

Politecnico di Torino

Politecnico di Torino - Dipartimento DAUIN

Corso di Laurea Magistrale in Ingegneria del Cinema e dei mezzi di comunicazione

Tesi di Laurea Magistrale in

Realtà Virtuale

Progettazione e sviluppo di ambienti interattivi di realtà virtuale 3D per il monitoraggio e il trattamento dei disturbi da uso di sostanze.

In collaborazione con l'Università degli Studi di Torino
e con il Colloca Lab, University of Maryland, Baltimore (USA)

Candidato: Paolo Arciprete

Relatore:
Prof. Andrea Bottino (Politecnico di Torino)

Co-relatori / Tutor:
Prof.ssa Elisa Carlino (Università degli Studi di Torino)
Prof.ssa Luana Colloca (University of Maryland, Baltimore)
Lola Akintola-Mala (University of Maryland, Baltimore)

Anno Accademico 2025/2026

Sommario

I disturbi da uso di sostanze (Substance Use Disorders, SUD) rappresentano una delle principali sfide per la salute globale. Affrontarne le dimensioni comportamentali, psicologiche e neurobiologiche richiede strumenti innovativi capaci di simulare scenari realistici in ambienti controllati, sicuri e replicabili. Questa tesi presenta la progettazione, lo sviluppo e la sperimentazione di ambienti tridimensionali interattivi in realtà virtuale, realizzati in Unreal Engine 5.7, per riprodurre scenari associati all'uso di sostanze e studiare le reazioni di soggetti con pregressa dipendenza all'esposizione a stimoli (cues) correlati al consumo.

Il percorso immersivo si articola in più ambienti, un ambiente clinico, un'abitazione e un ambiente esterno, nei quali un agente conversazionale MetaHuman fotorealistico, la psicologa virtuale "Nora", accoglie il paziente, conduce un colloquio e raccoglie dati. Il sistema integra, interamente in locale, un modello linguistico di grandi dimensioni per l'interpretazione del linguaggio e il riconoscimento vocale, insieme al lip-sync e alla gestualità del personaggio. La raccolta dati combina misure soggettive del craving tramite scale analogiche visive (VAS) somministrate come questionari diegetici in VR e misure comportamentali oggettive (tracciamento dello sguardo e della manipolazione degli oggetti-stimolo). Il progetto è condotto in collaborazione con il Colloca Lab dell'University of Maryland ed è realizzato presso il Politecnico di Torino.

Parole chiave: *Virtual Reality, Substance Use Disorder, Behavioral Monitoring, 3D Simulation, Interactive Environments.*

Indice

Sommario 2
Indice 3
1 Introduzione 5
1.1 La collaborazione e il disegno del progetto 5
1.2 Motivazione 6
2 Obiettivi 7
3 Background e stato dell'arte 8
3.1 Craving, cue-reactivity e bias attentivo nei disturbi da uso di sostanze 8
3.2 La realtà virtuale per la cue-exposure e la valutazione 8
3.3 La misura del craving: strumenti soggettivi e oggettivi 9
3.4 Agenti conversazionali e umani virtuali (MetaHuman) 10
4 Progettazione degli ambienti virtuali 11
4.1 Pre-produzione: moodboard e ricerca degli asset 11
4.2 L'ambiente clinico, lo studio della psicologa (al chiuso) 11
4.3 L'abitazione (al chiuso) 12
4.4 L'ambiente esterno (all'aperto) 15
4.5 Illuminazione e ottimizzazione per la realtà virtuale 16
4.6 Transizione tra ambienti 17
5 Il MetaHuman: la psicologa virtuale “Nora” 18
5.1 Creazione, abbigliamento e ottimizzazione del MetaHuman 18
5.2 Animazione facciale e lip-sync 18
5.3 Animazioni del corpo: idle e parlato 19
6 Il sistema di dialogo conversazionale 20
6.1 Architettura del sistema 20
6.2 Riconoscimento vocale e classificazione delle risposte 20
6.3 Voce del personaggio 21
7 Immersività, interazione e raccolta dati in VR 22
7.1 Framework VR e locomozione 22
7.2 Schema dei comandi e onboarding 22
7.3 Oggetti afferrabili e interazioni 23
7.3.1 Realismo dei cues e compromesso con il costo GPU 24
7.4 Questionari integrati in VR: la scala VAS 25
7.5 Tracciamento comportamentale: sguardo e manipolazione 26
8 Test di fattibilità e risultati preliminari 28
8.1 Tester e finalità dei test 28

8.2 Procedura 28
8.3 Misure e strumenti 28
8.4 Risultati 29
8.5 Discussione 31
9 Conclusioni e sviluppi futuri 32
Sitografia 34
Riferimenti scientifici 34
Documentazione ufficiale Unreal Engine 34
Plugin e strumenti per l'IA conversazionale 35
Risorse per l'ambiente esterno 35

1 Introduzione

I disturbi da uso di sostanze costituiscono una sfida clinica complessa, che coinvolge milioni di individui e le loro comunità. Uno degli aspetti più critici è il rischio di ricaduta, che rimane elevato anche dopo lunghi periodi di astinenza. In questo fenomeno gioca un ruolo centrale il craving, il desiderio intenso e improvviso di assumere la sostanza, spesso innescato da stimoli ambientali, i cosiddetti cues, che il soggetto ha nel tempo associato al consumo. Poter riprodurre tali stimoli in un contesto controllato e sicuro apre prospettive rilevanti sia per la ricerca sui meccanismi del craving sia per le strategie terapeutiche basate sull'esposizione.

La realtà virtuale (VR) immersiva si è affermata come una piattaforma particolarmente promettente per creare simulazioni realistiche e interattive: consente di collocare il soggetto in ambienti ecologicamente validi mantenendo il pieno controllo sperimentale sugli stimoli presentati, in modo sicuro e replicabile. Rispetto a immagini o video tradizionali, la VR aggiunge il senso di presenza, la sensazione di "essere davvero lì", che rafforza la risposta emotiva agli stimoli. Il presente lavoro di tesi descrive la progettazione, lo sviluppo e la sperimentazione, in Unreal Engine 5.7, di più ambienti VR interattivi, di cui uno che riproduce uno scenario associato all'uso di sostanze, e ne consentono lo studio, integrando tecnologie avanzate per la registrazione di dati comportamentali (eye-tracking, dati di movimento), valutazioni ecologiche momentanee (EMA) e modelli linguistici per interazioni dinamiche e context-aware.6

1.1 La collaborazione e il disegno del progetto

Il progetto si inserisce in una collaborazione interdisciplinare che unisce competenze di tecnologia VR, salute comportamentale e pratica clinica. È realizzato presso il Politecnico di Torino, in collaborazione con l'Università degli Studi di Torino e con il Colloca Lab dell'University of Maryland, Baltimore (USA).

Il Colloca Lab ha fornito le indicazioni cliniche che hanno guidato il disegno sperimentale: quali cues inserire negli ambienti, in quali punti collocarli e con quali modalità presentarli, oltre al tono (mood) con cui caratterizzare l'abitazione. Quest'ultima, in particolare, è stata progettata per non connotarsi né come appartenente a una classe sociale elevata né come marcatamente disagiata, ma in modo che un paziente con storia di dipendenza possa immedesimarsi e percepirla come la propria casa. Questa impostazione, definita insieme al partner clinico, è ciò che conferisce validità ecologica all'esposizione ai cues.

1.2 Motivazione

La scelta di questo tema di tesi nasce da un forte interesse personale per l'aspetto progettuale degli ambienti tridimensionali e, in particolare, per la loro resa immersiva a beneficio di un utente in realtà virtuale. A ciò si aggiunge la curiosità di conoscere e padroneggiare più a fondo il software Unreal Engine sotto molteplici aspetti, dalla modellazione dell'ambiente all'illuminazione, dall'interazione VR all'integrazione di sistemi conversazionali, competenze che si sono rivelate necessarie per rendere l'ambiente credibile e coinvolgente.

2 Obiettivi

In termini generali, l'obiettivo dell'esperienza è comprendere quale sia la risposta del paziente al contatto con determinati cues, osservando in particolare quali tra questi stimoli ne catturino maggiormente l'attenzione e quali invece risultino meno salienti. La possibilità di identificare gli stimoli più attrattivi per il singolo soggetto è di interesse sia per la ricerca sui meccanismi del craving sia per la personalizzazione degli interventi. Condurre questa valutazione in realtà virtuale, con il paziente pienamente immerso nell'ambiente, rende i dati raccolti più affidabili rispetto a stimoli presentati su uno schermo tradizionale: la risposta avviene infatti in un contesto percepito come reale, con un maggiore coinvolgimento emotivo e comportamentale.

