 2018), pp. 45–49. ISSN: 1072-5520, 1558-3449. DOI: 10.1145/3236673. URL: <https://dl.acm.org/doi/10.1145/3236673> (visitato il 25/03/2023) (cit. a p. 14).
- [28] Darragh Higgins, Donal Egan, Rebecca Fribourg, Benjamin Cowan e Rachel McDonnell. «Ascending from the valley: Can state-of-the-art photorealism avoid the uncanny?». In: *ACM Symposium on Applied Perception 2021*. SAP '21: ACM Symposium on Applied Perception 2021. Virtual Event France: ACM, 16 set. 2021, pp. 1–5. ISBN: 978-1-4503-8663-0. DOI: 10.1145/3474451.3476242. URL: <https://dl.acm.org/doi/10.1145/3474451.3476242> (visitato il 15/03/2023) (cit. a p. 14).
- [29] Nidhi Joshi, Niklas Beecken, Hawa Bah, Frank Steinicke e Juliane Degner. «Advanced Emotion Analytics of Virtual Group Meetings involving Intelligent Virtual Agents». In: *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). Christchurch, New Zealand: IEEE, mar. 2022, pp. 344–350. ISBN: 978-1-66548-402-2. DOI: 10.1109/VRW55335.2022.00077. URL: <https://ieeexplore.ieee.org/document/9757527/> (visitato il 15/03/2023) (cit. a p. 15).
- [30] Mohammed Hazim Alkawaz, Dzulkifli Mohamad, Ahmad Hoirul Basori e Tanzila Saba. «Blend Shape Interpolation and FACS for Realistic Avatar». In: *3D Research* 6.1 (mar. 2015), p. 6. ISSN: 2092-6731. DOI: 10.1007/s13319-015-0038-7. URL: <http://link.spr>

- inger.com/10.1007/s13319-015-0038-7 (visitato il 15/03/2023) (cit. a p. 15).
- [31] Matias Volonte, Eyal Ofek, Ken Jakubzak, Shawn Bruner e Mar Gonzalez-Franco. «HeadBox: A Facial Blendshape Animation Toolkit for the Microsoft Rocketbox Library». In: *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). Christchurch, New Zealand: IEEE, mar. 2022, pp. 39–42. ISBN: 978-1-66548-402-2. DOI: 10.1109/VRW55335.2022.00015. URL: <https://ieeexplore.ieee.org/document/9757409/> (visitato il 31/03/2023) (cit. a p. 15).
- [32] *GitHub - montybot/FACSHuman: FACSHuman plugin for Make-Human project — github.com*. <https://github.com/montybot/FACSHuman>. [Accessed 31-Mar-2023] (cit. a p. 15).
- [33] *www.makehumancommunity.org — makehumancommunity.org*. <http://www.makehumancommunity.org/>. [Accessed 31-Mar-2023] (cit. a p. 15).
- [34] Blender Foundation. *blender.org - Home of the Blender project - Free and Open 3D Creation Software — blender.org*. <https://www.blender.org/>. [Accessed 31-Mar-2023] (cit. a p. 15).
- [35] *VHTOOLKIT / Homepage — vhtoolkit.ict.usc.edu*. <https://vhtoolkit.ict.usc.edu/>. [Accessed 31-Mar-2023] (cit. a p. 15).
- [36] *Unreal Engine 5 — unrealengine.com*. <https://www.unrealengine.com/en-US/unreal-engine-5>. [Accessed 31-Mar-2023] (cit. a p. 16).
- [37] *Metahuman (UE5 plugin)*. Gen. 2023. URL: <https://www.unrealengine.com/en-US/metahuman> (cit. a p. 17).
- [38] *Rig Logic. Runtime facial rig evaluation solver system*. 2022. URL: <https://cdn2.unrealengine.com/rig-logic-whitepaper-v2-5c9f23f7e210.pdf> (cit. alle pp. 21, 22).
- [39] Michael A Sayette, Jeffrey F Cohn, Joan M Wertz, Michael A Perrott e Dominic J Parrott. «A psychometric evaluation of the facial action coding system for assessing spontaneous expression». In: *Journal of nonverbal behavior* 25 (2001), pp. 167–185.

