

國立臺北藝術大學  
電影與新媒體學院  
新媒體藝術學系碩士班

M.F.A.Program, Department of New Media Art  
School of Film and New Media  
Taipei National University of Arts

碩士論文

基於場景語意認知的互動藝術應用

Interactive Art Applications Based on Scene Semantic Cognition

研究生：林玥汝 撰  
Graduate Student: Yue-Ru Lin

指導教授：孫士韋 博士  
Thesis Supervisor: Shih-Wei Sun

中華民國 113 年 8 月

August, 2024

(題目)基於場景語意認知的互動藝術應用

指導教授：孫士韋

學生：林玥汝

## 中文摘要

本論文以心理學為其創作背景，探究人類受到社會影響(Social Influence)下，所產生的感受、行為的狀態，並透過互動科技的方式，如 TouchDesigner、OpenCV、Google ARCore 及 ChatGPT 等功能呈現。筆者提出三件作品〈高敏感的聽說讀寫練習〉、〈Hello World〉以及〈鏡中境〉，第一件作品〈高敏感的聽說讀寫練習〉，主要是利用 Touch Designer、OpenCV 完成相機校正及人體辨識等功能，其創作為探討高敏感個體的內心感受，著重在高敏感族群對言語的過度分析的行為，及其內心世界的複雜性。第二件作品〈Hello World〉，透過 Google ARCore 的 Scene Semantic 及 Geospatial 呈現 AR 效果，並透過 OpenAI 和 AWS 所提供的套件完成 Speech to Text、Text to Speech 功能，其創作是以認知心理學作為出發，當我們與整體社會互動下，所有的言語、事物的解讀，都是從我們自身的心理狀態出發，並以此作品探討觀看與看見之間的差異。第三件作品〈鏡中境〉，利用 ESP32 及 Hologram 呈現，概念為筆者的自身經歷結合心理學的視覺影響，將過往的生命經驗所感受到的情緒，再轉而以一個觀察者的角度觀看當時的狀態，並透過視覺影響下，將作品的影像從「主體物體」轉移為「整體」。

關鍵字：高敏感、認知心理學、知覺、認知、Touch Designer、OpenCV、相機校準、AR Core、Scene Semantic、Geospatial、Hologram

Interactive Art Applications Based on Scene Semantic Cognition

Thesis Supervisor: Shih-Wei Sun

Student: Yue-Ru Lin

## **ABSTRACT**

This thesis uses psychology as its creative background to explore the various thoughts, feelings, and even behavioral states that humans experience under social influence. It employs interactive technologies such as Touch Designer, OpenCV, Google ARCore, and ChatGPT. In this thesis, the author presents three works, "Highly Sensitive Listening, Speaking, Reading, and Writing Exercises", "Hello World" and "Realm in the Mirror". The first piece, "Highly Sensitive Listening, Speaking, Reading, and Writing Exercises". This piece use Touch Designer and OpenCV to perform camera calibration and human recognition. The concept is exploring the inner feelings of highly sensitive individuals, focusing on their behavior of over-analyzing language. This work aims to help viewers understand the complexity of the inner world of highly sensitive individuals through the process of trying to capture elusive messages while moving in front of the interactive installation. The second piece, "Hello World," utilizes Google ARCore's Scene Semantics and Geospatial features to present AR effects. Additionally, it incorporates Speech to Text and Text to Speech functionalities provided by OpenAI and AWS. This creation is based on cognitive psychology, emphasizing that all our interpretations of speech and events in our interactions with society originate from our psychological state. Through this work, I explore the difference between looking and seeing. The third piece, "Realm in the Mirror" employs ESP32 and Hologram for its presentation. The concept is rooted in the author's personal experiences combined with the psychological impact of visual perception. It reflects the emotions felt from past life experiences and re-examines them from the perspective of an observer.

Keywords : High Sensitivity, Cognitive Psychology, Perception, Cognition, Touch Designer, OpenCV, Camera Calibration, ARCore, Scene Semantic, Geospatial, Hologram

## 誌謝

在北藝大的這段時間，像是走在撒哈拉沙漠一樣，一眼望去盡是無限的寬廣與自由，然而徜徉在自由中，卻常讓人不小心就迷失了方向，像是沙漠中的旅人一樣，徒步地一直走著，有時怠惰了，想著：「算了，反正在遙遠的路程中，每一步就像是一粒沙，不走應該也無所謂」，但有時又忍不住抬頭望著北極星，嘗試在風沙星辰中，穿越到另一頭的終點。

慶幸的是在這段旅途，有家人一直在身旁的陪伴與照顧，有熊、駿廷、嘉璣、中誼、芷呈等摯友們的加油打氣，且總是包容我的難約和不定時人間蒸發的狀態，品翰、冠余長期聽我發牢騷、接收我的負能量，以及由於在北藝大期間幾乎是邊工作邊就讀的狀態，所以也非常感謝財政部財政資訊中心的科長及同事們的體諒，讓我可以休假去學校上課、前同事 Milly 在我最十萬火急的時候不吝嗇地給予技術諮詢、同事 Iris 默默幫我把我的工作分攤掉，減輕我的 Loading。在北藝大期間也要感謝卑鄙源之助的小卑鄙們，讓我的研究所生活有了歸屬感，以及新媒系的小民、佑任、巧芸在器材、空間、文件及時程等的熱心幫忙及提醒，讓兩光的我能夠趕上每一次 Deadline，和王連晟老師、游創文老師能夠擔任我的口試委員，給予鼓勵及建議，以及最要感謝的是孫士韋老師，在研究所期間不管是在技術上、概念上，總是給予無條件的協助，在每一次 Meeting 都花費非常多的時間幫助我釐清現況、癥結點，成為我在這段旅程中指引方向的北極星。

最後，感謝當初那個想要有始有終的我，讓我找回完整的自己，也期許未來能夠繼續加油，有機會回饋在這一路上曾給予我協助的每一個人。

# 目次

中文摘要.....	ii
<b>ABSTRACT</b> .....	iii
誌謝.....	iv
目次.....	v
表次.....	vii
圖次.....	viii
第一章 緒論.....	1
第二章 相關作品與文獻探討.....	2
第一節 心理學發想之探討.....	2
一、高敏感.....	2
二、認知心理學.....	4
第二節 相關技術之探究.....	8
一、人臉辨識及人體辨識.....	8
二、Google ARCore.....	11
三、相機校準.....	17
第三章 作品論述與執行.....	18
第一節 高敏感的聽說讀寫練習.....	18
一、創作理念.....	18
二、創作方法與技巧.....	19
三、呈現與紀錄.....	27
第二節 Hello World.....	29
一、創作理念.....	29
二、創作方法與技巧.....	30
三、呈現與紀錄.....	30
第三節 鏡中境.....	48
一、創作理念.....	48
二、創作方法與技巧.....	49
三、呈現與紀錄.....	62
第四章 結語.....	63