A partire da questo scopo, l'obiettivo generale della tesi è progettare, sviluppare e testare ambienti VR interattivi che permettano di esporre soggetti con pregressa dipendenza a stimoli associati al consumo di sostanze e di osservarne e misurarne le reazioni in condizioni controllate e sicure. Si individuano quindi i seguenti obiettivi specifici, in linea con le attività previste dalla proposta di progetto:

- progettare e sviluppare più ambienti immersivi e interattivi che riproducano fedelmente scenari reali associati all'uso di sostanze e al loro trattamento;
- realizzare un agente conversazionale MetaHuman, la psicologa virtuale Nora, in grado di accogliere il paziente, condurre un colloquio, metterlo a proprio agio e raccogliere dati;
- integrare un modello linguistico e il riconoscimento vocale in locale, con lip-sync e gestualità, per un'interazione credibile, adattiva e rispettosa della riservatezza dei dati;
- progettare questionari integrati in VR (scale VAS diegetiche) per la raccolta di dati e la valutazione dell'esperienza dell'utente;
- definire processi per registrare, estrarre e analizzare dati interattivi eterogenei: tracciamento dello sguardo, tracciamento del movimento e manipolazione degli oggetti, e feedback soggettivo;
- garantire la portabilità del sistema su hardware diverso, requisito necessario per l'impiego sperimentale al di fuori della macchina di sviluppo.

3 Background e stato dell'arte

La progettazione di un ambiente VR per lo studio del craving richiede di fondare sia l'impianto teorico sia le scelte di misura sulla letteratura esistente. Questo capitolo inquadra i disturbi da uso di sostanze e il fenomeno del craving, il ruolo della realtà virtuale nell'esposizione ai cues e nella loro valutazione, e gli strumenti con cui misurare la risposta del paziente. Gran parte di questo inquadramento, in particolare la scelta di come valutare il craving, si basa su tre studi indicati dal gruppo clinico dell'University of Maryland come riferimento per la metodologia e per gli aspetti visivi: Huang e collaboratori (2025) sulla cue-exposure in VR per la dipendenza da metanfetamina, Benvegnù e collaboratori (2025) sull'arricchimento ambientale virtuale e il craving da fumo, e Garland e collaboratori (2024) su un intervento VR di mindfulness per la dipendenza da oppioidi.

3.1 Craving, cue-reactivity e bias attentivo nei disturbi da uso di sostanze

Nei disturbi da uso di sostanze il craving e il bias attentivo sono considerati predittori chiave della ricaduta, il che li rende bersagli cruciali per l'intervento (Huang et al., 2025). Il craving può essere distinto in craving tonico, ovvero il desiderio di fondo presente durante l'astinenza, e craving indotto dai cues (cue-induced craving), che emerge in risposta a stimoli associati al consumo. Questi ultimi, la cosiddetta cue-reactivity, sono il risultato di un condizionamento: attraverso ripetute associazioni, oggetti, contesti e situazioni legati alla sostanza acquisiscono la capacità di evocare risposte fisiologiche, emotive e comportamentali anche in assenza della sostanza stessa.

Accanto al craving, il bias attentivo, la tendenza del soggetto ad allocare preferenzialmente l'attenzione verso gli stimoli associati alla droga, è un ulteriore marcatore del processo di dipendenza. Poiché tale processo determina quali cues catturano l'attenzione, la sua misura è strettamente connessa all'obiettivo di questa tesi, che mira proprio a osservare quali stimoli attraggano maggiormente il paziente. Ridurre craving e bias attentivo è, in questa prospettiva, centrale per la prevenzione della ricaduta.

3.2 La realtà virtuale per la cue-exposure e la valutazione

La realtà virtuale è oggi impiegata sia come motore di esposizione, capace di evocare in modo affidabile la cue-reactivity, sia come ambiente di valutazione, grazie alla combinazione di immersione, controllo sperimentale e validità ecologica. Rispetto agli stimoli bidimensionali, la VR consente di ricreare interi contesti, sociali, domestici, clinici, in cui il paziente si muove

e interagisce, ottenendo risposte più vicine a quelle del mondo reale in condizioni comunque sicure e replicabili.

I tre studi di riferimento illustrano bene questa duplice funzione. Huang e collaboratori (2025) hanno condotto uno studio randomizzato controllato su uomini con disturbo da uso di metanfetamina, mostrando che una terapia di cue-exposure basata su VR riduce il craving tonico, migliora l'autoefficacia nel rifiuto della sostanza e diminuisce l'ansia, senza eventi avversi rilevanti. Benvegnù e collaboratori (2025) hanno utilizzato la VR per studiare gli effetti dell'arricchimento ambientale sul craving da sigaretta in fumatori, esponendo i partecipanti a scenari virtuali con cues legati al fumo e misurandone la reattività. Garland e collaboratori (2024) hanno integrato l'esposizione ai cues in VR con pratiche di mindfulness in pazienti con dipendenza da oppioidi, evocando la cue-reactivity tramite l'interazione con droghe e paraphernalia virtuali e osservando risposte fisiologiche di attivazione. Complessivamente, questi lavori confermano che la VR è uno strumento sicuro ed efficace per esporre i pazienti ai cues e per misurarne le reazioni, e che l'immersione rafforza l'affidabilità delle risposte raccolte.

3.3 La misura del craving: strumenti soggettivi e oggettivi

La scelta di come valutare il craving in questo progetto deriva direttamente dalle indicazioni del gruppo clinico e dalla metodologia dei tre studi di riferimento. L'orientamento è verso una combinazione di misure soggettive (self-report all'interno dell'applicazione) e misure oggettive, comprendendo, oltre al craving, anche mood/affetto, ansia, arousal, autoefficacia e bias attentivo, dimensioni sempre più rilevanti in letteratura. All'interno dell'ambiente virtuale, tuttavia, lo strumento privilegiato per il craving è una semplice scala analogica visiva (VAS).

Sul versante soggettivo, la VAS a singolo item si conferma lo standard nei tre studi: Huang e collaboratori (2025) valutano il craving tonico con una VAS da 0 a 10; Benvegnù e collaboratori (2025) impiegano una scala a dieci punti costruita ad hoc, con item del tipo "Quanto desideri fumare in questo momento?"; Garland e collaboratori (2024) somministrano, direttamente all'interno dell'ambiente virtuale, una scala numerica a singolo item da 0 ("nessun craving") a 10 ("il craving più forte mai provato"), affiancata da un questionario più esteso. Oltre al craving, la letteratura raccoglie tipicamente misure di ansia (ad es. GAD-7), umore (ad es. POMS o PANAS), arousal e piacevolezza, autoefficacia nel rifiuto della sostanza e bias attentivo (valutato, ad esempio, con paradigmi Stroop).

Sul versante oggettivo, si distinguono misure fisiologiche e comportamentali. Tra le prime, Garland e collaboratori (2024) monitorano la frequenza cardiaca durante l'esposizione ai cues come indice di reattività autonomica. Tra le seconde, Huang e collaboratori (2025) segnalano i limiti dei paradigmi indiretti come lo Stroop e raccomandano l'adozione di strumenti di misura diretta, come l'eye-tracking, mentre Benvegnù e collaboratori (2025) codificano il comportamento di interazione e di spostamento del partecipante nell'ambiente virtuale come correlato della reattività ai cues.

Coerentemente con questo quadro, il presente lavoro adotta all'interno dell'ambiente una semplice VAS per il craving, somministrata in più momenti del percorso, e la affianca a misure comportamentali oggettive, il tempo di osservazione (sguardo) e il tempo di manipolazione (presa) degli oggetti-stimolo, che operazionalizzano direttamente, in VR, l'attenzione e l'interesse verso i cues, nella direzione della misura diretta raccomandata dalla letteratura. Le misure fisiologiche sono previste come possibile estensione futura, da definire con il partner clinico.

3.4 Agenti conversazionali e umani virtuali (MetaHuman)

Gli agenti conversazionali incarnati (embodied conversational agents) uniscono la comprensione del linguaggio a una rappresentazione visiva antropomorfa. Il framework MetaHuman di Epic Games consente di creare volti umani fotorealistici animabili in tempo reale, con elevata espressività facciale, rendendoli candidati ideali per il ruolo di interlocutore virtuale in un contesto clinico.

