

Dall'èkphrasis all'esperienza VR immersiva ne *Il Barone rampante* di Italo Calvino

Sofia Menconero
Chiara Florise Amadei
Giorgio Gosti
Bruno Fanini

Abstract

L'èkphrasis, tradizionalmente definita come descrizione verbale dettagliata di opere visive, è reinterpretata in questo studio attraverso un approccio multidisciplinare che integra letteratura, Intelligenza Artificiale generativa (AI) e Realtà Virtuale (VR). La ricerca esplora la traduzione intermediale di un testo tratto da *Il Barone rampante* di Italo Calvino, autore noto per il suo stile visivo e descrittivo. Attraverso l'estrazione di parole chiave dal testo sono state generate immagini panoramiche 360° con AI generativa. Queste immagini, insieme alla narrazione audio del brano, sono state integrate in un'esperienza VR fruibile con *Head-mounted display* (visore HMD).

Lo studio evidenzia come l'immersività della VR amplifichi la relazione tra immagine e narrazione, passando da una descrizione letteraria a una multisensoriale. L'esperienza è stata testata su un campione esplorativo, il cui comportamento visivo è stato tracciato con strumenti avanzati. I dati raccolti hanno permesso di analizzare la direzione dello sguardo e la correlazione tra elementi visivi e sonori, mentre un questionario sul senso di presenza ha fornito riscontri sull'efficacia immersiva dell'esperienza. Nonostante i limiti di un campione ristretto, i risultati preliminari dimostrano il potenziale delle tecnologie emergenti nella traduzione intermediale. La ricerca propone nuove prospettive per la fruizione letteraria, suggerendo sviluppi futuri che includano campioni più ampi e scene diversificate, esplorando ulteriormente il ruolo dell'èkphrasis in contesti immersivi contemporanei.

Parole chiave

Studi visuali, AI generativa, VR immersiva, Italo Calvino, comportamento utente.



Immagine evocativa di
sintesi dell'esperienza di
traduzione intermediale
(immagine degli autori).

Introduzione

L'èkphrasis, originariamente definita come una descrizione verbale dettagliata di un'opera d'arte o di una scena visiva, è un dispositivo retorico che connette il mondo delle parole con quello delle immagini. Negli ultimi decenni, l'èkphrasis ha assunto nuovi significati e applicazioni, grazie alla crescente intersezione tra letteratura, arte visiva e tecnologia digitale [Shiel 2013; Lindhé 2013; Jansson 2018].

Il presente studio esplora il concetto di èkphrasis su vari livelli e con un approccio multidisciplinare [1], focalizzandosi particolarmente sull'interazione tra testo letterario, generazione di immagini tramite Intelligenza Artificiale (AI) e Realtà Virtuale (VR) immersiva. Il metodo adottato si basa sulla selezione di una porzione di testo a carattere descrittivo dell'ambiente dove si svolge la narrazione e sull'estrazione di alcuni elementi chiave da utilizzare come *prompt* per la generazione di un'immagine panoramica a 360° mediante tecnologie di AI generativa. L'immagine risultante, insieme alla lettura audio della corrispondente parte del romanzo, è stata poi integrata in un'esperienza di VR proposta ad alcuni utenti, che hanno potuto immergersi in prima persona nell'ambiente virtuale attraverso un visore. Durante queste sessioni, il comportamento degli utenti è stato tracciato, consentendo di raccogliere dati sulla direzione di osservazione e ricavare una descrizione iconica dell'esperienza, chiudendo il cerchio e riportando il focus sull'èkphrasis, stavolta non più come ponte tra verbale e visuale ma tra visuale e visuale.

L'opera letteraria selezionata per la sperimentazione del metodo è *Il barone rampante* di Italo Calvino [Calvino 2024, pp. 93-338]. La scelta è ricaduta su questo autore per via della sua attitudine alla descrizione dettagliata dei contesti in cui opera la narrazione [Ricci 2001], tanto da venire talvolta associato al dispositivo dell'èkphrasis [Fucarino 2015] e stimolare la traduzione grafica delle sue ambientazioni [Cianci, Calisi 2014; Barenghi 2023]. Calvino è uno scrittore "visivo" [Belpoliti 2006] e proprio alla visibilità ha dedicato una delle sei *Lezioni americane* [Calvino 2020, pp. 81-100].

Il presente contributo si propone di analizzare l'intero processo, esplorando le sfide e le opportunità offerte dalla traduzione intermediale e dalla VR immersiva. In particolare, si illustrerà come l'èkphrasis possa evolversi in un contesto tecnologico contemporaneo, trasformandosi da fenomeno puramente letterario in un'esperienza multisensoriale capace di amplificare la comprensione e l'immersione nell'opera originale.

Dall'èkphrasis alla generazione AI dell'esperienza VR

Il barone rampante racconta una parte della peculiare e fantasiosa esistenza di Cosimo Piovasco di Rondò, narrata dal fratello Biagio. Siamo nella seconda metà del XVIII secolo, in un paese lungo la costa ligure. In particolare, la vicenda prende avvio a partire da un evento significativo nella vita del giovane protagonista, ovvero il momento in cui, dopo un diverbio col padre, Cosimo decide di vivere il resto della sua vita tra gli alberi, senza più poggiare i piedi in terra.

La parte di romanzo presa in considerazione ai fini dello studio si colloca tra la fine del primo capitolo e l'inizio del secondo [Calvino 2024, pp. 106-107]. È il momento in cui Cosimo per la prima volta sale su un albero e da lassù osserva il mondo che lo circonda (fig. 1).

Per tradurre questo testo letterario in un'esperienza multisensoriale immersiva e interattiva si è ricorso alla sfera visiva e uditiva. Da un lato c'era la necessità di ottenere un contesto grafico ispirato all'ambientazione del testo, all'altro una voce narrante e un paesaggio sonoro.

La scena è stata resa attraverso un panorama a 360° generato da un apposito servizio di AI *text-to-image* [2]. Dal testo originale di Calvino sono state selezionate una serie di parole chiave necessarie per rendere il paesaggio e l'atmosfera descritti nel brano. Queste parole chiave sono state poi ricucite in un semplice testo in linguaggio naturale che fungesse da *prompt* (fig. 2).

Cosimo salì fino alla forcella d'un grosso ramo dove poteva stare comodo, e si sedette lì, a gambe penzoloni, a braccia incrociate con le mani sotto le ascelle, la testa insaccata nelle spalle, il tricorno calcato sulla fronte.

Nostro padre si sporse dal davanzale: – Quando sarai stanco di star lì cambierai idea!

– gli gridò.

– Non cambierò mai idea, – fece mio fratello, dal ramo.

– Ti farò vedere io, appena scendi!

– E io non scenderò più! – E mantenne la parola.

Cosimo era sull'elce. I rami si sbracciavano, alti ponti sopra la terra. Tirava un lieve vento; c'era il sole. Il sole era tra le foglie, e noi per vedere Cosimo dovevamo farci schermo con la mano. Cosimo guardava il mondo dall'albero: ogni cosa, vista di lassù, era diversa, e questo era già un divertimento. Il viale aveva tutt'un'altra prospettiva, e le airole, le ortensie, le camelie, il tavolino di ferro per prendere il caffè in giardino. Più in là le chiome degli alberi si sfittivano e l'ortaglia digradava in piccoli campi a scala, sostenuti da muri di pietre; il dosso era scuro di oliveti, e, dietro, l'abitato d'Ombrosa sporgeva i suoi tetti di mattone sbiadito e ardesia, e ne spuntavano pennoni di bastimenti, là dove sotto c'era il porto. In fondo si stendeva il mare, alto d'orizzonte, ed un lento veliero vi passava.