參考文獻..... 66

## 表次

表 1：Scene Semantics 辨識物件.....	11
表 2：〈高敏感的聽說讀寫練習〉作品硬體清單.....	20
表 3：〈高敏感的聽說讀寫練習〉展出紀錄.....	28
表 4：〈Hello World〉硬體規格表.....	34
表 5：〈Hello World〉作品互動流程圖.....	44
表 6：〈顯微鏡〉硬體規格表.....	50
表 7：〈顯微鏡〉動作設計圖.....	56
表 8：〈顯微鏡〉作品互動流程圖.....	59

## 圖次

圖 1：感官刺激到知覺再到認知階段.....	4
圖 2：Jenny Holzer，《威尼斯 1990》（Venice 1990）。1990。.....	8
圖 3：OpenCV 區塊圖.....	9
圖 4：Rafael Lozano-Hemmer，《Tape Recorders》。2011。.....	10
圖 5：Rafael Lozano-Hemmer，《Tape Recorders》程式偵測畫面。2011。.....	10
圖 6：René Magritte，《形象的叛逆》（La trahison des images）。1928-1929。....	13
圖 7：Joseph Kosuth，《一把和三把椅子》（One and Three Chairs）。1965。.....	13
圖 8：ARCore 操作指引.....	14
圖 9：藝術家卡米爾·瓦拉拉 (Camille Walala) 的 AR 壁畫.....	15
圖 10：2019 年 Pipilotti Rist 的 International Liquid Finger Prayer.....	16
圖 11：Adrien M & Claire B 的 Acqua Alta.....	17
圖 12：〈高敏感的聽說讀寫練習〉硬體擺設.....	20
圖 13：實體環境的特徵點.....	21
圖 14：Blender 內的模型.....	22
圖 15：Touch Designer 內的 Cam Schnappr 模擬校準結果.....	22
圖 16：實體投影的對位結果.....	23
圖 17：〈高敏感的聽說讀寫練習〉軟體系統架構圖.....	24
圖 18：〈高敏感的聽說讀寫練習〉Server 端軟體流程圖.....	25
圖 19：〈高敏感的聽說讀寫練習〉Client 端軟體流程圖.....	26
圖 20：〈高敏感的聽說讀寫練習〉空間配置圖.....	27
圖 21：紅色區域為湖面、黃色區域為樹木.....	31
圖 22：當使用者站在較遠距離，樹木面積較小，可看到的文字範圍也較小.....	33
圖 23：當使用者站在較遠距離，樹當使用者走近後，隨著樹木在視覺上的面積變大，視覺可閱讀的文字範圍也越多.....	33
圖 24：ARCore Scene Semantic 效果.....	34
圖 25：ARCore 匯入 Unity 內的地理位置資訊.....	35
圖 26：ARCore 實際地理位置的 AR 結果.....	36
圖 27：Cesium 匯入 Unity 內的地理位置資訊.....	37
圖 28：對話生成架構.....	38
圖 29：對話生成架構.....	39
圖 30：〈Hello World〉作品關卡 1 的軟體架構圖.....	41
圖 31：〈Hello World〉作品關卡 2 的軟體架構圖.....	42
圖 32：〈Hello World〉作品關卡 3 的軟體架構圖.....	43
圖 33：〈Hello World〉實際展示畫面 1.....	46
圖 34：〈Hello World〉實際展示畫面 2.....	47
圖 35：〈鏡中境〉作品結構.....	50
圖 36：〈鏡中境〉作品軟硬體系統架構圖 1.....	52

圖 37：〈鏡中境〉作品 Server 與 ESP32 之流程圖.....	53
圖 38：〈鏡中境〉作品 Server 與 App 之流程圖.....	54
圖 39：影片的製作流程.....	55
圖 40：筆者的動態影片與 Rokoko Vision 結果畫面 .....	56
圖 41：〈鏡中境〉實際展示畫面.....	62

# 第一章 緒論

「我們的知識和信仰會影響我們觀看事物的方式。」

(約翰·伯格, 2021) [1]

筆者過去背景為幼兒教育學系，故在大學時期接觸到心理學、社會學、人類發展等學科，畢業專題以心理學為基礎，研究「彼得潘症候群」[2]，畢業後對於社會心理等議題相當感興趣，尤其在智慧型手機的普及、社群軟體的蓬勃發展下，資訊的可及性變得更加唾手可得，不再只仰賴傳播媒體、書籍等，透過智慧型手機，在搜尋引擎上鍵入關鍵字，不用一秒隨即出現數萬筆資訊正等待被瀏覽。然而，資訊是如此便利且多元，不論何種議題的多元論述都可立即找到資料來源，讓各方得以在網路世界上訊息交流，既然如此，我們又為何替自己築起一道又一道無形的牆，讓自己躲在同溫層中，愈加根深柢固地相信自己的立論。

John Berger 在《觀看的方式》中提及「我們注視的從來不只是事物本身，我們注視的永遠是事物與我們之間的關係。」[1]，觀者在注視訊息的同時，要能脫離自身立場、擺脫原有認知，談何容易？更不用說因陷入關係中，增加了複雜化的可能性。

本篇創作論文之架構分為四章，第一章為創作發想之論述；第二章為文獻探討，以認知心理學(Cognitive Psychology)作為主要的學理基礎及創作發想，以及相關技術的研究，作為實現其創作的基石；第三章是對應第二章作品的創作理念、創作方法與技巧，記錄作品中的硬體及軟體的細部說明；第四章為筆者日後的作品計畫與相關的技術研究。

## 第二章 相關作品與文獻探討

資訊的演進下，透過科技使我們可以輕易地拉近人與人之間的距離，理解不同的背景、不同的文化，大幅縮小個體間的差距，在此條件之下，原本應該感受到觀看的角度變得更加理性化的痕跡，回歸內心對於他人也應該有更深切的體會。

然而，即使在這樣的條件下，我們仍然不時遇到人際相處間的誤解，造成個體間的嫌隙，甚至因過度延伸他人意圖，而使得自身感到鬱悶。同時，因現代個人意識抬頭，開始關注自我身心的人，也越來越多。

相關的心理學的理論協助我們了解這些現況為什麼存在，我們需要意識到這些感知、認知和自我實驗預言的影響，並努力超越它們，以更客觀和全面的方式理解他人和事件，這樣才可能真正回歸內心的平靜。

本論文是觀察存在大眾日常生活間的「高度敏感」及「認知心理」，並利用認知心理學的學理基礎剖析人類行為及背後動機，進而作為互動科技創作發想，目的是希望讓觀者從行為上反思自我看待事物的角度。