- [40] S. Zhang, C. Zhu, J. K. O. Sin e P. K. T. Mok. «A Novel Ultrathin Elevated Channel Low-temperature Poly-Si TFT». In: 20 (nov. 1999), pp. 569–571.
- [41] Darragh Higgins, Katja Zibrek, Joao Cabral, Donal Egan e Rachel McDonnell. «Sympathy for the digital: Influence of synthetic voice on affinity, social presence and empathy for photorealistic virtual humans». In: *Computers & Graphics* 104 (2022), pp. 116–128.
- [42] Arturo S Garcia, Patricia Fernandez-Sotos, Miguel A Vicente-Querol, Guillermo Lahera, Roberto Rodriguez-Jimenez e Antonio Fernandez-Caballero. «Design of reliable virtual human facial expressions and validation by healthy people». In: *Integrated Computer-Aided Engineering* 27.3 (2020), pp. 287–299.
- [43] Anne-Sophie Milcent, Erik Geslin, Abdelmajid Kadri e Simon Richir. «Expressive Virtual Human: Impact of expressive wrinkles and pupillary size on emotion recognition». In: *Proceedings of the 19th acm international conference on intelligent virtual agents*. 2019, pp. 215–217.
- [44] Emmanuel Ayedoun e Masataka Tokumaru. «Towards Emotionally Expressive Virtual Human Agents to Foster L2 Production: Insights from a Preliminary Woz Experiment». In: *Multimodal Technologies and Interaction* 6.9 (2022), p. 77.
- [45] Michael A Sayette, Jeffrey F Cohn, Joan M Wertz, Michael A Perrott e Dominic J Parrott. «A Psychometric Evaluation of the Facial Action Coding System for Assessing Spontaneous Expression». In: *JOURNAL OF NONVERBAL BEHAVIOR* ().
- [46] Fariba Mostajeran, Nadia Burke, Nazife Ertugrul, Kilian Hildebrandt, Joshua Matov, Noemie Tapie, Wilhelm Gottlieb Zittel, Pia Reisewitz e Frank Steinicke. «Anthropomorphism of Virtual Agents and Human Cognitive Performance in Augmented Reality». In: *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). Christchurch, New Zealand: IEEE, mar. 2022, pp. 329–332. ISBN: 978-1-66548-402-2. DOI: 10 . 1109 / VRW55335 . 2022 . 00074. URL:

- <https://ieeexplore.ieee.org/document/9757379/> (visitato il 15/03/2023).
- [47] Matthew Groh, Craig Ferguson, Robert Lewis e Rosalind Picard. *Computational Empathy Counteracts the Negative Effects of Anger on Creative Problem Solving*. 15 Ago. 2022. arXiv: 2208.07178[cs]. URL: <http://arxiv.org/abs/2208.07178> (visitato il 15/03/2023).
- [48] Aldo Laurentini e Andrea Bottino. «Computer analysis of face beauty: A survey». In: *Computer Vision and Image Understanding* 125 (ago. 2014), pp. 184–199. ISSN: 10773142. DOI: 10.1016/j.cviu.2014.04.006. URL: <https://linkinghub.elsevier.com/retrieve/pii/S1077314214000939> (visitato il 15/03/2023).
- [49] Joanna Kołodziej e Horacio González-Vélez, cur. *High-Performance Modelling and Simulation for Big Data Applications: Selected Results of the COST Action IC1406 cHiPSet*. Vol. 11400. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019. ISBN: 978-3-030-16271-9 978-3-030-16272-6. DOI: 10.1007/978-3-030-16272-6. URL: <http://link.springer.com/10.1007/978-3-030-16272-6> (visitato il 15/03/2023).
- [50] Masaaki Kurosu, cur. *Human-Computer Interaction. Technological Innovation: Thematic Area, HCI 2022, Held as Part of the 24th HCI International Conference, HCII 2022, Virtual Event, June 26 – July 1, 2022, Proceedings, Part II*. Vol. 13303. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2022. ISBN: 978-3-031-05408-2 978-3-031-05409-9. DOI: 10.1007/978-3-031-05409-9. URL: <https://link.springer.com/10.1007/978-3-031-05409-9> (visitato il 15/03/2023).
- [51] Natalie Hube, Kresimir Vidackovic e Michael Sedlmair. «Using Expressive Avatars to Increase Emotion Recognition: A Pilot Study». In: *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. CHI '22: CHI Conference on Human Factors in Computing Systems. New Orleans LA USA: ACM, 27 apr. 2022, pp. 1–7. ISBN: 978-1-4503-9156-6. DOI: 10.1145/3491101.3519822. URL: <https://dl.acm.org/doi/10.1145/3491101.3519822> (visitato il 15/03/2023).