Nel processo di creazione dell'agente virtuale (l'NPC) l'obiettivo primario è stato renderlo il più immersivo possibile. A questo fine concorre non solo l'estetica diegetica, il personaggio e il suo contesto appartengono in modo credibile alla scena, ma anche una precisa scelta di interazione: l'assenza di sottotitoli. L'interazione si fonda unicamente su un dialogo parlato in lingua inglese, senza testo a schermo, così che il paziente non sia mai distolto dalla lettura e resti immerso nell'esperienza. La modalità di interazione è inoltre volutamente minimale: al paziente è richiesto solo il push-to-talk, la pressione di un tasto, per attivare la propria risposta vocale. Questa combinazione di estetica diegetica, dialogo esclusivamente parlato e interazione essenziale mira a massimizzare il senso di presenza e la naturalezza del colloquio.

4 Progettazione degli ambienti virtuali

Il percorso sperimentale si articola in più ambienti, progettati per accompagnare il paziente da una condizione di neutralità e sicurezza verso un'esposizione progressiva ai cues, per poi ricondurlo a una condizione di calma. La scelta di Unreal Engine 5.7 è motivata dalla qualità del rendering in tempo reale, dal supporto nativo al framework MetaHuman e dalla disponibilità di strumenti maturi per la realtà virtuale.

4.1 Pre-produzione: moodboard e ricerca degli asset

Per ciascun ambiente la fase di sviluppo è stata preceduta da una pre-produzione dedicata, con la definizione di una moodboard che ne fissasse tono, atmosfera e riferimenti visivi, e da una conseguente ricerca mirata degli asset. Le scelte relative all'atmosfera e ai contenuti sono state concordate con l'University of Maryland, in modo particolare per l'ambiente dell'abitazione, dove la caratterizzazione dello spazio e la collocazione dei cues hanno un impatto diretto sulla validità dell'esposizione.

4.2 L'ambiente clinico, lo studio della psicologa (al chiuso)

Il primo ambiente è lo studio della psicologa, concepito come spazio di accoglienza e neutralità. Si tratta di un ambiente luminoso, dotato di una finestra che si affaccia sulla città, così da trasmettere un senso di serenità e di calma: è qui che il paziente dialoga con la psicologa Nora, che avvia la conversazione, spiega l'esperienza e conduce il colloquio raccogliendo dati. In questo ambiente non sono presenti cues stimolanti: esso funge da baseline neutra e da introduzione all'interazione in VR. Al termine del colloquio viene somministrata la prima misura del craving (VAS1).



Figura 4.1, Lo studio della psicologa, l'ambiente di accoglienza iniziale: uno spazio luminoso e sobrio in cui il paziente viene messo a proprio agio prima dell'esposizione ai cues.

4.3 L'abitazione (al chiuso)

Il secondo ambiente è un'abitazione, nella quale il paziente esplora liberamente lo spazio. A differenza dello studio, l'atmosfera è deliberatamente più buia e stressante, con l'obiettivo di immedesimare l'utente in una situazione in cui si trova solo. È in questo scenario che avviene l'esposizione principale ai cues: alcuni oggetti-stimolo afferrabili sono distribuiti nell'ambiente e il sistema traccia, per ciascuno, il tempo di osservazione e il tempo di manipolazione.

La caratterizzazione dell'abitazione, la tipologia di cues e il tono complessivo dell'ambiente, è stata definita seguendo le indicazioni del Colloca Lab. L'abitazione è volutamente "media": non lussuosa né degradata, così da massimizzare la possibilità di immedesimazione del paziente, che dovrebbe percepirla come la propria casa. In corso di sviluppo l'ambiente è stato rivisto, passando da un set di asset più cupi e datati a uno più moderno, per rendere lo spazio più credibile e attuale.

Per accrescere il realismo e il senso di uno spazio vissuto, la resa dei materiali è stata arricchita con decals e con asset fotorealistici Megascan; per contenere il costo di questi ultimi in VR, alle relative texture ad alta risoluzione è stato applicato un LOD bias, riducendone il peso senza comprometterne in modo percepibile la qualità. Per rendere l'ambiente più disordinato e dunque più realistico, coerentemente con l'atmosfera decisa in pre-produzione, sono stati scaricati capi d'abbigliamento da Sketchfab, la cui caduta o la cui disposizione appesa all'attaccapanni è stata simulata in Marvelous Designer, ottenendo pieghe e drappeggi naturali.

I quadri appesi alle pareti sono stati realizzati dall'autore, combinando il modello della cornice con quello del vetro, sporcato con dei decals, e scegliendo di volta in volta poster di film o riproduzioni di dipinti famosi. Più in generale, gli asset e i modelli dell'abitazione sono stati selezionati cercando di mantenere una palette di colori coerente; nei casi in cui non è stato possibile reperire modelli adatti a questo vincolo cromatico, alcune texture sono state realizzate direttamente dall'autore su Substance Painter, come nel caso dell'armadio.

L'abitazione è un appartamento compatto, articolato in poche stanze, tra cui un soggiorno con letto e una cucina, così da non risultare dispersivo e da mantenere l'attenzione del paziente concentrata sui cues e sugli altri oggetti afferrabili. A differenza dello studio, la finestra non si affaccia su una città: all'esterno si intravede soltanto una serie di lampioni nel buio della strada, che lasciano filtrare un po' di luce calda dando contrasto all'ambiente. Dopo la fase di esplorazione libera è prevista una seconda misura del craving (VAS2), seguita da una fase di esplorazione guidata da una voce fuori campo e da una terza misura (VAS3) che conclude l'esperienza.



Figura 4.2, Soggiorno dell'abitazione: sul tavolino sono collocati i cues (oggetti-stimolo afferrabili); alle pareti poster e attaccapanni con capi d'abbigliamento simulati in Marvelous Designer.



Figura 4.3, La cucina dell'abitazione.



Figura 4.4, La camera da letto dell'abitazione: si notano l'armadio (la cui anta è resa apribile), i capi d'abbigliamento a terra e la luce calda che filtra dalla finestra.

4.4 L'ambiente esterno (all'aperto)

Il terzo ambiente è uno spazio all'aperto con funzione di rilassamento, che porta il paziente in uno scenario simile a una casa nella foresta, completamente estraneo a quello della casa buia in città. Si tratta del cortile antistante una baita immersa in una foresta di cui non si scorge la fine, circondata da alberi. In questo ambiente il paziente può muoversi soltanto nel cortile esterno: non rappresenta infatti una zona utile all'esperienza, ma solo uno spazio di stazionamento momentaneo, e si può rientrare nell'abitazione attraverso la maniglia della stessa porta.



Figura 4.5, L'ambiente esterno: il cortile davanti alla baita nella foresta, con funzione di spazio sicuro e rilassante.

Sul piano realizzativo, la vegetazione è stata gestita con particolare attenzione al costo in VR: gli alberi più vicini alla baita hanno una qualità maggiore, mentre quelli più lontani scendono progressivamente di dettaglio man mano che ci si allontana dalla casa. Anche in questo caso erano inizialmente stati impiegati alberi Megascan, la cui qualità è però stata drasticamente ridotta; è stata inoltre rimossa la funzione di wind dalle mesh e dai materiali, poiché troppo onerosa. L'ambiente è stato costruito modellando il landscape e utilizzando lo strumento di landscape painting per delineare la strada sterrata davanti alla casa; un grafo PCG è stato impiegato per distribuire ulteriori dettagli naturali, pietre, foglie, erba e simili, con mesh a basso numero di poligoni. L'illuminazione è stata anch'essa precalcolata tramite bake, con uno Sky Light abbinato a una Directional Light statica e a un Volumetric Cloud.

Questo ambiente riprende, in forma alleggerita, un progetto realizzato in origine per un corso: da quest'ultimo sono stati rimossi numerosi elementi, mantenendo soltanto ciò che serve alla funzione di spazio di decompressione. Se ne descrivono qui le scelte realizzative principali.

Il terreno è stato scolpito manualmente con gli strumenti Sculpt, Smooth e Flatten per definire collinette e dislivelli, dopo aver creato un apposito layer di modellazione dedicato. Al landscape è stato poi applicato un materiale costruito con il nodo Landscape Layer Blend, che consente di dipingere sul terreno texture differenti: in Paint Mode è stata così tracciata la strada sterrata che conduce alla baita e che funge da linea guida visiva verso di essa.