### 第一節 心理學發想之探討

#### 一、高敏感

「高敏感」是指高度敏感，英文名稱為「Highly Sensitive Person」，伊爾斯·山德(Ilse Sand)(2017)《高敏感是種天賦》[3]一書提及此名稱由美國心理學與研究專家伊蓮·艾融(Elaine N. Aron)於1996年所提出，高敏感和「五感」特別敏感有關，五感對於接受到的刺激和反應都會特別深刻，相較於他人更容易受到外在環境的干擾，如聲音、味道、光線，甚至溫度等影響。

根據網站療日子健康新聞[4]所統整，這類型人格通常具有以下 3 點特徵

1. 感官敏感：視覺、聽覺、嗅覺、味覺和觸覺會比一般人來得更加敏銳，例如一般人在進行對話過程中，通常會專注於對話內容本身，但是聽覺較敏感的人在對話的同時，還會察覺到周遭的環境音；或是視覺敏感的人會更容易受到光線的刺激。
2. 情緒敏感：相較於一般人，更容易察覺到他人的情緒，就像朋友即便沒有表達很多，但很快就能夠感受對方當下的情緒狀態。但與此同時，在這樣的狀態下，高敏感的人，也容易受到他人的情緒左右。
3. 反應敏感：高敏感的人通常會接收大量的資訊去做內在處理，所以在想事情或是在表達自己時，就可能會有非常多不同的角度去做分析，或是會有許多感受需要先消化才能做表達，在這樣的狀態下，反應或表達的時間就會相較於一般人更長一點。

在外在表現上，高敏感的人常見以下 4 種人格特質

1. 處理事務的深度較一般人高：想事情很全面、較一般人成熟、觀察事情較一般人透徹。但想太多、需要很多的時間消化、容易想到事情的負面情形因而帶來情緒上的困擾、內在情緒難以被他人理解。
2. 容易接收到龐大的刺激訊息：在環境中，可以同時觀察到很多細節，對於細節的掌握較一般人高很多。但也因為耗費太多的能量及注意力在掌握比一般人還要多的細節，容易疲累或感到不舒服、不自在。
3. 同理心及情緒反應：能夠理解其他人在想什麼，或是有著什麼樣的感受，但相反地，也容易過度配合、協助他人時，卻忽略了自己。
4. 對於細微刺激的敏感度很高，像是衣服上的毛線頭，刺刺的標籤等：容易被說想太多、過度反應等

## 二、 認知心理學

### (一)、感官刺激與知覺

在討論認知的解讀歷程上，可退一步思考在大腦解讀訊息之前，又有什麼  
是影響著我們的認知解讀。

威廉·詹姆士(William James) (1890)曾說「每個人都知道什麼是注意力，它以  
清晰而生動的形式佔據了心靈，從似乎同時存在的幾個可能的物件或思想聚焦、  
意識集中是其本質，它意味著為了有效地處理其他事情而放棄某些事情。」[5]然  
而，當人類將注意力關注於特定的物件或特徵時，將影響我們對於感覺世界中的  
刺激，進而改變我們的大腦對於訊息理解的過程。

在注意力的研究中，主要分為兩種類型的研究方向，分別為基於空間的注  
意力(Spatial Attention)和基於特徵點的注意力(Feature-Based Attention, FBA)，基於  
空間的注意力理論認為注意如同聚光燈會優先選擇空間位置，再對該位置的刺激  
進行知覺理解[6]，如整體式思考 (Holistic Processing)或分析式思考(Analytic  
Processing)。而基於特徵點的注意力理論則是認為視覺刺激是由空間和非空間特徵  
所建構而成，像是形狀、大小、顏色等，而不同的知覺理解方式也會產生不同的  
解讀[7]。

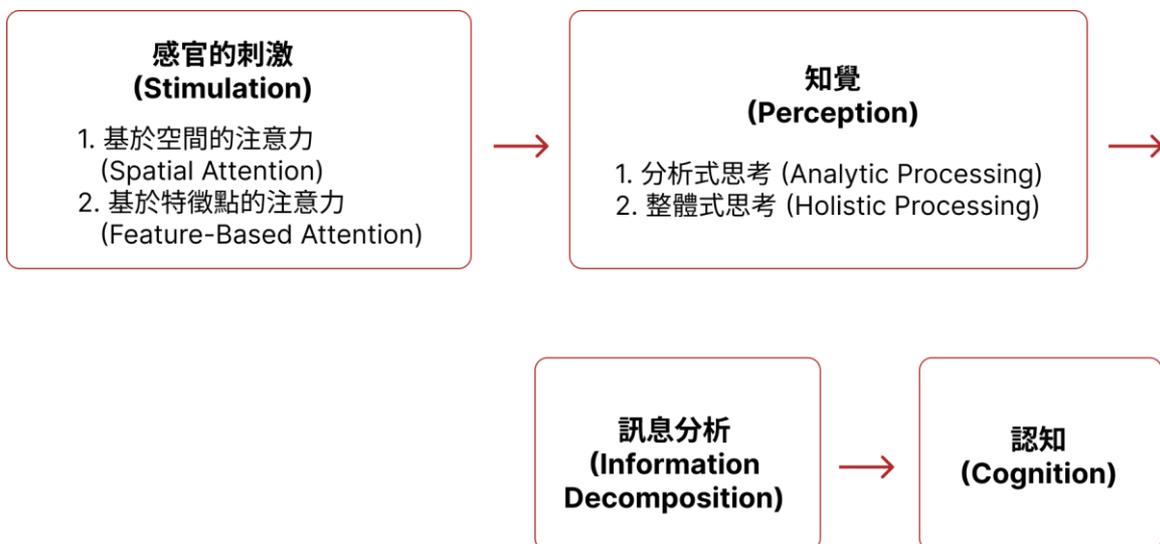


圖 1：感官刺激到知覺再到認知階段

分析式思考(Aalytic Processing)，較為強調主體特徵，並對涉及到的視覺訊息的細節進行分解和逐一分析，較未考慮主體與周遭物件的脈絡；而整體式思考(Holistic Processing)，則是對視覺訊息的整體性、全局性進行處理。

人類的視覺目光及思考歷程，不斷地動態切換，從注意力上的注視點到知覺理解的切換，使得人類的眼光彷彿如攝影機鏡一樣，看似自由地選擇運鏡的動線，切換著焦距對準那些隱藏在事物中的細節，而對於事物的解讀與理解則像是剪輯過的影片，但既使我們觀看著同一幅畫作，透過不同的視覺、基模的編導，最終還是用不同的鏡頭語言去闡述著各自的觀點。

## (二)、訊息分析與認知

「雖然我們堅信自己受到理性和意志力的引導，但其實是更常被自動化的過程牽著走。」

亞當·班福拉多(Adam Benforado)《不平等的審判》[8]

張淑美(2002)[9]提到 Robert L, Solso(1998)[10]指出認知，包含知覺、記憶與思考的過程，其心理過程包含感覺到知覺，再到腦神經學、組型再認識、注意力、意識、學習、記憶等等。而根據張春興(1990)[11]指出廣義的認知心理學包含皮亞傑(Jean Piaget)所提出的研究心智發展的認知心理學[12]，和里德(Reed)(1989)[13]所提出的訊息處理論。