Fig. 1. Brano de *Il Barone rampante* di Italo Calvino considerato ai fini della sperimentazione [Calvino 2024, pp. 106-107].

In the garden of an Italian villa on a midday in the 1700s. From the PoV of a high branch of an oak tree, one sees a garden with trees, pathways, flowerbeds, hydrangeas, camellias, and a wrought-iron coffee table. In the distance, the treetops reveal fields supported by stone walls and olive groves. On the other side, an Italian village with faded brick and slate rooftops is visible. Beyond the village lies the port, and in the background, the sea with a sailing ship. Just one person sitting on an oak branch.

Fig. 2. Prompt utilizzato per la generazione del panorama a 360° (immagine degli autori).

Oltre alla descrizione di Calvino, sono state aggiunte alcune informazioni (il contesto cronologico e geografico, il momento della giornata) non inserite nel brano di testo selezionato ma presenti nelle immediate vicinanze, per permettere all'AI di centrare meglio la generazione. Per garantire un punto di vista più alto rispetto al suolo, e quindi simulare il tipo di visione del protagonista, è stato utilizzato un modello di AI che simula la vista da drone. Il risultato ottenuto, nei tempi brevi che contraddistinguono le generazioni AI, è stato piuttosto buono, pur in mancanza di alcuni elementi topici del racconto (fig. 3). Tali elementi sono stati aggiunti successivamente tramite l'*editing* dell'immagine, che da equirettangolare è stata trasformata in proiezione cubica per poterci lavorare senza deformazioni. Sempre attraverso strumenti generativi [3] sono stati così aggiunti: Cosimo sulla forcella dell'albero con il tricorno in testa, alcune aiuole fiorite, il tavolino da caffè e il veliero in lontananza. Infine, è stata eseguita la trasformazione inversa da proiezione cubica a immagine equirettangolare (fig. 4).

Per quanto riguarda la componente audio dell'esperienza, si è fatto ricorso a un servizio AI *text-to-speech* [4]. Questo tipo di tecnologia permette di ottenere risultati professionali e una vasta gamma di scelta in termini di lingua, genere, età e stile. Per la lettura del brano tratto da *Il Barone rampante*, si è optato per una giovane voce maschile modulata, in lingua italiana, appartenente alla categoria narrativa, con l'obiettivo di simulare il racconto attraverso la voce di Biagio. Come sottofondo della lettura sono stati montati suoni del mondo naturale: il rumore del mare in lontananza e il cinguettio degli uccelli [5].



Fig. 3. Immagine equirettangolare originale ottenuta attraverso la generazione AI (immagine degli autori).



Fig. 4. Immagine equirettangolare dopo l'editing con l'aggiunta di alcuni elementi salienti (immagine degli autori).

L'esperienza VR è stata costruita grazie al framework *open source* ATON [6] [Fanini et al. 2021] che permette l'implementazione di scene 3D o 360° (arricchite da contenuti multimediali come video e audio), la loro fruizione via web attraverso molteplici dispositivi (desktop, smartphone, visori ecc.) anche in *extended reality* (WebXR), e la facile condivisione *online*. Attraverso un link persistente, gli utenti possono accedere liberamente all'esperienza e fruirla con qualsiasi dispositivo [7] (fig. 5). Data la natura immersiva e interattiva dell'esperienza, composta da immagini e suoni, è chiaramente preferibile una fruizione tramite visore HMD.

La descrizione del comportamento utente nelle sessioni VR

La distinzione tra 'vedere' e 'guardare' è cruciale per comprendere l'esperienza dell'*èkphrasis*. 'Vedere' implica una percezione visiva passiva, uno stimolo che raggiunge l'occhio e viene elaborato dal cervello senza necessariamente attivare un livello profondo di consapevolezza. Al contrario, 'guardare' è un atto intenzionale, che richiede attenzione e partecipazione attiva da parte dell'osservatore. In un contesto immersivo come quello della VR, questa distinzione si amplifica: gli utenti non si limitano a 'vedere' l'ambiente creato, ma sono chiamati a 'guardare' con attenzione e curiosità, esplorando lo spazio, soffermandosi sui dettagli e ricostruendo mentalmente il significato delle immagini in relazione al testo letterario, abbinando il carattere narrativo dell'immagine con il racconto vocale.



Fig. 5. Alcune viste dell'esperienza (immagini degli autori).

Ai fini del presente studio, si è reputato interessante soffermarsi anche sull'esperienza utente. Si tratta, da un lato, di un ulteriore sviluppo del concetto di *èkphrasis*, ovvero quello rivolto alla descrizione del comportamento utente per via iconica, tracciando la direzione del visore dei fruitori – e di conseguenza, con una buona approssimazione, dello sguardo – e visualizzando tali dati. Dall'altro lato, un questionario orientato a misurare il senso di 'presenza' in ambienti virtuali, da compilare al termine dell'esperienza, ha fornito interessanti riscontri. Un prototipo di questa sperimentazione ha visto la partecipazione di 16 individui [8]: un campione non abbastanza numeroso e disomogeneo ai fini di un'attendibile analisi dei

dati, per la quale si aspetterà un numero maggiore e variegato di utenti, ma sicuramente interessante per una prima analisi esplorativa e per testare il metodo.

Il tracciamento e la visualizzazione dei dati tracciati sono stati possibili grazie a Interlumo [9]: una *pipeline* di servizi open-source rivolta a tutti coloro che intendono studiare e comprendere grandi quantità di interazioni prodotte da applicazioni remote, installazioni pubbliche o di laboratorio, spazi virtuali o fisici, web app online e altro [Fanini, Gosti 2024].

Interlumo è suddiviso in tre macro-componenti: cattura, elaborazione e ispezione. Il primo componente (*Kapto*) permette di catturare le interazioni prodotte da applicazioni remote sottoforma di dati numerici [10]. Il secondo componente (*Procezo*) permette di processare e aggregare i dati (record) per analizzare i pattern degli utenti tramite Analytics. Include operazioni di filtraggio, pulizia, generazione di dati statistici e applicazione di modelli di *Machine Learning* (ML). Per esempio, si possono stimare le aree più visitate, i punti più osservati e altri comportamenti. Il terzo componente (*Merkhet*) permette di ispezionare tali dati in forma visuale/immersiva, facilitandone una lettura avanzata e l'analisi dei comportamenti direttamente attraverso un comune web browser. Fornisce, inoltre, strumenti avanzati per annotare sessioni e dati statistici, interfacce spaziali per modalità immersive e varie funzioni collaborative.