- [52] Katja Zibrek, Benjamin Niay, Anne-Helene Olivier, Julien Pettre, Ludovic Hoyet e Rachel McDonnell. «Proximity in VR: The Importance of Character Attractiveness and Participant Gender». In: *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). Christchurch, New Zealand: IEEE, mar. 2022, pp. 672–673. ISBN: 978-1-66548-402-2. DOI: 10.1109/VRW55335.2022.00187. URL: <https://ieeexplore.ieee.org/document/9757405/> (visitato il 25/03/2023).
- [53] Siddarth Ravichandran, Ondřej Texler, Dimitar Dinev e Hyun Jae Kang. *Synthesizing Photorealistic Virtual Humans Through Cross-modal Disentanglement*. 2 Set. 2022. arXiv: 2209.01320[cs]. URL: <http://arxiv.org/abs/2209.01320> (visitato il 25/03/2023).
- [54] Katja Zibrek, Elena Kokkinara e Rachel Mcdonnell. «The Effect of Realistic Appearance of Virtual Characters in Immersive Environments - Does the Character's Personality Play a Role?». In: *IEEE Transactions on Visualization and Computer Graphics* 24.4 (apr. 2018), pp. 1681–1690. ISSN: 1077-2626, 1941-0506, 2160-9306. DOI: 10.1109/TVCG.2018.2794638. URL: <https://ieeexplore.ieee.org/document/8267290/> (visitato il 25/03/2023).
- [55] Andrea Kleinsmith, Diego Rivera-Gutierrez, Glen Finney, Juan Cendan e Benjamin Lok. «Understanding empathy training with virtual patients». In: *Computers in Human Behavior* 52 (nov. 2015), pp. 151–158. ISSN: 07475632. DOI: 10.1016/j.chb.2015.05.033. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0747563215004045> (visitato il 25/03/2023).
- [56] Mike Seymour, University of Sydney, Australia, Lingyao (Ivy) Yuan, Iowa State University, USA, Alan R. Dennis, Indiana University, USA, Kai Riemer e University of Sydney, Australia. «Have We Crossed the Uncanny Valley? Understanding Affinity, Trustworthiness, and Preference for Realistic Digital Humans in Immersive Environments». In: *Journal of the Association for Information Systems* 22.3 (2021), pp. 591–617. ISSN: 15369323. DOI: 10.17705/1jais.00674. URL:

- <https://aisel.aisnet.org/jais/vol22/iss3/9/> (visitato il 25/03/2023).
- [57] Katja Zibrek e Rachel McDonnell. «Social presence and place illusion are affected by photorealism in embodied VR». In: *Motion, Interaction and Games*. MIG '19: Motion, Interaction and Games. Newcastle upon Tyne United Kingdom: ACM, 28 ott. 2019, pp. 1–7. ISBN: 978-1-4503-6994-7. DOI: 10.1145/3359566.3360064. URL: <https://dl.acm.org/doi/10.1145/3359566.3360064> (visitato il 25/03/2023).
- [58] Katja Zibrek, Sean Martin e Rachel McDonnell. «Is Photorealism Important for Perception of Expressive Virtual Humans in Virtual Reality?» In: *ACM Transactions on Applied Perception* 16.3 (31 lug. 2019), pp. 1–19. ISSN: 1544-3558, 1544-3965. DOI: 10.1145/3349609. URL: <https://dl.acm.org/doi/10.1145/3349609> (visitato il 25/03/2023).
- [59] Faculty of Art, Computing and Creative Industry,Universiti Pendidikan Sultan Idris, 35900 Tanjong Malim, Perak, MALAYSIA, Kogilathah Segaran, Ahmad Zamzuri Mohamad Ali, Faculty of Art, Computing and Creative Industry,Universiti Pendidikan Sultan Idris, 35900Tanjong Malim, Perak, MALAYSIA, Wee Hoe Tan e Faculty of Art, Computing and Creative Industry,Universiti Pendidikan Sultan Idris, 35900Tanjong Malim, Perak, MALAYSIA. «Mediating Effects of Avatar Realism on Motivation inGame Based Learning - Review». In: *Online Journal for TVET Practitioners* 04.2 (30 set. 2019). ISSN: 22897410. DOI: 10.30880/ojtp.2019.04.02.010. URL: <https://publisher.uthm.edu.my/ojs/index.php/ojtp/article/view/5096/3160> (visitato il 25/03/2023).
- [60] David DeVault et al. «SimSensei Kiosk: A Virtual Human Interviewer for Healthcare Decision Support». In: ().
- [61] Kevin el Haddad, Francois Zajega e Thierry Dutoit. «An Open-Source Avatar for Real-Time Human-Agent Interaction Applications». In: *2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*. 2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW). Cambridge, United

- Kingdom: IEEE, set. 2019, pp. 79–80. ISBN: 978-1-72813-891-6. DOI: 10.1109/ACIIW.2019.8925115. URL: <https://ieeexplore.ieee.org/document/8925115/> (visitato il 29/03/2023).
- [62] Catherine Zambaka, Amy Ulinski, Paula Goolkasian e Larry F Hodges. «Social Responses to Virtual Humans: Implications for Future Interface Design». In: *Social Influence* (2007).
- [63] Michael Seymour, Kai Riemer e Judy Kay. «Interactive Realistic Digital Avatars - Revisiting the Uncanny Valley». In: Hawaii International Conference on System Sciences. 2017. DOI: 10.24251/HICSS.2017.067. URL: <http://hdl.handle.net/10125/41216> (visitato il 30/03/2023).