La foresta è stata popolata in modo procedurale con un PCG Graph: sfruttando il nodo Landscape Data, uno Static Mesh Spawner distribuisce sul terreno diverse tipologie di alberi, organizzate in sette indici distinti, a ciascuno dei quali sono assegnati un peso e una densità specifici per ottenere una flora il più possibile eterogenea. Il posizionamento manuale, tramite Foliage, è stato riservato ai soli dettagli narrativi, come i rampicanti sulla baita e le rocce lungo il sentiero, attivando per essi l'opzione Align to Normal affinché la vegetazione aderisca in modo realistico alle pendenze del tetto e del legno.

Gli alberi provengono dalla libreria Quixel Megascans (nello specifico gli European Beech), scelti per l'elevata qualità fotorealistica ma, come già ricordato, ridotti di dettaglio e privati della funzione di wind per contenerne il costo in VR. I massi principali, importati da Blender, sono stati raccolti in un Blueprint dedicato a cui sono state associate diverse decals di muschio e foglie cadute per aggiungere micro-dettagli realistici, mentre l'asset della baita è stato reperito online.

4.5 Illuminazione e ottimizzazione per la realtà virtuale

Il mantenimento di un framerate elevato e stabile è un requisito critico in VR, sia per il comfort del paziente sia per evitare fenomeni di malessere da simulazione. Per questo motivo Nanite e Lumen, che in questo progetto risultavano troppo onerosi per l'esecuzione in realtà virtuale, sono stati disattivati. In loro luogo si è scelto di precalcolare l'illuminazione tramite bake: gli oggetti sono stati resi statici e le loro lightmap sono state calibrate una ad una, oggetto per oggetto. La risoluzione delle lightmap varia da 256 a 1024 a seconda delle dimensioni dell'oggetto e del fatto che esso sia o meno afferrabile dall'utente, così da concentrare la qualità dove serve. Per il calcolo dell'illuminazione indiretta è stata inoltre inserita nell'ambiente una Sphere Reflection Capture (nell'abitazione con una brightness di 0,3).

La caratterizzazione più scura dell'abitazione ha richiesto scelte specifiche sui corpi illuminanti per evitare ombre incoerenti. Per le luci a soffitto è stato adottato un tipo di lampada che non proietta ombra verso l'alto (Figura 4.6), così da non generare artefatti sul soffitto; le lampade a parete che puntano verso il basso sono state realizzate con Rect Light, mentre per le plafoniere

a soffitto e per la lampada da tavolo sono state usate Point Light. Nella reception iniziale, infine, le lampade appese a soffitto sono state circondate da una Rect Light per ciascun lato, in modo da simularne la luce in maniera coerente: una singola Rect Light o Point Light avrebbe infatti prodotto ombre incoerenti.



Figura 4.6, Esempio di illuminazione nell'abitazione: la plafoniera a soffitto è stata scelta per non proiettare ombre verso l'alto.

4.6 Transizione tra ambienti

Il passaggio dallo studio della psicologa all'abitazione avviene tramite un elemento diegetico, una porta, che, all'interazione con la maniglia, innesca una dissolvenza e il caricamento del livello successivo. È importante notare che questa transizione si attiva soltanto dopo che il dialogo con la psicologa è terminato, così da garantire che il colloquio e la prima misura siano completati prima di procedere. La soluzione mantiene la continuità dell'esperienza ed evita rotture dell'immersione nel passaggio da un ambiente all'altro.

5 Il MetaHuman: la psicologa virtuale "Nora"

Il personaggio virtuale che accoglie il paziente e conduce il colloquio è un MetaHuman denominato "Nora", nel ruolo di psicologa. La sua funzione è duplice: sul piano clinico, accogliere il paziente, metterlo a proprio agio e condurre il colloquio raccogliendo i dati; sul piano tecnico, fungere da interfaccia conversazionale multimodale che unisce voce, espressione facciale e gestualità. Come descritto nel capitolo precedente, l'intero personaggio è stato pensato per massimizzare l'immersione, con dialogo esclusivamente parlato e senza sottotitoli.

5.1 Creazione, abbigliamento e ottimizzazione del MetaHuman

La scelta estetica del volto del MetaHuman non risponde a un criterio particolare. L'abbigliamento, invece, ha richiesto una lavorazione dedicata: il modello base del vestito è stato reperito online (da un sito che forniva la sola mesh) e adattato al personaggio. Il capo è stato dapprima importato in Marvelous Designer per farlo aderire correttamente al corpo del MetaHuman e per rifilarne le parti eccessivamente lunghe; è stato quindi riportato in Unreal Engine sotto forma di skeletal mesh, in modo da poter essere applicato come vestito al blueprint di Nora e da ereditare i pesi delle ossa dell'armatura del personaggio, così da muoversi in sincronia con esso. Un procedimento analogo è stato seguito per le scarpe, assegnate alle ossa dei piedi del MetaHuman, ma senza il passaggio intermedio in Marvelous Designer.

Per contenere il carico sulla GPU è stato infine forzato un livello di dettaglio ridotto sul personaggio: un LOD 3 sul volto e un LOD 2 sul corpo.

5.2 Animazione facciale e lip-sync

La sincronizzazione labiale è realizzata tramite un insieme di 41 sequenze di animazione (AnimSequence) dei visemi, che agiscono sulle curve di controllo delle espressioni facciali (CTRL_expressions_*) a 30 fotogrammi al secondo. All'interno dell'Animation Blueprint del personaggio, un nodo di tipo Slot ("DefaultSlot") è inserito nella catena ModifyCurve → Slot → RigLogic, in modo che l'animazione dei visemi si sovrapponga correttamente alla posa di base gestita da RigLogic. All'atto della riproduzione di una battuta, il Blueprint del dialogo carica dinamicamente la sequenza di lip-sync corrispondente al file vocale e la riproduce come montage dinamico sullo slot dedicato, ottenendo un lip-sync allineato all'audio.

5.3 Animazioni del corpo: idle e parlato

La gestualità del corpo è affidata a quattro animazioni scaricate da Mixamo e retargetate sullo scheletro del MetaHuman: una posa di idle, riprodotta quando il personaggio non parla, e tre animazioni di parlato che si alternano durante il discorso. La logica di selezione risiede nell'Animation Blueprint del corpo e distingue in primo luogo se il personaggio stia parlando o meno, attraverso una variabile booleana (IsTalking); quando il personaggio parla, un indice (TalkIndex) seleziona di volta in volta una delle tre animazioni di parlato, facendole ruotare in modo variabile a seconda di ciò che Nora sta dicendo.

Un aspetto di rilievo è il modo in cui viene determinato lo stato di parlato: anziché essere impostato manualmente, esso è ricavato in tempo reale dall'analisi dell'audio riprodotto. L'Animation Blueprint avvia l'analisi dell'uscita audio e ne estrae l'ampiezza in corrispondenza di una frequenza di riferimento (1000 Hz); il valore, opportunamente lisciato nel tempo, attiva lo stato IsTalking quando supera una soglia (0,03). In questo modo il corpo del personaggio si anima come se stesse effettivamente parlando ogni volta che una battuta viene riprodotta, in sincronia con la voce e, di conseguenza, con il lip-sync.

Poiché le animazioni provengono da Mixamo e sono state retargetate, la posa di base è stata inoltre rifinita direttamente nell'Animation Blueprint tramite una catena di correzioni ossee: nodi di modifica additiva della rotazione applicati alla colonna (dallo spine_01 allo spine_05), alle clavicole, alle braccia, alle cosce e ai polpacci, insieme a due controlli di cinematica inversa (Two Bone IK) sui piedi. Questi accorgimenti correggono le imperfezioni introdotte dal retargeting e ancorano correttamente i piedi, restituendo una posa naturale e stabile.

Un ultimo intervento ha riguardato le gambe del personaggio: l'Animation Blueprint è stato modificato in modo che restino quasi completamente ferme. Nel passaggio da un'animazione all'altra, infatti, capitava spesso che le gambe di Nora "fluttuassero" sul pavimento, spostandosi in modo innaturale; bloccandole si ottiene invece il comportamento tipico di una persona che parla da ferma, senza movimenti innaturali degli arti inferiori.

6 Il sistema di dialogo conversazionale

Il cuore dell'interazione è un sistema di dialogo che integra, interamente in locale, un modello linguistico di grandi dimensioni e il riconoscimento vocale. La scelta di eseguire i modelli sul dispositivo, anziché su servizi cloud, risponde a tre esigenze: tutelare la riservatezza dei dati sensibili raccolti da soggetti con storia di dipendenza, rendere le sessioni indipendenti dalla connessione di rete e, non da ultimo, contenere i costi. Le soluzioni commerciali reperite in rete che offrivano questo tipo di servizio erano infatti a pagamento, a partire da quella che, sul piano qualitativo, sarebbe stata la scelta ideale: Convai.