La visualizzazione dei dati tracciati con Interlumo si basa sempre sul *framework* ATON e, nel caso di sessioni con panorami 360°, è possibile esaminare due informazioni diverse: la traiettoria di osservazione dell'utente e la *heatmap* associata ai maggiori punti di fissazione [11] (fig. 6). Nel primo caso, una volta caricato il record è anche possibile ancorare

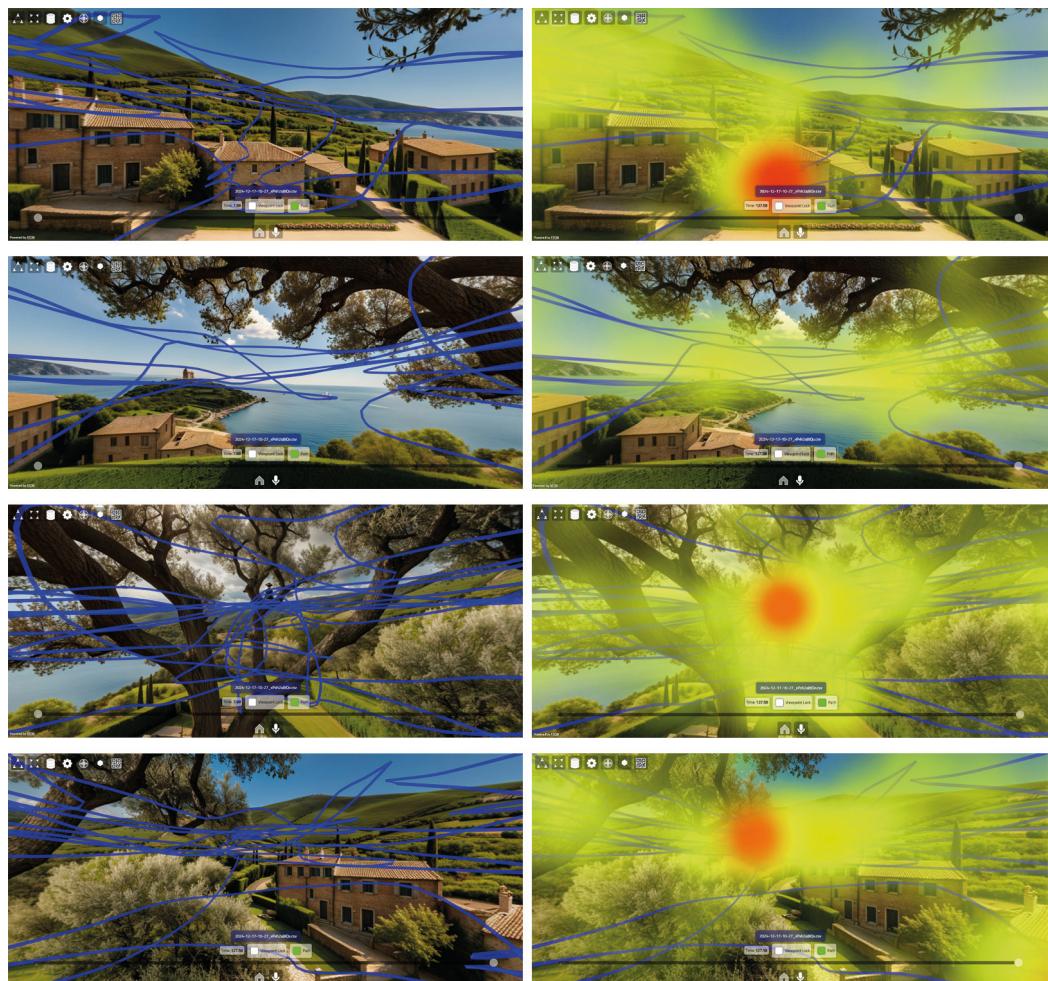


Fig. 6. A sinistra, il tracciato in blu con la direzione di osservazione di un utente; a destra, la heatmap con i punti di fissazione dello stesso utente (immagini degli autori).

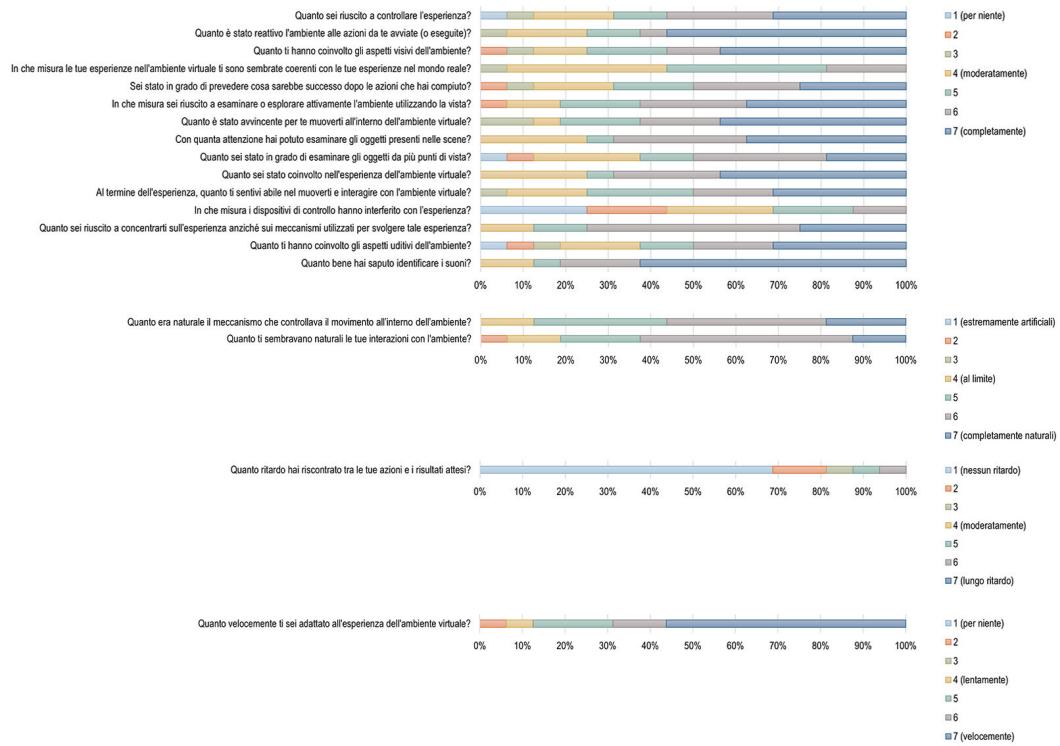


Fig. 7. Esiti del questionario sul senso di 'presenza' compilato dai partecipanti alla sperimentazione (immagini degli autori).

il punto di vista della scena a quello registrato e osservare lo spostamento della direzione visiva lungo la traiettoria muovendo uno *slider* lungo una *timeline*. Nel secondo caso, il computo delle aree di maggiore fissazione può essere eseguito anche per più record contemporaneamente. Queste due informazioni sono servite a capire quali punti della scena sono stati osservati, per quanto tempo, e la relazione tra la componente visiva e sonora dell'esperienza. Alcuni esiti parziali [12] di queste osservazioni dimostrano due aree particolarmente osservate: il gruppo di edifici e Cosimo sull'albero. Mentre nel primo caso la motivazione dipende dal fatto che l'esperienza prende avvio dall'inquadratura verso gli edifici, e dunque si riscontra un soffermarsi maggiore su questa posizione, nel secondo caso è la presenza dell'unico elemento umano all'interno della figurazione a catturare l'attenzione degli osservatori.

Al termine della fruizione, gli utenti sono stati chiamati a rispondere a un sondaggio sul senso di 'presenza' avvertito, ovvero su uno dei tre assi intimamente correlati che caratterizzano le esperienze immersive iconiche: 'scorniciamento', 'presenzialità' e 'immediatezza' [Pinotti 2024, p. 276-281]. La 'presenza' è definita come "l'esperienza soggettiva di trovarsi in un luogo o ambiente, anche quando ci si trova fisicamente in un altro" [traduzione da Witmer, Singer 1998, p. 225] e per misurarla è stato utilizzato un apposito questionario in scala Likert, lievemente modificato per adattarlo all'esperienza su *Il Barone rampante* [13]. Seppure, come anticipato, i risultati non possono essere considerati plausibili per via dell'omogeneità e dimensione del campione, per completezza dello studio si riportano le domande poste e le risposte in forma aggregata (fig. 7).

Conclusioni

"All'origine di ogni storia che ho scritto c'è un'immagine che mi gira per la testa, nata chissà come e che mi porto dietro magari per anni. A poco a poco mi viene da sviluppare questa immagine in una storia con un principio e una fine, e nello stesso tempo – ma i due processi sono spesso paralleli e indipendenti – mi convinco che essa racchiude qualche significato" [Calvino 2024, p. 461].