- 皮亞傑認知心理學：由皮亞傑(Jean Piaget)所提出，認知發展是一個逐步成熟的心理過程，由個體不斷同化和重組外在經驗所引起（因此也被稱為「建構主義」）。兒童的心理建構過程基於對周圍世界的不斷理解，通過經歷成長過程中發現事物的差異並相應地調整他們的認知來實現。認知發展過程是人類的核心，而人類的語言形成正是通過認知發展中獲得的知識和理解為基礎的。

- 訊息處理論[14]：訊息處理理論的基本形式是將人腦比作電腦或基本處理器，跟電腦一樣以序列的方式工作，這個序列的順序是「接收輸入、處理訊息、傳送輸出」。

然而，當我們看到關鍵字時，常會不自覺地對該關鍵字自動產生聯想，甚至貼上標籤，而自動歸類是因為人類的知覺與記憶間存有一套複雜的組織系統，稱之為基模(Schemas)。基模是由個體習得的各種經驗、意識、概念等構成一個與外界現實世界相對應的抽象認知架構，掌控了我們對外在世界的概念，由於這些概念的屬性，以及這些屬性之間的關係，它對我們所注意的、所思考的及所記憶的資訊產生了極大的影響力。有賴於基模的運作，使得我們在接受訊息時，不必重新詳實的記錄每一個細節，而是只需要抓取關鍵特徵，並提取記憶中過去的相似經驗做比較及歸類，便可快速地組織大量訊息。

由於面對大量繁雜的訊息同時湧入，人們的大腦會為了降低大腦思考的機制，會自動地分析情境、環境、人物，這是基於過去經驗的世界的認知，這樣的行為被稱之為「自動化思考」(Automatic Thinking)[15]，以 Elliot Aronson 等人在《社會心理學》一書中對自動化思考的定義是下意識的、非自覺、非自主且不費力的思考方式，並引述了 Bargh JA 及 Ferguson MJ 在 2000 年的研究[16]，認定受測者對於社會干預、評估、判斷及思考運作都可以在沒有意識的情境下完成上述動作，確信人類的思考模式等同於自動化過程。

隨著人們的年紀漸長，基模也累積得越多，因此在自動化思考的歷程中，「促發作用」(Priming)[16]就扮演一個重要的角色，促發作用是指當前的經驗，導致某些相關的基模、特質或概念更快速地浮現出來，它會選擇哪些基模會成為對這件事情的解讀，往好處想它可以使我們更加快速地產生判斷的依據，但反向思考，人類是否真如我們所想的理性？舉例來說在 1982 年 John Bargh 和 Paula Pietromonaco，他們根據 Higgins 和同僚(1977)的研究[16]，找了兩群受試者，並讓受試者觀看螢幕上快速閃爍的文字，一群受試者觀看具有敵意的字眼，例如

Unkind、Hostile，另一群受試者觀看中性字眼，例如 Water、Between，文字閃爍速度之快到受試者只看到一個閃光，接著讓受試者閱讀同樣一段描述個人行為的文字，例如推銷員按了門鈴後，James 拒絕讓他進來，並判斷文中的行為是有敵意或無敵意，實驗結果為這兩群受試者都沒有意識到自己看到什麼字眼，但是看到敵意字眼的受試者對於閱讀後的文中行為判斷比中性字眼的受試者們都更具敵意。因此可以推斷出不為意識所理解的訊息，卻依然可能影響人們的判斷、態度與行為。

綜合上述的人類思考歷程，平心而論，錯誤的認知，背後或許不是蓄意、抱持惡意的心態出發，也不是純粹的無知、愚昧與輕信，這只是人類在演化出高度智能下，發展出得以迅速準確地處理大量資訊的弊端。

在藝術創作中，作品是透過藝術家的眼去述說其觀點，並企圖影響觀者進一步思考，Jenny Holzer 為美國著名女性藝術家，其創作風格是將明確而簡短的文字以不同媒材的載體傳遞出來，例如 LED 燈、建築投影...等[17]，其創作的觀點包含社會生活的各個層面，包括生存、道德、政治、權力...等，讓觀者從理解到連結自身生活經驗，引起人們的注意，進而開始思考這些語句背後所代表的意涵。

其中作品《威尼斯 1990》（Venice 1990）（圖 2），Jenny Holzer 在小房間內使用了 LED 長型裝置固定於周遭牆上，而 Jenny Holzer 之所以使用該媒材是為了呈現更模糊且更具有挑戰性的訊息，其文字以不同的色彩呈現，是為了喚起觀者的各種情緒，而 LED 上移動的文字或滾動或暫停，就像是口語表達時聲音的變化。[18]