6.1 Architettura del sistema

La logica del dialogo è organizzata su due livelli distinti. Un componente in C++ (UReceptionistDialogueComponent, il "cervello" del dialogo) contiene il contenuto e il flusso del colloquio: carica la definizione del dialogo, implementa la macchina a stati, costruisce i prompt per il classificatore, effettua il matching delle risposte, cattura le variabili ed esporta la sessione. Il Blueprint (BP_ReceptionistDialogue) si occupa esclusivamente del cablaggio e della presentazione: carica i modelli, riproduce l'audio delle battute, innesca il lip-sync, mostra il pannello VAS e gestisce microfono e puntatore. Questa separazione tiene distinta la logica del colloquio, che risiede nel codice, dalla sua presentazione, gestita a livello visuale.

Il contenuto del colloquio è definito in modo dichiarativo in un file JSON (Dialogue_Receptionist.json), strutturato come un grafo di nodi con ramificazioni: ogni battuta di Nora è associata a un file vocale e a un insieme di possibili risposte, ciascuna con un identificatore, un significato e le variabili da impostare. Questa impostazione consente di modificare domande, ordine e ramificazioni senza ricompilare il codice. Al termine, le variabili raccolte, la trascrizione grezza e i timestamp sono esportati in un file di sessione (session_*.json).

6.2 Riconoscimento vocale e classificazione delle risposte

Il paziente risponde parlando liberamente. L'interazione avviene in modalità push-to-talk: tenendo premuto un tasto del controller si attiva la cattura del microfono, che viene disattivata al rilascio. Il parlato è trascritto in tempo reale da un modello Whisper (whisper.cpp) eseguito in locale. La trascrizione è quindi interpretata da un modello linguistico locale (qwen2.5-1.5b, tramite il plugin Llama-Unreal), il cui compito non è generare le frasi di Nora, bensì classificare la risposta libera del paziente associandola a una delle opzioni previste dal nodo corrente del dialogo, oppure segnalarla come non chiara. La classificazione è resa deterministica

(temperatura nulla, sequenze di stop e reset del contesto a ogni domanda), così da garantire coerenza e ripetibilità. In base all'esito, la macchina a stati determina la battuta successiva.

È importante precisare che Nora non reagisce alle azioni compiute dal paziente sugli oggetti, ma unicamente alle sue frasi, interpretate dal modello. Se il paziente pronuncia una risposta che il classificatore non riesce ad associare ad alcuna opzione del nodo corrente, il dialogo prosegue comunque con la battuta successiva. Il colloquio si arresta e ripete l'ultima battuta soltanto in un caso: quando, durante la pressione del tasto di push-to-talk, non viene rilevato alcun parlato. Questo controllo, gestito nell'Event Graph, evita che il dialogo avanzi se il paziente preme il tasto senza dire nulla.

6.3 Voce del personaggio

Poiché le battute di Nora sono predefinite, la voce non è sintetizzata a runtime ma pre-registrata, così da evitarne la latenza e garantire una qualità costante. Le clip vocali sono generate offline con una voce neurale (edge-tts, voce en-US-AvaNeural, con un ritmo lievemente rallentato per un tono calmo e clinico) e convertite nel formato richiesto dal motore. L'interazione è realizzata in lingua inglese (en-US) per il migliore supporto dei modelli linguistici e di sintesi vocale rispetto all'italiano, oltre che per la scelta, descritta nel capitolo 3, di un dialogo esclusivamente parlato e privo di sottotitoli.

7 Immersività, interazione e raccolta dati in VR

L'efficacia di un ambiente di cue-exposure dipende in misura significativa dal grado di immersione e dalla naturalezza dell'interazione. Questo capitolo descrive i meccanismi che consentono al paziente di muoversi e interagire, i questionari integrati in VR e il sistema che registra il comportamento durante l'esposizione.

7.1 Framework VR e locomozione

Il pawn del giocatore è basato sull'XR Framework di Unreal e dispone dei controller destro e sinistro con cui il paziente interagisce con l'ambiente e con i widget tridimensionali. Durante lo sviluppo è emerso un problema per cui, componendo lo spostamento con la direzione della camera, la componente verticale poteva far "salire" il giocatore sugli arredi; la soluzione adottata vincola il movimento al piano orizzontale, garantendo che il paziente rimanga all'altezza del pavimento in un ambiente a superficie piana. Per favorire una navigazione più libera, inoltre, la collisione di alcuni piccoli elementi presenti a terra, sia nel primo sia nel secondo ambiente, è stata volutamente ignorata, così che non ostacolino il movimento del paziente.

7.2 Schema dei comandi e onboarding

Per rendere l'esplorazione semplice e immediata, l'interazione è stata ridotta a soli tre comandi: gli analogici (thumbstick) per spostarsi nell'ambiente e per ruotare la visuale, il tasto A per il push-to-talk con cui parlare, e il tasto laterale di grab del controller per afferrare gli oggetti. Lo stesso tasto di grab è impiegato sia per prendere gli oggetti sia per interagire con la scala VAS: questa scelta riduce il numero di pulsanti e rende più intuitiva l'esperienza, lasciando che il paziente compia tutto il resto con i soli analogici e i due tasti.

Per guidare il paziente, i comandi sono segnalati nel momento in cui servono tramite modelli tridimensionali dei controller che fluttuano nello spazio e riproducono in loop un'animazione indicante il tasto da premere; a ciascuno è sovrapposto un testo bidimensionale (un ulteriore widget) che spiega la funzione del comando. All'avvio, ad esempio, alcuni suggerimenti mostrano come spostarsi e come ruotare la visuale (Figura 7.1).

All'avvio dell'esperienza i due controller-tutorial restano visibili finché il paziente non utilizza entrambi i comandi analogici, quello di spostamento del corpo e quello di rotazione della visuale; solo dopo che li ha impiegati i suggerimenti scompaiono. A quel punto, avvicinandosi

alla psicologa Nora, e unicamente quando le è sufficientemente vicino, il dialogo si avvia e prosegue fino alla conclusione.



Figura 7.1, Lo studio della psicologa con i suggerimenti di onboarding: i comandi (spostamento e rotazione della visuale) sono indicati da testi e da modelli dei controller.

7.3 Oggetti afferrabili e interazioni

Nell'abitazione tutti i cues sono resi afferrabili e, insieme ad essi, anche numerosi altri oggetti non-stimolo. La ragione è metodologica: lasciando il paziente libero di esplorare e manipolare qualunque cosa, è possibile osservare in modo più sincero quanto egli sia effettivamente interessato ai cues, e non perché indotto a interagire soltanto con quelli. Va inoltre notato che, anche senza afferrare o aprire alcunché, l'ambiente contiene di per sé un numero di oggetti sufficiente a essere osservato nel tempo di permanenza previsto; le interazioni servono soprattutto a invogliare l'esplorazione.

A rafforzare questo comportamento, al termine di VAS2 una voce fuori campo invita esplicitamente il paziente a esplorare la casa e a interagire con gli oggetti: anche chi, nella prima fase, non lo avesse fatto spontaneamente viene così incoraggiato a prenderli e a spostarli.



Figura 7.2, Un tavolo dell'abitazione con diversi cues afferrabili (siringa, pillole, cucchiaio, bustina, flacone) accanto a oggetti di uso quotidiano.

Alcuni oggetti offrono interazioni più articolate del semplice afferrare. L'anta dell'armadio, ad esempio, è stata resa apribile: nel suo blueprint le mesh sono state separate e collegate da un physics constraint, così che l'utente possa aprirla afferrandola. Il flacone di pillole, invece, può essere svitato: il sistema misura l'angolo di rotazione impresso dalla mano e ne accumula il valore complessivo; quando questo supera una soglia prestabilita, il tappo viene considerato svitato, si stacca dal flacone, passa alla simulazione fisica e attiva l'apertura del contenitore. Meccaniche di questo tipo accrescono la naturalezza e il coinvolgimento dell'interazione.

La realizzazione del tappo svitabile ha richiesto anche una modifica al sistema di presa. Inizialmente, tentando di afferrare il tappo veniva afferrata l'intera bottiglia, che passava così da una mano all'altra senza che l'utente riuscisse a toccare e ruotare il tappo. Per risolvere il problema è stato modificato il blueprint del pawn (BP_XRPawn), che controlla il personaggio-utente e i suoi comandi, incluso quello di grab: alla pressione del tasto il sistema individua ora il componente afferrabile più vicino alla mano e afferra quello, consentendo di prendere il tappo indipendentemente dalla bottiglia.