Il presente studio si ispira alla forte affinità di Calvino con le immagini, proponendo un metodo di traduzione intermediale della scena di una sua celebre opera. L'integrazione tra letteratura, AI e VR ha permesso di esplorare nuove modalità di traduzione intermediale, aprendo interessanti prospettive sulla fruizione e sull'interpretazione delle opere letterarie. Il metodo proposto si basa sulla generazione di un panorama visivo e sonoro attraverso strumenti di AI generativa. È stato preferito l'utilizzo di un'immagine 360° piuttosto che di un modello 3D immaginando che il *workflow* possa essere applicato in modo massivo su più parti di opere letterarie, rendendo la modellazione 3D non compatibile in termine di tempi-stistiche [14]. L'utilizzo di un'immagine 360° monoscopica invece che di un modello 3D, se da un lato riduce l'esplorazione a 3 gradi di libertà (*Degree of Freedom, DoF*) consentendo solo una navigazione stazionaria, dall'altro lato non sembra rappresentare un limite alla fruizione da parte degli utenti. Questo dato è stato registrato proprio grazie alla sperimentazione su *Il Barone rampante*, i cui fruitori hanno dichiarato nel questionario sul senso di 'presenza' un ottimo livello di coinvolgimento rispetto agli aspetti visivi, uditivi e, in generale, di tutta l'esperienza.

I risultati preliminari, sebbene limitati a un campione ristretto e omogeneo, offrono spunti interessanti per comprendere come il comportamento degli utenti e la loro percezione dell'ambiente virtuale possano essere tracciati, analizzati e interpretati. I più recenti strumenti tecnologici di tracciamento e visualizzazione dell'interazione utente in contesti immersivi aprono nuovi orizzonti per studi futuri, che potranno ampliare il campione, diversificare l'esperienza in più scene e raffinare ulteriormente le tecniche di traduzione intermediale.

Crediti

La ricerca rientra all'interno del Progetto H2IOSC – *Humanities and cultural Heritage Italian Open Science Cloud* (<https://www.h2iosc.cnr.it>) finanziato dall'Unione Europea – NextGenerationEU – Piano Nazionale di Ripresa e Resilienza (PNRR) – Mission 4 "Education and Research", Component 2 "From research to business", Investment 3.1 "Fund for the realization of an integrated system of research and innovation infrastructures", Action 3.1.1 "Creation of new research infrastructures strengthening of existing ones and their networking for Scientific Excellence under Horizon Europe" – Codice progetto IR0000029, CUP B63C22000730005. Ente attuatore CNR.

Concettualizzazione, S. M.; metodologia, C.F. A., B. F., G. G., S. M.; software, B. F., G. G.; validazione, C.F. A., B. F., G. G., S. M.; cura dei dati, C.F. A., S. M.; user experience, C.F. A., S. M.; redazione-preparazione della bozza originale, S. M.; redazione-revisione e editing, C.F. A., B. F., G. G., S. M. Tutti gli autori hanno letto e accettato la versione pubblicata del manoscritto.

Note

[1] All'interno dello studio convergono competenze di *analytics*, visualizzazione 3D interattiva, rappresentazione visiva e valutazione della *user experience*.

[2] Per la generazione AI del panorama a 360° è stato utilizzato *Skybox AI* (<https://skybox.blockadelabs.com/>).

[3] Per l'*editing* dell'immagine con strumenti generativi è stata utilizzata l'estensione *Firefly* integrata nelle recenti versioni di *Adobe Photoshop*.

[4] La lettura del brano è stata eseguita tramite la funzione *Text to Speech* di *ElevenLabs* (<https://elevenlabs.io/>).

[5] I suoni sono stati presi dalle librerie di effetti sonori della BBC (<https://sound-effects.bbcrewind.co.uk/>).

[6] ATON è un *framework open source* sviluppato dal DHiLab del CNR-ISPC (<https://osiris.itabc.cnr.it/aton/>).

[7] Il link per accedere all'esperienza VR è il seguente: <https://aton.ispc.cnr.it/s/smenconero/20241113-tpxfaj8i>.

[8] L'opportunità di proporre l'esperienza a un gruppo di utenti si è presentata in occasione di un evento di Terza Missione che ha coinvolto alcuni studenti e studentesse del Liceo Scientifico "Charles Darwin" di Roma il 17 dicembre 2024 presso il DHiLab del CNR-ISPC.

[9] *Interlumo* (<https://interlumo.ispc.cnr.it/>) è una *pipeline open source* sviluppata dal DHiLab del CNR-ISPC all'interno del progetto H2IOSC (<https://www.h2iosc.cnr.it/>).

[10] Nel caso studio *Il Barone rampante*, gli attributi tracciati sono la direzione e posizione del visore, la direzione del selettori che si attiva impugnando i *controller*, e la direzione, posizione e orientamento delle mani dell'utente mentre impugna i *controller*. La frequenza di campionamento è pari a 200 millisecondi per una durata massima di 3 minuti (che è leggermente superiore alla traccia audio dell'esperienza).

[11] Il link per visionare i risultati del tracciamento è il seguente: <https://aton.ispc.cnr.it/a/merkhet/?s=smentcone-ro%2F20241113-tpxfaj8i>.

[12] Una analisi più approfondita verrà svolta in seguito quando il campione di utenti sarà numericamente più elevato.

[13] È stato preso come riferimento il questionario pubblicato dal Cyberpsychology Lab dell'Université du Québec en Outaouais, che a sua volta riprende Witmer, Singer 1998 e Witmer et al. 2005.

[14] La scelta delle panoramiche 360° non è dipesa da questioni relative alla successiva analisi dei tracciamenti poiché *Interlumo* è in grado di operare anche all'interno di scene 3D.

Riferimenti bibliografici

Barenghi, M. (a cura di). (2023). *Favoloso Calvino. Il mondo come opera d'arte: Carpaccio, de Chirico, Gnoli, Melotti e altri*. Milano: Electa.

Belpoliti, M. (2006). *L'occhio di Calvino*. Nuova edizione ampliata. Torino: Giulio Einaudi.

Calvino, I. (2020). *Lezioni americane*. Milano: Mondadori.

Calvino, I. (2024). *I nostri antenati*. Milano: Mondadori.

Cianci, M. G., Calisi, D. (2014). Il Mondo è un libro: visioni ispirate da "Le città invisibili" di Italo Calvino. In A. Melián García (ed.) *El dibujo de viaje de los arquitectos. Actas del 15 Congreso Internacional de Expresión Gráfica Arquitectónica*. Las Palmas de Gran Canaria, 22-23 maggio 2014, pp. 751-759. Las Palmas de Gran Canaria: Universidad de Las Palmas de Gran Canaria, Servicio de Publicaciones y Difusión Científica.

Fanini, B., Ferdani, D., Demetrescu, E., Berto, S., d'Annibale, E. (2021). ATON: An Open-Source Framework for Creating Immersive, Collaborative and Liquid Web-Apps for Cultural Heritage. In *Applied Sciences*, vol. 11, n. 22. <https://doi.org/10.3390/app112211062>.

Fanini, B., Gosti, G., (2024). A New Generation of Collaborative Immersive Analytics on the Web: Open-Source Services to Capture, Process and Inspect Users' Sessions in 3D Environments. In *Future Internet*, vol. 16, n. 5. <https://doi.org/10.3390/fi16050147>.

Fucarino, E. A. (2015). *Le immagini raccontate. Immagini-testo in Italo Calvino*. Tesi di laurea in Teorie della Comunicazione. xUniversità degli Studi di Palermo. https://www.academia.edu/30047526/Le_immagini_raccontate_Immagini_testo_in_Italo_Calvino.