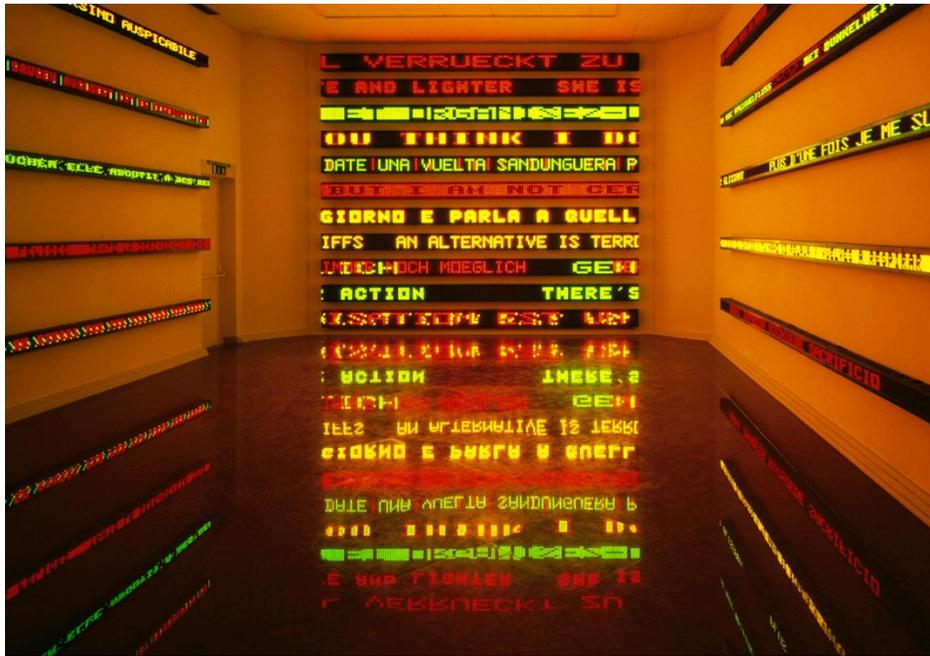


圖 2：Jenny Holzer，《威尼斯 1990》（Venice 1990）。1990。

圖片來源：<https://projects.jennyholzer.com/LEDs/venice-1990>。

## 第二節 相關技術之探究

本章節分為 3 個子章節來討論論文作品之相關技術應用。

### 一、人臉辨識及人體辨識

針對人臉辨識及人體辨識，現階段已有相當多的工具可以做為開發使用，本論文採用的技術以 OpenCV 為主。

OpenCV 的全名為 Open Source Computer Vision Library，是由 Intel 與 Google 公司所支援，依據 OpenCV 官方論文[19]提到在 1996 年時，Gary Bradski 在 Intel 公司服務時，為了促進電腦視覺與人工智慧的發展，於是發起 OpenCV 專案，以能夠提供一個跨平台的電腦視覺庫，協助人們快速建立視覺應用程式，目前 OpenCV 程式庫有超過 500 個函式，且此視覺庫可以在商業和研究領域中免費使用，因此現今的應用橫跨許多視覺領域，包括工廠產品檢驗、醫學影像、保全、相機校準等，最常被拿來用於開發即時的圖像處理、電腦視覺，以及圖型識別程式，常應用於擴增實境、臉部辨識、手勢辨識、動作辨識、運動跟蹤、物體

辨識或圖像分割等。

OpenCV 是以軟體層組成的，最上面的是運行 OpenCV 的作業系統，接下來是語言綁定與示範應用程式，它的下面在 opencv\_contrib 之中由別人貢獻的程式，多數都是高階功能，接下來是 OpenCV 的核心，而最底下的是硬體加速層 (Hardware acceleration layer, HAL) 的各種硬體最佳化程式(如圖 3)。

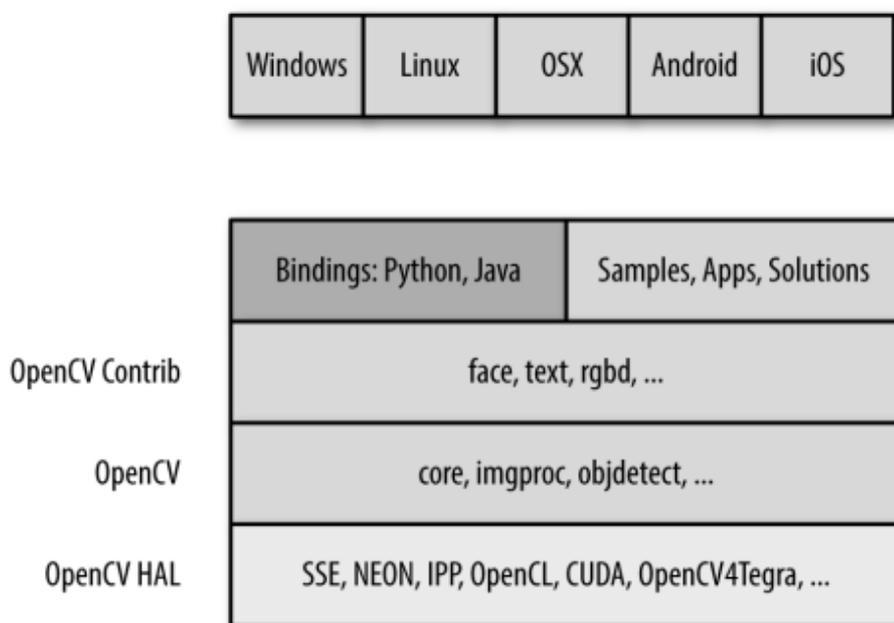


圖 3：OpenCV 區塊圖

圖片來源：<https://img2020.cnblogs.com/blog/1425036/202105/1425036-20210507161416712-2032477025.png>

在互動科技中，人臉辨識及人體辨識常被用來作為與觀者互動的手法，像是 Rafael Lozano-Hemmer 在 2011 年的作品《Tape Recorders》(圖 4)，其作品於空間中擺放數個電動測量卷尺以記錄訪客在裝置中停留的時間。當 Kinect 檢測到空間中有觀者存在時(圖 5)，會將距離觀者最近的測量卷尺往上延伸，當卷尺達到約 3 公尺高時，就會墜落並向後彈回，且每一個小時，系統會印出所有訪客所花費的總共時數。[20]

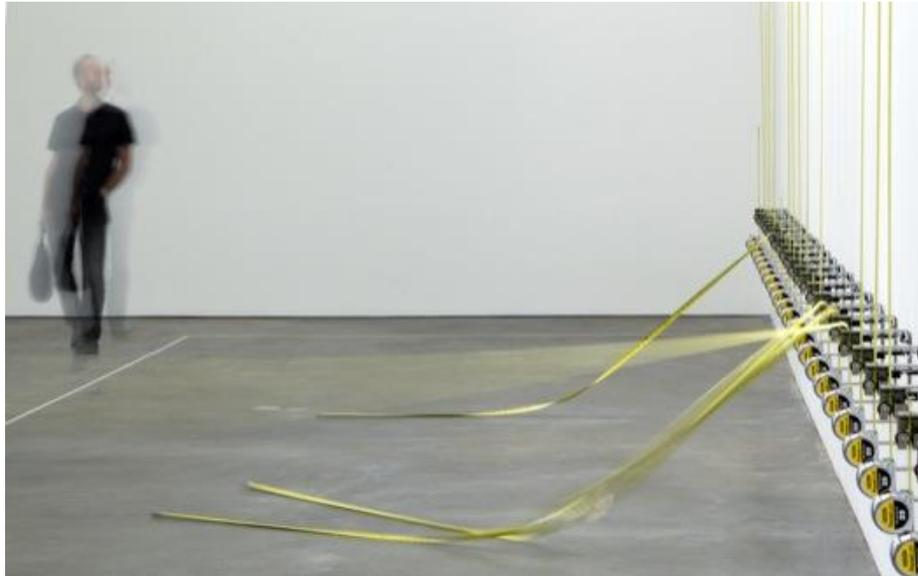


圖 4：Rafael Lozano-Hemmer，《Tape Recorders》。2011。

圖片來源：[https://www.lozano-hemmer.com/artworks/tape\\_recorders.php](https://www.lozano-hemmer.com/artworks/tape_recorders.php)