7.3.1 Realismo dei cues e compromesso con il costo GPU

La realizzazione dei cues ha richiesto un continuo compromesso tra realismo e costo computazionale: renderli credibili senza gravare eccessivamente sugli fps, e quindi sulla GPU, non è stato semplice, soprattutto per gli oggetti dotati di parti riflettenti o trasparenti. I tre cues più onerosi in termini di carico sono stati la bustina di pillole, la bustina di erba e la siringa, tutti caratterizzati da componenti trasparenti, notoriamente costose da renderizzare per via della gestione della traslucenza.

Le due bustine sono state modellate appositamente. I loro contenuti, le pillole e l'erba, sono stati reperiti come singoli modelli da Sketchfab, mentre l'involucro è stato costruito dall'autore. Nel caso dell'erba, il modello è stato portato in Marvelous Designer per simularvi attorno un involucro di plastica in modo volutamente irregolare; il risultato è stato quindi riportato in Blender per l'assegnazione dei materiali corretti e infine reimportato in Unreal per verificarne la resa. Per le pillole, invece, il numero di elementi è stato moltiplicato (da uno a circa duecento) e la loro caduta su un piano è stata simulata in Blender; il gruppo così ottenuto è stato esportato come blocco unico verso Marvelous Designer, dove un lembo di stoffa, a cui sono stati assegnati parametri fisici simili a quelli della plastica, è stato simulato sopra il mucchio, con una leggera pressione, così da avvolgerlo in una bustina più ampia e ordinata. La differenza tra i due procedimenti riflette anche la diversa natura degli involucri: disordinato per l'erba, più regolare per le pillole.

In molti casi è stato inoltre necessario creare in Unreal un materiale di plastica ad hoc, traslucido ma "sporco": l'aggiunta di piccole imperfezioni e di uno strato di sporcizia, ottenuto anche tramite decals, rende più percepibile la presenza dell'involucro di plastica attorno all'oggetto, evitando l'aspetto eccessivamente pulito che una trasparenza perfetta avrebbe conferito.

Un discorso a parte merita il flacone di pillole apribile. Poiché l'oggetto è cavo e deve poter essere aperto, la collisione generata automaticamente da Unreal, una forma semplice che avvolge l'intera mesh, non era adeguata, e si è dovuta definire una collisione personalizzata. In Blender sono state quindi modellate alcune piccole mesh convesse che, nel loro insieme, approssimano la forma del flacone. Seguendo la convenzione di denominazione di Unreal, tali mesh ausiliarie sono state nominate con il prefisso UCX_ (seguito dal nome della mesh di rendering): all'importazione, il motore le riconosce automaticamente come collisione personalizzata dell'oggetto, le impiega come collider e non le disegna a schermo. In questo modo si ottiene una collisione fedele anche per oggetti concavi o complessi, impossibile da rappresentare con i volumi di collisione primitivi standard.

7.4 Questionari integrati in VR: la scala VAS

La misura soggettiva del craving è raccolta tramite una scala analogica visiva (VAS), presentata come questionario diegetico: il widget (WBP_VAS), accompagnato da alcune mesh, tra cui quella di una clipboard su cui è montato, compare come una sorta di modulo clinico da compilare, coerente con l'ambiente dello studio (Figura 7.3). L'interfaccia è stata resa il più

semplice possibile, con pulsanti grandi e due sole azioni previste: lo spostamento dello slider (su una scala da 0 a 10) e la pressione del pulsante di conferma in basso. Entrambe le azioni sono attivate tramite il tasto di grab del controller, lo stesso usato per afferrare gli oggetti, segnalato al paziente nel momento in cui la VAS compare per la prima volta. Per rendere l'operazione ancora più comprensibile, come emerso dai test descritti nel capitolo 8, è la stessa psicologa Nora a spiegare a voce come compilare la scala, offrendo un aiuto che non interrompe l'immersione. Alla conferma, il valore è registrato e il dialogo prosegue.



Figura 7.3, La scala VAS diegetica, presentata come un modulo su clipboard all'interno dello studio della psicologa.

La scala VAS è somministrata in tre momenti del percorso: al termine del colloquio iniziale (VAS1, craving baseline), dopo l'esplorazione libera dell'abitazione (VAS2) e al termine dei compiti guidati (VAS3). Il confronto tra le tre misure permette di osservare l'andamento del craving in relazione all'esposizione ai cues. Questa scelta è coerente con la letteratura di riferimento, in cui una VAS o una scala a singolo item da 0 a 10 è lo strumento privilegiato per la misura del craving all'interno dell'ambiente virtuale.

Va infine precisato che nell'abitazione non è presente alcun MetaHuman e il paziente non dialoga con nessuno: l'unico elemento sonoro di guida è la voce fuori campo (extradiegetica) che accompagna la comparsa di VAS2 e VAS3. Diversamente dallo studio della psicologa, quindi, in questo ambiente non vi è interazione conversazionale.

7.5 Tracciamento comportamentale: sguardo e manipolazione

Nell'abitazione, un sistema di tracciamento registra per ciascun oggetto-stimolo due grandezze principali: il tempo di sguardo (dwell), ovvero quanto a lungo e quante volte il paziente osserva

l'oggetto, e il tempo di manipolazione (hold), ovvero quanto a lungo e quante volte lo afferra. Il tracciamento dello sguardo è realizzato tramite un trace sferico che parte dalla posizione e dalla direzione del visore (che in VR seguono i movimenti della testa), mentre la manipolazione è rilevata dagli eventi di presa e rilascio dei componenti afferrabili. Questi indicatori comportamentali forniscono una misura oggettiva dell'attenzione e dell'interesse verso i cues, complementare alla misura soggettiva del craving raccolta con la VAS, e permettono di individuare quali stimoli catturino maggiormente il paziente.

Al termine dell'esplorazione i dati raccolti sono esportati in un file di testo leggibile, all'interno di una cartella scrivibile e portabile del progetto, con una riga per ciascun oggetto che riporta tempo e numero di osservazioni e di prese. La portabilità del percorso di salvataggio è un requisito esplicito, poiché il sistema deve funzionare anche su altri PC e nella build pacchettizzata.

Esempio di riga prodotta dal sistema di tracciamento:

```
Oggetto: Bottle_of_pills | Osservato: 3.6 s (3 volte) | In mano: 5.1 s (2 prese).
```

8 Test di fattibilità e risultati preliminari

8.1 Tester e finalità dei test

La sperimentazione condotta finora non ha coinvolto la popolazione target, bensì un insieme di tester non in target, cioè persone senza una storia di dipendenza da sostanze. Lo scopo di questi test non è quindi valutare l'efficacia clinica del sistema, ma verificarne la validità pratica: l'assenza di problemi di motion sickness, la linearità e la comprensibilità del percorso e il corretto funzionamento delle interazioni. Complessivamente sono state svolte circa dieci sessioni di test informali: le prime prove sono state condotte dall'autore stesso e, successivamente, da un collega in veste di tester, mentre le altre hanno coinvolto ulteriori volontari non in target. Da questo insieme di prove sono emerse diverse osservazioni utili al miglioramento del sistema, di cui le principali sono riportate nel seguito.

La validazione con la popolazione target, con i relativi criteri di inclusione ed esclusione, il consenso informato e l'approvazione del comitato etico, costituisce il passo successivo e necessario, non ancora affrontato in questa fase.

8.2 Procedura

La sessione sperimentale segue il percorso attraverso gli ambienti. Indossato il visore, il paziente viene accolto nello studio della psicologa Nora, che avvia il dialogo e conduce il colloquio; al termine è richiesta la prima valutazione del craving (VAS1). Solo a colloquio concluso, interagendo con la maniglia della porta, il paziente accede all'abitazione, dove esplora liberamente lo spazio ed è esposto ai cues mentre il sistema traccia sguardo e manipolazione degli oggetti; segue la seconda valutazione del craving (VAS2). Viene poi proposta una fase di compiti guidati, al termine della quale è raccolta la terza misura (VAS3). Il percorso si conclude nell'ambiente esterno, il cortile della baita nella foresta, che riporta il paziente a una condizione di calma.

8.3 Misure e strumenti

Le misure raccolte comprendono: la valutazione soggettiva del craving tramite scala VAS nei tre momenti del percorso; il tempo e il numero di osservazioni (dwell/gaze) per ciascun oggetto-stimolo; il tempo e il numero di manipolazioni (hold/grab) per ciascun oggetto. A queste si aggiungono la trascrizione e le variabili del colloquio, esportate nel file di sessione.