Jansson, M. (2018). Ekphrasis and Digital Media. In *Poetics Today*, vol. 39, n. 2, pp. 299-318. <https://doi.org/10.1215/03335372-4324468>.

Lindhé, C. (2013). "A visual sense is born in the fingertips": towards a digital ekphrasis. In *Digital Humanities Quarterly*, vol. 7, n. 1, pp. 1-14. <https://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-64434>.

Pinotti, A. (2024). *Il primo libro di teoria dell'immagine*. Torino: Einaudi.

Ricci, F. (2001). *Painting with Words, Writing with Pictures. Word and Image in the Work of Italo Calvino*. Toronto-Buffalo-London: University of Toronto Press.

Shiel, N. (2013). A Second Level Pictorial Turn? The Emergence of Digital Ekphrasis from the Visuality of New Media. In C. Fowley, C. English, S. Thouësny (eds). *Internet Research, Theory, and Practice: Perspectives from Ireland*, pp. 75-93. Dublin: Research Publishing. <https://files.eric.ed.gov/fulltext/ED576208.pdf>.

Witmer, B. G., Singer, M. J., (1998). Measuring Presence in Virtual Environments: A Presence Questionnaire. In *Presence*, vol. 7, n. 3, pp. 225-240. <https://doi.org/10.1162/105474698565686>.

Witmer, B. G., Jerome, C. J., Singer, M. J., (2005). The Factor Structure of the Presence Questionnaire. In *Presence*, vol. 14, n. 3, pp. 298-312. <https://doi.org/10.1162/105474605323384654>.

Autori

Sofia Menconero, CNR - Istituto di Scienze del Patrimonio Culturale, sofia.menconero@cnr.it
Chiara Florise Amadei, CNR - Istituto di Scienze del Patrimonio Culturale, chiaraflorise.amadei@cnr.it
Giorgio Gosti, CNR - Istituto di Scienze del Patrimonio Culturale, giorgio.gosti@cnr.it
Bruno Fanini, CNR - Istituto di Scienze del Patrimonio Culturale, bruno.fanini@cnr.it

Per citare questo capitolo: Sofia Menconero, Chiara Florise Amadei, Giorgio Gosti, Bruno Fanini (2025). Dall'èkphrasis all'esperienza VR immersiva nel Barone rampante di Italo Calvino. In L. Carlevaris et al. (a cura di), *èkphrasis. Descrizioni nello spazio della rappresentazione/èkphrasis. Descriptions in the space of representation*. Atti del 46° Convegno Internazionale dei Docenti delle Discipline della Rappresentazione. Milano: FrancoAngeli, pp. 3971-3990. DOI: 10.3280/oa-1430-c961.

From *Ekphrasis* to Immersive VR Experience in *Il barone rampante* by Italo Calvino

Sofia Menconero
Chiara Florise Amadei
Giorgio Gosti
Bruno Fanini

Abstract

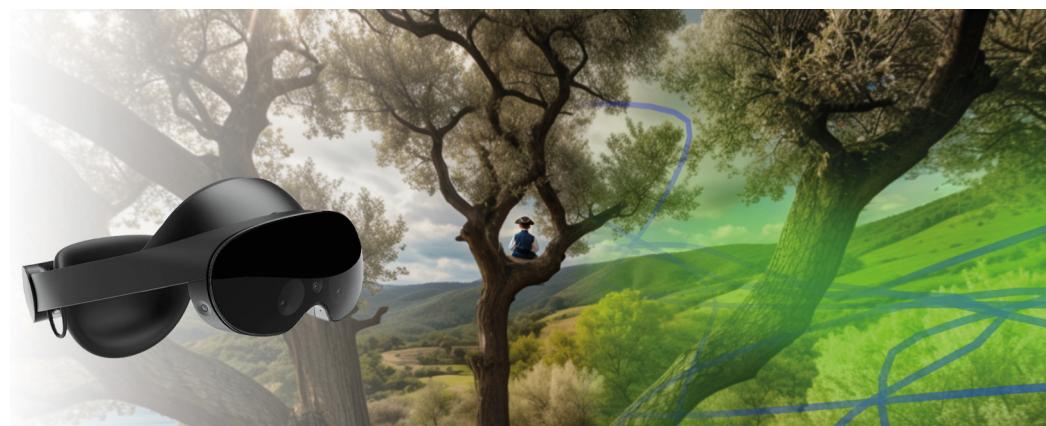
Ekphrasis, traditionally defined as the detailed verbal description of visual works, is reinterpreted in this study through a multidisciplinary approach integrating literature, generative artificial intelligence (AI), and virtual reality (VR). The research explores the intermedial translation of a text excerpted from *Il barone rampante* (*The baron in the Trees*) by Italo Calvino, renowned for his visual and descriptive style. By extracting keywords from the text, 360-degree panoramic images were generated using generative AI. These images, combined with an audio narration of the passage, were incorporated into a VR experience accessible via an HMD headset.

The study highlights how the immersive nature of VR enhances the relationship between image and narration, evolving from a literary description to a multisensory experience. The VR experience was tested on an exploratory sample, with participants' visual behaviour tracked using advanced tools. The collected data enabled the analysis of gaze direction and the correlation between visual and auditory elements, while a questionnaire on the sense of 'presence' provided feedback on the immersive effectiveness of the experience.

Despite the limitations of a small sample size, preliminary results demonstrate the potential of emerging technologies in intermedial translation. The research offers new perspectives on literary engagement, suggesting future developments involving larger and more diverse samples and expanded scene exploration, further investigating the role of *ékphrasis* in contemporary immersive contexts.

Keywords

Visual Studies, Generative AI, Immersive VR, Italo Calvino, User Behaviour.



Evocative image
summarising the
experience of intermedial
translation (authors' own
image).

Introduction

Èkphrasis, originally defined as a detailed verbal description of an artwork or visual scene, is a rhetorical device that bridges the worlds of words and images. In recent decades, èkphrasis has acquired new meanings and applications due to the growing intersection between literature, visual art, and digital technology [Shiel 2013; Lindhé 2013; Jansson 2018].

This study explores the concept of èkphrasis across multiple levels and through a multidisciplinary approach [1], with a particular focus on the interaction between literary text, AI-generated imagery, and immersive virtual reality (VR). The proposed method involves selecting a descriptive passage from a narrative, extracting key elements to serve as prompts for generating a 360° panoramic image using generative AI. The resulting image, combined with an audio reading of the corresponding novel excerpt, was then integrated into a VR experience presented to users, who could immerse themselves in the virtual environment using a headset.

During these sessions, user behaviour was tracked, allowing for data collection on gaze direction and the derivation of an iconic description of the experience –thus completing the circle and shifting the focus of èkphrasis from a bridge between verbal and visual to one between visual and visual.

The literary work selected for this methodological experiment was *Il barone rampante* by Italo Calvino (*The Baron in the Trees*) [Calvino 2024, pp. 93-338]. Calvino was chosen due to his propensity for detailed descriptions of narrative settings [Ricci 2001], to the extent that he is sometimes associated with èkphrasis [Fucarino 2015] and has inspired graphic interpretations of his scenes [Cianci and Calisi 2014; Barenghi 2023]. Calvino is a “visual” writer [Belpoliti 2006], and he dedicated one of his *Sei Lezioni Americane* (*Six Memos for the Next Millennium*) to the concept of visibility [Calvino 2020, pp. 81-100].