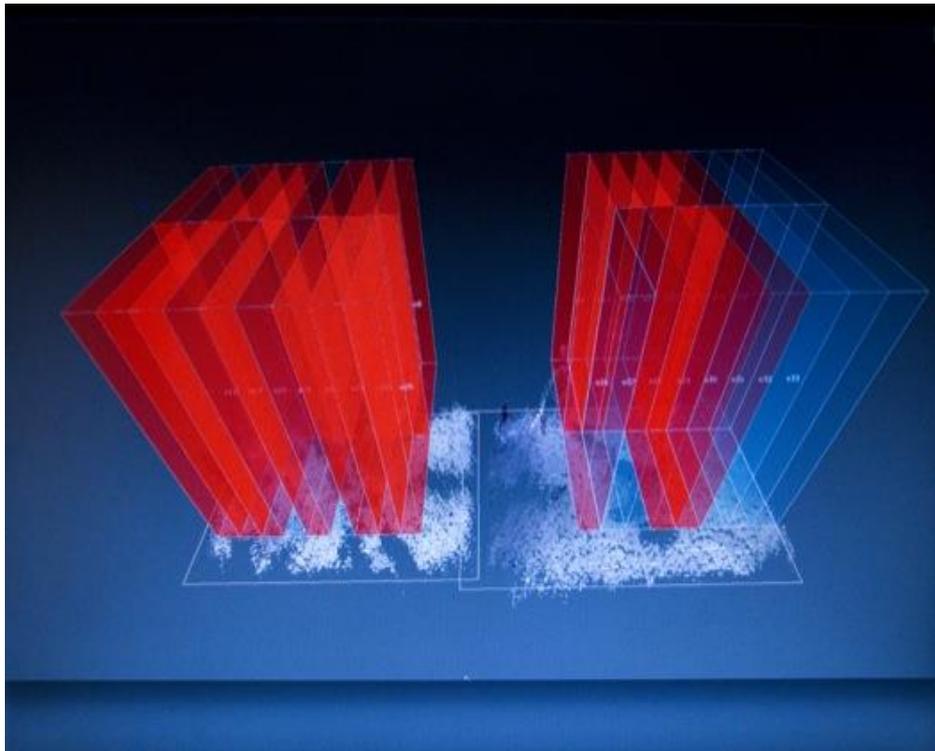


圖 5：Rafael Lozano-Hemmer，《Tape Recorders》程式偵測畫面。2011。

圖片來源：[https://www.lozano-hemmer.com/artworks/tape\\_recorders.php](https://www.lozano-hemmer.com/artworks/tape_recorders.php)

## 二、Google ARCore

### (一)、Scene Semantic

Semantic Segmentation 為語意分割，其目的是將影像中的所有像素點進行分類，可用於智慧機器人、自駕車、車用輔助系統、擴增實境等。現今的 Semantic Segmentation 有多種辨識方法，如 Pixel-Level Semantic Segmentation、Scene Semantic Segmentation 等。

本論文使用的是 Scene Semantic 為場景語義分割，是指對 2D 影像或 3D 場景中的物體、人物等不同類型的元素進行分類，並為不同類型的元素賦予相應的語義標籤或描述，其特點為更加關注場景的整體理解和上下文的訊息。由於近年來深度學習的興起，Scene Semantic 的成果也取得了巨大的突破。[21]，其研究方法多為使用 Convolutional Neural Network (CNNs) 達到影像分類及物件辨識，並透過常見模型框架，如 Fully Convolutional Network (FCN) 用來端對端 (End-to-end) 的模型訓練，可有效的提高邊緣細化的精準度[22]。

Google 公司所推出的 ARCore 近年也提供 Scene Semantics API 功能，讓開發者能了解使用者所處的環境，並提供標籤結果。其所提供的標籤可信度會根據物體面積越大及物體越常見於生活中而提高，辨別結果也更具可靠性。以下為辨識結果高到低的排名。[23]

表 1：Scene Semantics 辨識物件

辨識可信度	辨識標籤
主要場景組	1. 天空 2. 建築物 3. 樹木 4. 道路 5. 車輛
主要場景細節	1. 人行道 2. 地形 3. 結構 4. 水
次要場景細節	1. 物體

近年，為了能對 2D 影像或 3D 場景中的元素進行分類，機器學習在這塊上可說是功不可沒，為了能讓模型獲得正確的分類結果，開發人員需要對於訓練圖像加上標籤或描述，然而在貼上標籤的過程中，究竟是客觀結果，還是主觀意識？

AI 研究員凱特·克勞福德 (Kate Crawford) 及藝術家特雷弗·帕格倫 (Trevor Paglen) 在 2019 年提出《ImageNet Roulette》藝術計畫，該作品是透過大型視覺資料庫 ImageNet 當中類別為「人類」的資料集所延伸的攝影展，研究員及藝術家發現在 ImageNet 影像數據庫中，喝啤酒的年輕人被標籤為「酒鬼、酗酒者」，而戴墨鏡的孩子被歸類為「失敗者、不成功的人」，對於穿著清涼及身材姣好的女性，更是被標籤為女魔頭(ball-breaker)。因此意識到當研究人員替這些影像加上標籤的同時，不知不覺也加入了系統性的偏見、意識形態在其中，這樣的數據庫的建構和分析，在未來極有可能成為有害的分類依據，除了會降低社會對於事物的多元性外，更加深了歧視的可能性。[24]

況且，回顧藝術歷史，藝術家們不斷地展示「圖像不能描述其自身」的特點，例如比利時超現實主義畫家雷內·馬格利特(René Magritte)於 1928 年至 1929 年的著名創作《形象的叛逆》(La trahison des images) (圖 6)，作品繪製了一支菸斗，但底下的文字卻寫著這不是一支菸斗，因為圖中的煙斗無法拿來使用，讓圖像與物件之間形成矛盾，以及約瑟夫·科蘇斯 (Joseph Kosuth) 在 1965 年的裝置藝術作品，《一把和三把椅子》(One and Three Chairs) (圖 7)，作品中擺放真實的椅子，以及椅子的放大照和從字典上影印有關椅子的文字，才讓世人顛覆了原有的「定義」，試圖讓是人們不同的方式看待物件，然而影像的標籤，反而又讓一切回歸到最簡化、直接的思考。



圖 6：René Magritte，《形象的叛逆》（La trahison des images）。1928-1929。

圖片來源：<https://zh.wikipedia.org/zh-tw/%E5%BD%A2%E8%B1%A1%E7%9A%84%E5%8F%9B%E9%80%86#/media/File:MagrittePipe.jpg>



圖 7：Joseph Kosuth，《一把和三把椅子》（One and Three Chairs）。1965。

圖片來源：[https://zh.wikipedia.org/zh-tw/%E4%B8%80%E6%8A%8A%E5%92%8C%E4%B8%89%E6%8A%8A%E6%A4%85%E5%AD%90#/media/File:One\\_and\\_Three\\_Chair.jpg](https://zh.wikipedia.org/zh-tw/%E4%B8%80%E6%8A%8A%E5%92%8C%E4%B8%89%E6%8A%8A%E6%A4%85%E5%AD%90#/media/File:One_and_Three_Chair.jpg)

## (二)、Geospatial

ARCore 為 Google 公司所支援，隨著手機在世界中移動(如圖 8)，ARCore 會使用「即時定位與地圖建構」功能(或稱 SLAM)判斷手機與周遭世界的相對位置。ARCore 會在拍攝的相機影像中，偵測出稱為「地圖項目」的視覺獨特特徵，並使用這些點來計算其位置的變化。視覺資訊結合了裝置 IMU 的儀器測量數據，以便隨時間估計相機相對於世界的姿勢(位置和方向)。