8.4 Risultati

I test pilota condotti finora hanno restituito indicazioni incoraggianti sulla fattibilità del sistema. Non si sono manifestati problemi rilevanti di motion sickness durante le sessioni, e il percorso attraverso i tre ambienti è risultato sufficientemente lineare e comprensibile, senza che i tester perdessero l'orientamento o non sapessero come procedere. Le interazioni fondamentali, ovvero lo spostamento nello spazio, la presa degli oggetti e la compilazione della scala VAS, si sono rivelate utilizzabili senza particolari difficoltà anche da persone con poca o nessuna esperienza pregressa di realtà virtuale.

Il sistema di tracciamento comportamentale e l'esportazione dei dati di sessione hanno funzionato come previsto, producendo per ciascun oggetto-stimolo i tempi e il numero di osservazioni e di prese in un file leggibile e portabile. Trattandosi di prove con tester non in target, i valori raccolti (le misure VAS e gli indicatori comportamentali) non hanno rilevanza clinica e non sono quindi riportati come risultati sperimentali, ma sono serviti a verificare che l'intera catena di raccolta e salvataggio dei dati fosse corretta e affidabile.

Un problema significativo è invece emerso a livello del dialogo con la psicologa Nora. Nella logica iniziale il sistema tentava di interpretare la risposta verbale del paziente prima di far avanzare il colloquio: quando il modello non comprendeva o non classificava correttamente la frase pronunciata, il colloquio si bloccava e Nora ripeteva l'ultima domanda, generando ripetizioni innaturali e frustranti. La causa è riconducibile alla natura del modello linguistico impiegato, volutamente leggero per poter essere eseguito in locale in tempo reale, che non garantisce una comprensione semantica affidabile del parlato in ogni condizione.

Per risolvere il problema si è modificata la logica di avanzamento del dialogo: il colloquio prosegue ora alla ricezione di una qualsiasi risposta verbale percepita dal sistema, senza subordinare l'avanzamento alla corretta interpretazione del suo contenuto. Questa scelta è resa possibile dal fatto che, durante la sessione, un operatore umano è presente e trascrive le risposte fornite dal paziente, per cui la comprensione semantica in tempo reale da parte del modello non è un requisito necessario ai fini della raccolta dei dati. La modifica ha reso il colloquio fluido e privo delle ripetizioni riscontrate in precedenza.

Un secondo gruppo di problemi ha riguardato la capsula di collisione del Pawn, ovvero il volume che rappresenta l'utente all'interno del gioco. Nei primi due ambienti, lo studio e l'abitazione, non è mai previsto un dislivello, uno scalino o un punto in cui il paziente debba salire o scendere, e l'utente si muove sempre sullo stesso piano. Nel passaggio al terzo

ambiente, invece, all'uscita dell'abitazione è presente un gradino che immette nel cortile, e in questa situazione il fatto che l'utente non scendesse effettivamente il gradino risultava poco verosimile.

La soluzione ha dovuto tenere conto anche di un effetto indesiderato collaterale. Abilitando la gravità nei primi due ambienti si rendeva necessario sbloccare anche la rotazione della capsula, e in alcuni casi questa ruotava o si inclinava su assi diversi da quello verticale, generando un movimento innaturale del punto di vista e un rischio concreto di motion sickness. Si è quindi scelto di mantenere l'utente vincolato a un unico piano nei primi due ambienti, senza gravità e senza possibilità di salire o scendere, bloccando la rotazione della capsula su tutti gli assi a eccezione di quello verticale (Z), attorno al quale l'utente deve comunque potersi girare. Solo nel momento in cui, attraverso la porta, si accede al terzo ambiente, la gravità viene sbloccata, così che l'utente possa salire e scendere il gradino in modo realistico.

Un ulteriore problema, anch'esso legato alla capsula del Pawn, riguardava gli oggetti afferrati. Quando l'utente prendeva un oggetto, lo avvicinava a sé e provava a spostarsi, si verificava una collisione tra il collider dell'oggetto e quello della capsula dell'utente, con comportamenti anomali del movimento. Una prima soluzione, far scomparire il collider dell'oggetto durante la presa, si è rivelata inadeguata, poiché al momento del rilascio l'oggetto riacquisiva il collider in modo instabile, comportandosi in maniera imprevedibile. La soluzione definitiva è stata impostare il collider di ciascun oggetto afferrabile in modo che ignori esclusivamente il collider della capsula del Pawn, mantenendo per il resto le proprie collisioni. In questo modo l'oggetto e l'utente non possono mai entrare in conflitto, senza che sia necessaria una collisione fisica tra gli oggetti impugnati e il volume dell'utente, di fatto superflua.

Un'ultima indicazione emersa dai test riguarda la chiarezza delle interazioni. Alcuni tester hanno trovato poco intuitivo l'uso del tasto di grab per muovere lo slider della scala VAS e per azionare la maniglia della porta. Per rendere più comprensibili queste azioni senza compromettere l'immersione, si è scelto di sfruttare la voce della psicologa Nora: poiché sta già dialogando con il paziente, è lei stessa a spiegare a voce come compilare il pannello del craving, offrendo un aiuto aggiuntivo che resta coerente con la finzione dell'esperienza. Inoltre, il suggerimento tridimensionale del comando di grab non compare soltanto accanto alla scala VAS nel momento in cui questa appare, ma anche accanto alla maniglia della porta, così da indicare al paziente dove essa si trova e come utilizzarla. Fanno eccezione i primi due comandi, lo spostamento e la rotazione della visuale tramite gli analogici, e il push-to-talk, che il paziente

deve individuare autonomamente: i relativi controller-tutorial fluttuanti restano infatti visibili finché non vengono effettivamente utilizzati, segnalandone così la presenza e la funzione.

8.5 Discussione

Nel loro complesso, i test pilota indicano che il sistema è praticabile e usabile: gli ambienti si esplorano senza disagio percettivo, il percorso è chiaro e le interazioni sono alla portata anche di utenti inesperti. Questo risultato, pur non riguardando l'efficacia clinica, è il presupposto necessario per poter condurre in futuro uno studio con la popolazione target, poiché un sistema che provocasse motion sickness o che risultasse di difficile utilizzo comprometterebbe qualsiasi misurazione successiva.

La revisione della logica di avanzamento del dialogo merita una riflessione. La soluzione adottata, far proseguire il colloquio a ogni risposta verbale percepita, è un compromesso consapevole tra le prestazioni ottenibili con un modello linguistico locale e leggero e la fluidità dell'esperienza. Nel contesto d'uso previsto tale compromesso è accettabile e anzi opportuno, perché la trascrizione affidabile delle risposte è affidata all'operatore umano presente durante la sessione, mentre l'agente conversazionale ha soprattutto la funzione di rendere il colloquio naturale e di mettere a proprio agio il paziente.

I limiti di questa fase sono espliciti. Il campione è ridotto e composto da tester non in target, per cui non è possibile trarre alcuna conclusione sull'efficacia del sistema nel suscitare o misurare il craving, né sulla validità psicometrica delle misure raccolte. I risultati qui riportati hanno pertanto natura esclusivamente tecnica e di usabilità, e vanno intesi come propedeutici a una successiva validazione con la popolazione clinica di riferimento.

9 Conclusioni e sviluppi futuri

Questa tesi ha presentato la progettazione, lo sviluppo e la sperimentazione di ambienti VR interattivi per l'esposizione controllata a cues legati alle sostanze, integrando un agente conversazionale MetaHuman fotorealistico, la psicologa virtuale Nora, con modelli linguistici e di riconoscimento vocale eseguiti in locale, questionari diegetici per la misura del craving e un sistema di tracciamento comportamentale. Il lavoro ha affrontato e risolto una serie di sfide tecniche non banali, dalla sincronizzazione labiale all'interazione affidabile con l'interfaccia in VR, dalla locomozione alla portabilità del progetto, costruendo, in collaborazione con il Colloca Lab, una piattaforma funzionante e riutilizzabile per la ricerca sul craving.

È importante precisare la portata di questi risultati. La sperimentazione condotta ha coinvolto un numero ridotto di tester non appartenenti alla popolazione target, e non consente pertanto di trarre conclusioni sull'efficacia clinica del sistema né sulla sua capacità di suscitare e misurare il craving. Il contributo principale di questo lavoro è quindi di natura tecnica e metodologica: la realizzazione di una piattaforma immersiva completa, portabile e funzionante, che integra ambienti realistici, un agente conversazionale eseguito interamente in locale, strumenti di misura soggettiva del craving e un sistema di tracciamento comportamentale. La verifica di fattibilità e di usabilità qui presentata costituisce la base necessaria su cui potrà innestarsi la successiva validazione clinica.