This contribution aims to analyse the entire process, exploring the challenges and opportunities presented by intermedial translation and immersive VR. Specifically, it will illustrate how èkphrasis can evolve in a contemporary technological context, transforming from a purely literary phenomenon into a multisensory experience capable of enhancing comprehension and immersion in the original work.

From èkphrasis to AI-Generated VR experience

Il barone rampante tells a portion of the peculiar and imaginative life of Cosimo Piovacco di Rondò, narrated by his brother Biagio. Set in the second half of the 18th century in a Ligurian coastal village, the story starts with a pivotal moment in the young protagonist's life: after an argument with his father, he decides to spend the rest of his days living among the trees, never setting foot on the ground again.

The section of the novel examined in this study spans the end of the first chapter and the beginning of the second –the moment when Cosimo first climbs a tree and observes the world around him from above (fig. 1) [Calvino 2024, pp. 106-107].

To translate this literary text into an immersive, interactive, and multisensory experience, both visual and auditory elements were employed. On one hand, there was a need to create a graphical context inspired by the text's setting; on the other, a narrated voice and soundscape were incorporated.

The scene was rendered as a 360° panorama generated by a specialised AI text-to-image service [2]. From Calvino's original text, a series of keywords were selected to capture the landscape and atmosphere described in the passage. These keywords were then woven into a simple natural-language prompt (fig. 2).

In addition to Calvino's description, supplementary details (chronological and geographical context, time of day) were included –though not explicitly stated in the selected passage– to guide the AI's generation more effectively. To simulate the protagonist's elevated perspective, an AI model mimicking drone vision was used. The

Cosimo salì fino alla forcella d'un grosso ramo dove poteva stare comodo, e si sedette lì, a gambe penzoloni, a braccia incrociate con le mani sotto le ascelle, la testa insaccata nelle spalle, il tricorno calcato sulla fronte.

Nostro padre si sporse dal davanzale: – Quando sarai stanco di star lì cambierai idea!

– gli gridò.

– Non cambierò mai idea, – fece mio fratello, dal ramo.

– Ti farò vedere io, appena scendi!

– E io non scenderò più! – E mantenne la parola.

Cosimo era sull'elce. I rami si sbracciavano, alti ponti sopra la terra. Tirava un lieve vento; c'era il sole. Il sole era tra le foglie, e noi per vedere Cosimo dovevamo farci schermo con la mano. Cosimo guardava il mondo dall'albero: ogni cosa, vista di lassù, era diversa, e questo era già un divertimento. Il viale aveva tutt'un'altra prospettiva, e le aiole, le ortensie, le camelie, il tavolino di ferro per prendere il caffè in giardino. Più in là le chiome degli alberi si sfittivano e l'ortaglia digradava in piccoli campi a scala, sostenuti da muri di pietre; il dosso era scuro di oliveti, e, dietro, l'abitato d'Ombrosa sporgeva i suoi tetti di mattone sbiadito e ardesia, e ne spuntavano pennoni di bastimenti, là dove sotto c'era il porto. In fondo si stendeva il mare, alto d'orizzonte, ed un lento veliero vi passava.

Fig. 1. Excerpt from *Il barone rampante* by Italo Calvino, selected for the experiment [Calvino 2024, pp. 106-107].

In the garden of an Italian villa on a midday in the 1700s. From the PoV of a high branch of an oak tree, one sees a garden with trees, pathways, flowerbeds, hydrangeas, camellias, and a wrought-iron coffee table. In the distance, the treetops reveal fields supported by stone walls and olive groves. On the other side, an Italian village with faded brick and slate rooftops is visible. Beyond the village lies the port, and in the background, the sea with a sailing ship. Just one person sitting on an oak branch.

Fig. 2. Prompt used for generating the 360° panorama (authors' own image).

resulting output, produced in the short timespan typical of AI generation, was reasonably successful, though it lacked some key narrative elements (fig. 3). These were later added through image editing: the equirectangular image was converted into a cubemap projection for distortion-free adjustments.

Using generative tools [3], the following elements were incorporated: Cosimo perched on a tree fork wearing a tricorn hat, flower beds, a coffee table, and a distant sailing ship. Finally, the image was converted back to an equirectangular projection (fig. 4).

For the auditory component, an AI text-to-speech service was employed [4]. This technology enables professional-quality results with a wide range of options in terms of language, gender, age, and style. For the reading of *Il barone rampante*, a modulated young male Italian voice was selected from the narrative category, aiming to simulate Biagio's storytelling. Natural soundscapes –distant waves and birdsong [5]– were layered beneath the narration.

The VR experience was built using the open-source framework ATON [6] [Fanini et al. 2021], which facilitates the implementation of 3D or 360° scenes (enriched with multi-media content such as video and audio) and their web-based accessibility across multiple devices (desktop, smartphones, headsets etc.), including extended reality (WebXR). Through a persistent link, users could freely access and experience the content on any device [7] (fig. 5).

Given the immersive and interactive nature of the experience –comprising both visuals and sound– a head-mounted display (HMD) was recommended for optimal engagement.



Fig. 3. Original equirectangular image obtained through AI generation (authors' own image).



Fig. 4. Edited equirectangular image with the addition of key elements (authors' own image).

Describing user behaviour in VR sessions

The distinction between 'seeing' and 'looking' is crucial to understanding the experience of *èkphrasis*. 'Seeing' implies passive visual perception – a stimulus reaching the eye and being processed by the brain without necessarily triggering deep awareness. 'Looking', conversely, is an intentional act requiring attention and active participation from the observer. In an immersive context like VR, this distinction is amplified: users do not merely see the created environment but are invited to 'look' attentively and curiously, exploring the space, lingering on details, and mentally reconstructing the meaning of the images in relation to the literary text – pairing the narrative quality of the visuals with the vocal storytelling.

For this study, user experience was deemed an additional point of interest. On one hand, it represents a further development of *èkphrasis* – namely, an iconic description of user behaviour by tracking headset direction (and, by extension, gaze) and visualising this data. On the other, a post-experience questionnaire measuring the sense of 'presence' in virtual environments provided valuable feedback.

A prototype of this experiment involved 16 participants [8] – a sample neither large nor diverse enough for statistically reliable analysis (which will require a broader user base) but sufficient for preliminary exploratory assessment and methodological testing.



Fig. 5. Selected views from the experience (authors' own images).

Tracking and data visualisation were enabled by *Interlumo* [9]: an open source service pipeline designed for researchers studying large-scale interactions from remote applications, public or lab installations, virtual or physical spaces, and online web apps [Fanini, Gosti 2024]. *Interlumo* comprises three macro-components: capture, processing, and inspection. The first (*Kapto*) records interactions from remote applications as numerical data [10]. The second (*Procezo*) processes and aggregates records to analyse user patterns via analytics, including filtering, cleaning, generating statistics, and applying ML (Machine Learning) models (e.g., estimating most-viewed areas or

fixation points). The third (Merkhet) allows for immersive/visual inspection of this data via a standard web browser, facilitating advanced analysis and behavioural assessment. It also offers tools for annotating sessions, spatial interfaces for immersive modes, and collaborative features.

Interlumo's visualisation, built on the ATON framework, provides two key insights for 360° panoramas: the user's observation trajectory and a heatmap of primary fixation points [11] (fig. 6). For the former, loading a record enables anchoring the scene's viewpoint to the recorded perspective, with a timeline slider tracking gaze shifts. For the latter, fixation areas can be computed across multiple records simultaneously. These metrics helped identify which scene elements were observed, for how long, and their relationship to the auditory component. Partial findings [12] revealed two highly viewed areas: the cluster of buildings and Cosimo in the tree. The former likely stems from the experience's initial framing, while the latter's prominence is attributed to being the sole human figure in the scene.