藉由對齊虛擬相機的姿勢，讓虛擬相機以 ARCore 提供的裝置相機姿勢轉譯 3D 內容，開發人員就能從正確的視角算繪虛擬內容。轉譯的虛擬圖片可以重疊在從裝置相機取得的影像上，看起來就像虛擬內容是真實世界的一部分。[25]



圖 8：ARCore 操作指引

圖片來源：

<https://developers.google.com/static/ar/images/MotionTracking.jpg?hl=zh-tw>

ARCore 同時還支援深度理解的功能，使用支援裝置的主要 RGB 相機，建立深度地圖，當中包含表面與指定點之間的距離資料。使用者可以使用深度地圖提供的資訊，提供身歷其境的沉浸式使用者體驗，例如讓虛擬物件與觀察到的表面精確融合，或讓虛擬物件顯示在實際物體前方或後方。[25]

本論文使用的技術著重在 ARCore Geospatial API 功能，此功能可將虛擬內容附加到 Google 街景服務所涵蓋的任何區域，打造一個全球規模的 AR 體驗。當使用這項功能時，會使用裝置感應器和 GPS 資料偵測裝置環境，然後將該環境中可

辨識的部分與 Google 視覺定位系統 (VPS) 提供的本地化模型進行比對，判斷使用者裝置的精確位置。此外，這個 API 也會將使用者的本機座標與 VPS 的地理座標合併，以便在單一座標系統中運作。

VPS 是 Google 地圖的街景服務圖像，也是 VPS 的基礎。透過深層類神經網路能識別並描述長時間可能辨識出的組成部分。接著，系統會將這些零件合併至數十億張圖像，以運算出全球環境的 3D 點雲(3D Point Cloud)。這個本地化模型包含數兆個點數，並涵蓋幾乎所有國家及地區。

使用者的裝置向地理空間 API 發出要求時，類神經網路會處理像素，找出使用者環境中可辨識的部分，並將這些部分與 VPS 本地化模型進行比對。接著，電腦視覺演算法會計算裝置的位置和方向，提供的位置比單純使用 GPS 時更為準確的結果。

類似的 AR 作品如 2023 年 Google 和藝術家卡米爾·瓦拉拉 (Camille Walala) 的合作(圖 9)，米爾·瓦拉拉在倫敦東部的肖爾迪奇以 AR 的方式創作了一幅 AR 壁畫，將她自身帶有的標誌性的大膽而充滿活力的風格融入在倫敦街頭。



圖 9：藝術家卡米爾·瓦拉拉 (Camille Walala) 的 AR 壁畫

圖片來源：[https://artsandculture.google.com/story/4QWh1epJ2D\\_bJQ](https://artsandculture.google.com/story/4QWh1epJ2D_bJQ)

2019 年 Apple 蘋果也利用自家的 ARKit 功能與多位藝術家合作，當中包含 Pipilotti Rist 的 International Liquid Finger Prayer(圖 10)，此作品邀請觀者在全球城市中追逐閃閃發光的形式。



圖 10：2019 年 Pipilotti Rist 的 International Liquid Finger Prayer

圖片來源：<https://www.apple.com/sn/newsroom/2019/07/apple-offers-new-augmented-reality-art-sessions/>

另外，若以單以 AR 效果的藝術創作來說，如 Adrien M & Claire B 在 2019 年的 Acqua Alta(圖 11)，此件作品是藉由 AR 述說一個故事。故事內容為一個女人、一個男人和一棟房子，他們的日常生活，充滿著荒謬且矛盾，然在一個潮濕的雨天，他們的生活發生了翻天覆地的變化。作品是用一個災難故事，欲傳達迷失和尋找的過程。



圖 11：Adrien M & Claire B 的 Acqua Alta

圖片來源：<https://www.am-cb.net/en/projets/acqua-alta>

### 三、相機校準

相機校準(Camera Calibration)是指由於照相機及投影機都需要藉由透鏡來蒐集足夠的光線，才能夠成像，但是透鏡會造成物理上的透鏡畸變的問題，使得影像中間及四周會有扭曲的狀況，因此會需要透過相機校準來修正影像的偏差。本作品是透過 Touch Designer 的 Cam Schnappr 完成相機校準的動作，其原理則是使用 OpenCV 的 Calibrate 功能，透過帶入 3D 物件及 2D 影像的相同特徵點進行運算。

[26]

## 第三章 作品論述與執行

### 第一節 高敏感的聽說讀寫練習

#### 一、創作理念

「高敏感」在過去幾年間，曾一度為流行的討論議題，訪間更出版了眾多帶領大眾了解高敏感及如何面對此性格特質的書籍，例如《高敏感是種天賦》、《高敏人的職場放鬆課》...等，此概念是美國精神分析專家伊蓮·艾融(Elaine N. Aron)提出，主要是指五感相較於他人特別敏感，然而對於事物的認知，在社會時空不同的背景下，也有不同的理解，過去或許是社會偏見的影響，許多內向者或敏感者會都會被聯想為「壓抑的」、「靦腆害羞的」、「神經質的」，甚至不少人為自尊心低落所苦。[3]

高敏感的人容易累積心理上的負擔，經常會卡在自我否定及懷疑的過程，而使得最後精疲力盡，由於高敏感族的神經敏感纖細，所以比較容易受困於此種狀況。且現今社會高度讚揚外向特質的社會文化，使得高敏感的人會不自覺得感到備感壓力、壓抑。同時，高敏感族對於痛覺的忍態度較低，卻又熱衷於探索自己的內在世界，所以每當問題發生時，往往會糾結過頭、鑽牛角尖無法收手。

近年來，筆者發現自己此項特質越來越強烈，因此希望以這個概念作為創作的起點，作品的概念的是以刻意將言語的力量，透過文字及影像，以投射在空間中呈現，目的是讓觀者能理解使用者傳達的訊息意圖，但卻無法完全感受到原始語句的意涵，無法逐一推敲每個語句和用字的含義，並將文字呈現在空間中。這個過程就像一個抽象的心靈建築過程，展現了內心敏感的經歷給觀者看。語句的解讀和接受能承受的訊息量，以及麻痺自身的感受能力，重新映照所接受的訊息。這樣的視覺展示可能充滿直接而暴力的元素，就像一種象徵，避免高敏感個體內心世界的複雜性。