Il primo e più importante sviluppo futuro è dunque la validazione del sistema con la popolazione target, ovvero con utenti con una storia di dipendenza da sostanze, nel rispetto dei necessari criteri di inclusione ed esclusione, del consenso informato e dell'approvazione del comitato etico. Solo uno studio di questo tipo, condotto su un campione adeguato, potrà chiarire se e in che misura gli ambienti sviluppati siano in grado di elicitare risposte di craving misurabili e di discriminare l'effetto dell'esposizione ai cues.

Un secondo, promettente sviluppo riguarda l'integrazione di misure fisiologiche a corredo di quelle soggettive e comportamentali già implementate. La letteratura sul craving indica infatti come particolarmente informative due famiglie di segnali: l'attività cardiaca, in particolare la frequenza cardiaca e la sua variabilità (heart rate variability, HRV), e l'attività elettrodermica, cioè la conduttanza cutanea legata alla sudorazione. La frequenza cardiaca è già stata impiegata come indicatore di attivazione all'interno di protocolli di realtà virtuale per i disturbi da uso di sostanze (Garland et al., 2024), la variabilità della frequenza cardiaca è stata associata in modo consistente ai processi di autoregolazione e al craving nei soggetti con disturbo da uso di

sostanze (Moon et al., 2024), e l'impiego combinato di più segnali fisiologici durante compiti di esposizione ai cues si è dimostrato utile a caratterizzarne la reattività (Kroczek et al., 2024). L'aggiunta di sensori indossabili per la frequenza cardiaca, la HRV e la conduttanza cutanea permetterebbe di affiancare alle misure già presenti una componente oggettiva e continua dell'attivazione fisiologica, rafforzando la validità complessiva della misura del craving.

Un ulteriore sviluppo, suggerito dal gruppo dell'University of Maryland, riguarda la personalizzazione dell'ambiente di esposizione. L'idea è disporre i cues all'interno del secondo ambiente, l'abitazione, in modo differente a seconda del profilo del paziente, ad esempio in base al tipo di sostanza utilizzata e alle relative modalità di consumo, collocando in posizione più prominente gli oggetti-stimolo pertinenti alla sua storia. Una simile adattività permetterebbe di calibrare l'esposizione sul singolo paziente e di osservare in modo più mirato da che cosa dipenda il suo craving, incrociando ciò che egli osserva e manipola con maggiore interesse durante l'esperienza con quanto dichiarato nel colloquio iniziale, così da verificare se le due fonti di informazione convergano o divergano.

Ulteriori direzioni di sviluppo comprendono l'ampliamento del ventaglio di ambienti e di cues, così da coprire contesti d'uso e sostanze diverse, il miglioramento dell'agente conversazionale con modelli linguistici locali più capaci, e l'affinamento del protocollo sperimentale sulla base dei dati raccolti nello studio con la popolazione target.

Sitografia

Il progetto non prevede una bibliografia tradizionale, bensì una sitografia. Di seguito sono raccolti i riferimenti scientifici, la documentazione tecnica ufficiale e le principali risorse online utilizzate nello sviluppo.

Riferimenti scientifici

Benvegnù, G., Perotti, S., Vegher, A., & Chiamulera, C. (2025). Virtual Reality Environmental Enrichment Effects on Craving for Cigarettes in Smokers. *Games for Health Journal*, 14(1), 21–28. https://doi.org/10.1089/g4h.2023.0188

Garland, E. L., Recasens, M., Taple, B. J., Donaldson, G. W., & Weisberg, R. B. (2024). Virtual reality-based Mindfulness-Oriented Recovery Enhancement (MORE-VR) as an adjunct to medications for opioid use disorder: a Phase 1 trial. *Annals of Medicine*, 56(1), 2392870. https://doi.org/10.1080/07853890.2024.2392870

Huang, Q., Chen, X., Wang, X., Tang, Y., Wang, W., Chao, L., Nie, Y., Peng, S., Zhao, L., Shen, H., & Liao, Z. (2025). Virtual reality-based cue exposure therapy reduces psychological craving in men with methamphetamine use disorder: a randomized controlled trial. *Translational Psychiatry*, 15, 318. https://doi.org/10.1038/s41398-025-03553-7

Kroczek, A. M., Häußinger, F. B., Rohe, T., Schneider, S., Plewnia, C., Batra, A., Fallgatter, A. J., & Ehlis, A. C. (2024). Multimodal assessment of smoking cue reactivity during a smoking cue exposure task. *Clinical EEG and Neuroscience*, 55(3), 287–295. https://doi.org/10.1177/15500594221138273

Moon, S. J. E., Schlenk, E. A., & Lee, H. (2024). Heart rate variability in adults with substance use disorder: A comprehensive narrative review. *Journal of the American Psychiatric Nurses Association*, 30(2), 240–251. https://doi.org/10.1177/10783903221145142

Documentazione ufficiale Unreal Engine

Documentazione Unreal Engine (hub): https://dev.epicgames.com/documentation/unreal-engine

MetaHuman: https://dev.epicgames.com/documentation/en-us/metahuman

Illuminazione statica e lightmap: https://dev.epicgames.com/documentation/en-us/unreal-engine/understanding-lightmapping-in-unreal-engine

Landscape: https://dev.epicgames.com/documentation/en-us/unreal-engine/landscape-outdoor-terrain-in-unreal-engine

Procedural Content Generation (PCG): https://dev.epicgames.com/documentation/en-us/unreal-engine/procedural-content-generation-framework-in-unreal-engine

Collisioni delle Static Mesh e mesh di collisione personalizzate (UCX): https://dev.epicgames.com/documentation/en-us/unreal-engine/setting-up-collisions-with-static-meshes-in-unreal-engine

Physics Constraint Component: https://dev.epicgames.com/documentation/en-us/unreal-engine/physics-constraint-component-user-guide-in-unreal-engine

Enhanced Input: https://dev.epicgames.com/documentation/en-us/unreal-engine/enhanced-input-in-unreal-engine

Scripting di base con i Blueprint: https://dev.epicgames.com/documentation/en-us/unreal-engine/basic-scripting-with-blueprints-in-unreal-engine

Animation Blueprint, nodi di blend (Blend Poses by bool): https://dev.epicgames.com/documentation/en-us/unreal-engine/animation-blueprint-blend-nodes-in-unreal-engine

Plugin e strumenti per l'IA conversazionale

Llama-Unreal, plugin di integrazione di llama.cpp in Unreal Engine (getnamo, GitHub): https://github.com/getnamo/Llama-Unreal

whisper.cpp, implementazione in C/C++ del modello Whisper per il riconoscimento vocale (ggml-org, GitHub): https://github.com/ggml-org/whisper.cpp

edge-tts, sintesi vocale tramite il servizio Text-to-Speech di Microsoft Edge (rany2, GitHub): https://github.com/rany2/edge-tts

MetaHuman AI Lip Sync con Text-to-Speech locale in UE 5.6+ (Epic Developer Community): https://dev.epicgames.com/community/learning/tutorials/72ma/unreal-engine-metahuman-ai-lip-sync-with-local-text-to-speech-in-ue-5-6

Blend Pose by Bool in Unreal Engine 5 (Epic Developer Community): https://dev.epicgames.com/community/learning/tutorials/l754/blend-pose-by-bool-unreal-engine-5

Tutorial video seguito per l'animazione e la sincronizzazione labiale del MetaHuman (YouTube): https://www.youtube.com/watch?v=dAaIvyG2tE8

Risorse per l'ambiente esterno

Asset della baita (CGTrader): https://www.cgtrader.com/free-3d-models/architectural/architectural-street/house-in-forest-39cdb168-7c30-4742-9283-708d9727abed

Alberi della foresta (Fab): https://www.fab.com/listings/d11cc01d-9422-41b7-950f-416c9ce79caf

Costruire una foresta con il PCG (YouTube): https://www.youtube.com/watch?v=cD7Q-tu1DgU

Creare una foresta in Unreal (Reddit): https://www.reddit.com/r/unrealengine/comments/1bgav7n/

Nodo Landscape Data nel PCG (YouTube): https://www.youtube.com/watch?v=RiEUQQXa4xA

Dipingere il landscape con texture diverse (YouTube): https://www.youtube.com/watch?v=_XLe8sZwdj8

How to make a forest (YouTube): https://youtu.be/8dWflkhqc9k