After the experience, users completed a questionnaire on their sense of 'presence' –one of three interrelated axes defining iconic immersive experiences: *scorniciamento* (framing), 'presenza' (presence), and 'immediatezza' (immediacy) [Pinotti 2024, pp. 276-281]. 'Presence' is defined as "the subjective experience of being in one place or environment, even when physically situated in another" [Witmer, Singer 1998, p. 225]. A Likert-scale questionnaire, slightly adapted for *Il barone rampante* [13], was used to measure this. Though the results lack statistical robustness due to sample size and homogeneity, the aggregated responses are included for completeness (fig. 7).

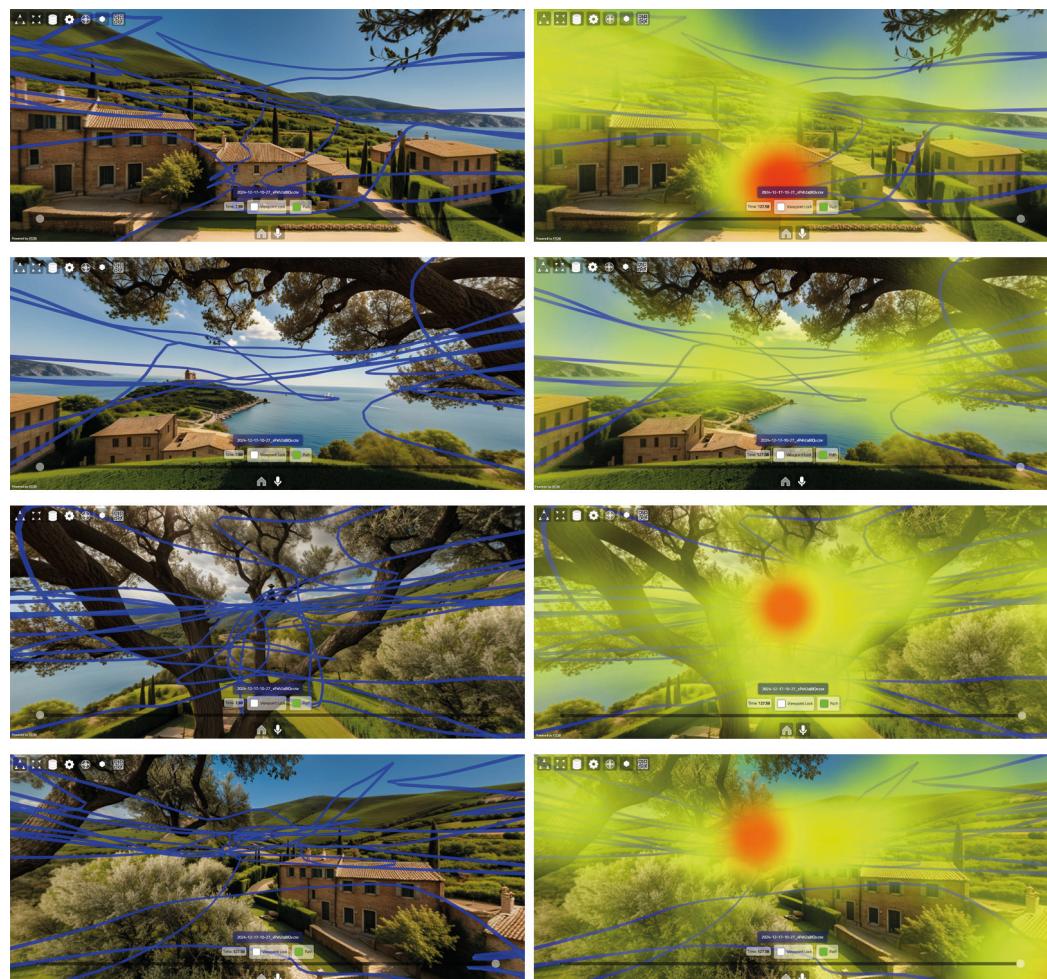


Fig. 6. Left: blue trajectory indicating a user's viewing direction; right: heatmap showing the same user's fixation points (authors' own images).

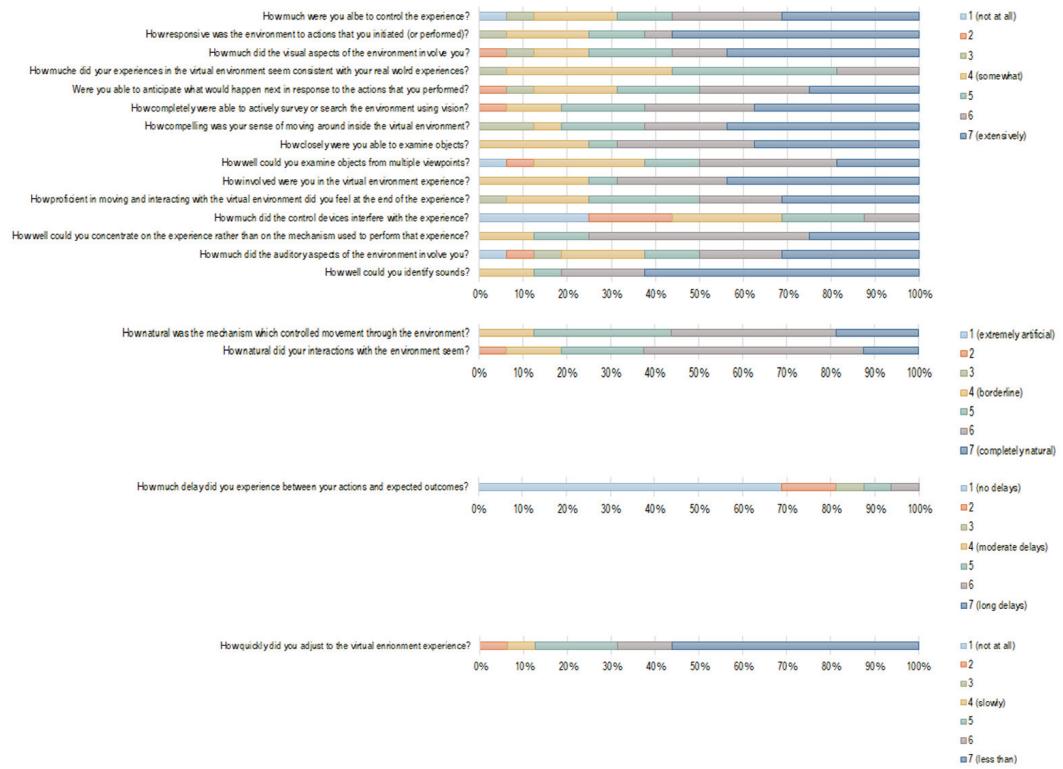


Fig. 7. Results from the 'presence' questionnaire completed by experiment participants (authors' own images).

Conclusions

"At the origin of every story I've written, there's an image that spins in my mind – born who knows how – and which I carry with me for years. Gradually, I develop this image into a story with a beginning and an end, and at the same time – though the two processes are often parallel and independent – I become convinced that it holds some meaning" [translated from Calvino 2024, p. 461].

This study draws inspiration from Calvino's strong affinity for images, proposing a method for intermedial translation of a scene from one of his celebrated works. The integration of literature, AI, and VR has enabled new approaches to intermedial translation, opening intriguing possibilities for the engagement and interpretation of literary texts.

The proposed method relies on generative AI to produce a visual and auditory panorama. A 360° image was preferred over 3D modelling to ensure scalability across multiple literary passages, as 3D reconstruction would be time-prohibitive [14]. While monoscopic 360° imagery limits exploration to three DoF (Degrees of Freedom), permitting only stationary navigation, this did not appear to hinder user immersion – a finding supported by questionnaire responses indicating high engagement with visual, auditory, and overall experiential aspects.

Preliminary results, though constrained by a small and homogenous sample, offer valuable insights into how user behaviour and environmental perception can be tracked, analysed, and interpreted. Emerging tools for immersive interaction tracking and visualisation pave the way for future studies with expanded participant pools, diversified scenes, and refined intermedial translation techniques.

Credits

This research is part of the H2IOSC Project – *Humanities and cultural Heritage Italian Open Science Cloud* (<https://www.h2iosc.cnr.it>) funded by the European Union – NextGenerationEU – National Recovery and Resilience Plan (NRRP) – Mission 4 "Education and Research", Component 2 "From research to business", Investment 3.1 "Fund for the realization of an integrated system of research and innovation infrastructures", Action 3.1.1 "Creation of new research infrastructures strengthening of existing ones and their networking for Scientific Excellence under Horizon Europe" – Project code IR0000029, CUP B63C22000730005. Implementing Entity CNR.
Conceptualisation, S.M.; methodology, C.F. A., B.F., G. G., S. M.; software, B. F., G. G.; validation, C.F. A., B.F., G. G., S. M.; data curation, C.F. A., S. M.; user experience, C.F. A., S. M.; writing, original draft, S. M.; writing, review, editing, C.F. A., B.F., G. G., S. M. All authors have read and agreed the published version of the manuscript.

Notes

- [1] The study integrates expertise in analytics, interactive 3D visualisation, visual representation, and user experience evaluation.
- [2] Skybox AI (<https://skybox.blockadelabs.com/>) was used for 360° panorama generation.
- [3] Adobe Photoshop's *Firefly* extension was used for generative image editing.
- [4] ElevenLabs' Text to Speech (<https://elevenlabs.io/>) generated the narration.
- [5] Sound effects were sourced from BBC's sound libraries (<https://sound-effects.bbcrewind.co.uk/>).
- [6] ATON, an open-source framework by CNR-ISPC's DHiLab (<https://osiris.itabc.cnr.it/aton/>).
- [7] VR experience link: <https://aton.ispc.cnr.it/s/smencenero/20241113-tpxfaj8i>.
- [8] Participants included students from Liceo Scientifico "Charles Darwin" (Rome) during a 17/12/2024 outreach event at CNR-ISPC's DHiLab.
- [9] *Interlumo* (<https://interlumo.ispc.cnr.it/>) is an open-source pipeline developed by CNR-ISPC's DHiLab under H2IOSC project (<https://www.h2iosc.cnr.it/>).
- [10] For *Il barone rampante*, tracked attributes included headset direction/position, controller selector direction, and hand position/orientation (200ms sampling rate, max 3-minute duration).
- [11] Tracking results: <https://aton.ispc.cnr.it/a/merkhet/?s=smencenero%2F20241113-tpxfaj8i>.
- [12] Deeper analysis will follow with a larger user sample.
- [13] The questionnaire was adapted from Cyberpsychology Lab (UQO), based on Witmer; Singer 1998 and Witmer et al. 2005.
- [14] 360° panoramas were chosen for scalability; *Interlumo* also supports 3D scene analysis.

Reference List

Barenghi, M. (a cura di). (2023). *Favoloso Calvino. Il mondo come opera d'arte: Carpaccio, de Chirico, Gnoli, Melotti e altri*. Milano: Electa.

Belpoliti, M. (2006). *L'occhio di Calvino*. Nuova edizione ampliata. Torino: Giulio Einaudi.

Calvino, I. (2020). *Lezioni americane*. Milano: Mondadori.

Calvino, I. (2024). *I nostri antenati*. Milano: Mondadori.

Cianci, M. G., Calisi, D. (2014). Il Mondo è un libro: visioni ispirate da "Le città invisibili" di Italo Calvino. In A. Melián García (ed.). *El dibujo de viaje de los arquitectos*. Actas del 15 Congreso Internacional de Expresión Gráfica Arquitectónica. Las Palmas de Gran Canaria, 22-23 maggio 2014, pp. 751-759. Las Palmas de Gran Canaria: Universidad de Las Palmas de Gran Canaria, Servicio de Publicaciones y Difusión Científica.

Fanini, B., Ferdani, D., Demetrescu, E., Berto, S., d'Annibale, E. (2021). ATON: An Open-Source Framework for Creating Immersive, Collaborative and Liquid Web-Apps for Cultural Heritage. In *Applied Sciences*, vol. 11, n. 22. <https://doi.org/10.3390/app112211062>.

Fanini, B., Gosti, G., (2024). A New Generation of Collaborative Immersive Analytics on the Web: Open-Source Services to Capture, Process and Inspect Users' Sessions in 3D Environments. In *Future Internet*, vol. 16, n. 5. <https://doi.org/10.3390/fi16050147>.

Fucarino, E. A. (2015). *Le immagini raccontate. Immagini-testo in Italo Calvino*. Tesi di laurea in Teorie della Comunicazione. xUniversità degli Studi di Palermo. https://www.academia.edu/30047526/Le_immagini_raccontate_Immagini_testo_in_Italo_Calvino.

Jansson, M. (2018). Ekphrasis and Digital Media. In *Poetics Today*, vol. 39, n. 2, pp. 299-318. <https://doi.org/10.1215/03335372-4324468>.

Lindhé, C. (2013). "A visual sense is born in the fingertips": towards a digital ekphrasis. In *Digital Humanities Quarterly*, vol. 7, n. 1, pp. 1-14. <https://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-64434>.

Pinotti, A. (2024). *Il primo libro di teoria dell'immagine*. Torino: Einaudi.

Ricci, F. (2001). *Painting with Words, Writing with Pictures. Word and Image in the Work of Italo Calvino*. Toronto-Buffalo-London: University of Toronto Press.

Shiel, N. (2013). A Second Level Pictorial Turn? The Emergence of Digital Ekphrasis from the Visuality of New Media. In C. Fowley, C. English, S. Thouësny (eds). *Internet Research, Theory, and Practice: Perspectives from Ireland*, pp. 75-93. Dublin: Research Publishing. <https://files.eric.ed.gov/fulltext/ED576208.pdf>.

Witmer, B. G., Singer, M. J., (1998). Measuring Presence in Virtual Environments: A Presence Questionnaire. In *Presence*, vol. 7, n. 3, pp. 225-240. <https://doi.org/10.1162/105474698565686>.

Witmer, B. G., Jerome, C. J., Singer, M. J., (2005). The Factor Structure of the Presence Questionnaire. In *Presence*, vol. 14, n. 3, pp. 298-312. <https://doi.org/10.1162/105474605323384654>.

Authors

Sofia Menconero, CNR - Institute of Heritage Science, sofia.menconero@cnr.it
Chiara Florise Amadei, CNR - Institute of Heritage Science, chiaraflorise.amadei@cnr.it
Giorgio Gosti, CNR - Institute of Heritage Science, giorgio.gosti@cnr.it
Bruno Fanini, CNR - Institute of Heritage Science, bruno.fanini@cnr.it

To cite this chapter: Sofia Menconero, Chiara Florise Amadei, Giorgio Gosti, Bruno Fanini (2025). From Ékphrasis to Immersive VR Experience in *Il barone rampante* by Italo Calvino. In L. Carlevaris et al. (Eds.). *ékphrasis. Descrizioni nello spazio della rappresentazione/ékphrasis. Descriptions in the space of representation*. Proceedings of the 46th International Conference of Representation Disciplines Teachers. Milano: FrancoAngeli, pp. 3971-3990. DOI: 10.3280/oa-1430-c961.