這個裝置背後隱含著一個強烈的主題——對言語的過度分析。高敏感的個體

常常深入思考言語的含義，並尋找情感共鳴。然而，這種過度分析可能導致情感的放大，使每個字詞都變得極具深意，並且每個字詞背後都隱藏著動機，由於高敏感的個體常常陷入內耗的循環，這種內耗源於對外界刺激的強烈敏感，他們容易將周遭的感知細節深入思考，試圖從中獲取更多的信息和情感，這種過度分析使得他們經常反覆思考每個詞句的可能含義，甚至揣摩說話者的意圖。這個裝置試圖捕捉這種過程，並讓高敏感的個體練習減弱這些行為，它將這些行為展示給觀者，通過互動的過程，嘗試捕捉不可得的文字，而此文字是藉由另一位觀者透過麥克風所生成，並投影在空間中，希望以文字賦予空間新的語意，讓觀者與在文字所產生的語意下互動，試圖將場景語意具像地呈現出來，讓觀者在空間場景下的互動過程中嘗試理解因文字而生成訊息。

總體而言，此作品為高敏感個體的內心感受為創作出發，情感內耗為主題，透過科技的媒介呈現了語言與肢體之間的互動。通過這個作品，希望能傳達並使觀者得以深入了解高敏感個體內心世界的複雜性，並思考情感內耗對於人際關係和自我成長的影響。

## 二、創作方法與技巧

### (一)、硬體架構說明

本作品在展場上的硬體擺設示意(圖 12)，該作品展示於地美館入口處，一共有 4 座展架，作品入口處放有擺放麥克風的裝置，此麥克風連接到 1 台筆記型電腦，此電腦同時連接 Arduino 的伺服馬達，控制油煙機的開關時間。

空間中同時擺放 3 座展台，此 3 座展台上皆有 1 台投影機及 1 台 WebCam，該投影機及 WebCam 同時連接到同 1 台筆記型電腦，為了讓 1 台筆記型電腦能夠傳送不同的螢幕畫面到 3 台投影機，選擇使用螢幕分配器，讓螢幕畫面可以輸出至指定的投影機，以避免 WebCam 的偵測結果和搭配的投影機畫面不一致，確保 WebCam 所偵測的結果可以和投影機畫面相同。

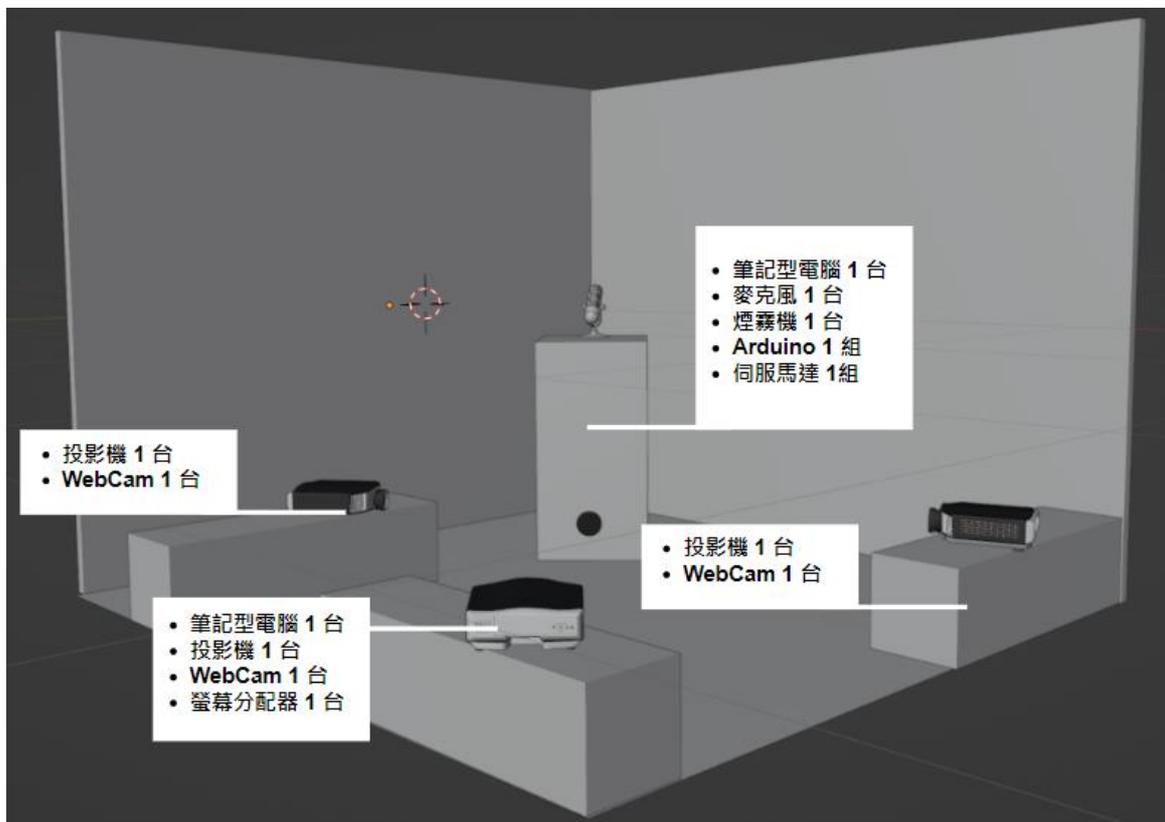


圖 12：〈高敏感的聽說讀寫練習〉硬體擺設

表 2：〈高敏感的聽說讀寫練習〉作品硬體清單

硬體名稱	規格/型號	數量
筆記型電腦	Windows	2
WebCam	Logitech C270 HD WEBCAM HD 720p	3
麥克風	SHURE VP83F 槍型指向錄音麥克風	1 支
短焦投影機	Epson EB-530	3 台
煙霧機	Z-380 Fazer 薄型煙機	1 台
螢幕分配器	1 進 4 出	1 台
Arduino UNO、伺服馬達		各 1

## (二)、軟體架構說明

本作品在軟體的部分一共分為 3 個部分

### 1. 語音辨識：

利用 Python 匯入 speech\_recognition 套件，撰寫開起麥克風及接收音源檔，當接收到音源時，會透過語音辨識功能，將音源轉為文字，並在該台電腦架設 Server，若完成語音辨識功能後，即會將文字以 TCP/IP 的方式傳送至另一台電腦。

### 2. 相機校準：

先開啟 WebCam，取得 WebCam 拍攝到的位置，並測量實體環境中，拍攝範圍的長寬、比例，於拍攝範圍內的實體環境上，用膠帶貼上特徵點(如圖 13)，利用 Blender 建出與拍攝範圍相同的長寬比模型(如圖 14)，將模型匯入 Touch Designer，開啟其軟體內的 Cam Schnappr 的功能，先讓投影機投影至實體空間中，依照特徵點位置，點擊在 Touch Designer 內的 Cam Schnappr 的相對位置，點擊超過 6 個點後，Touch Designer 內的 Cam Schnappr 就會模擬校準結果(如圖 15)，該結果則為 WebCam 實際的位置(如圖 16)[27]。



圖 13：實體環境的特徵點

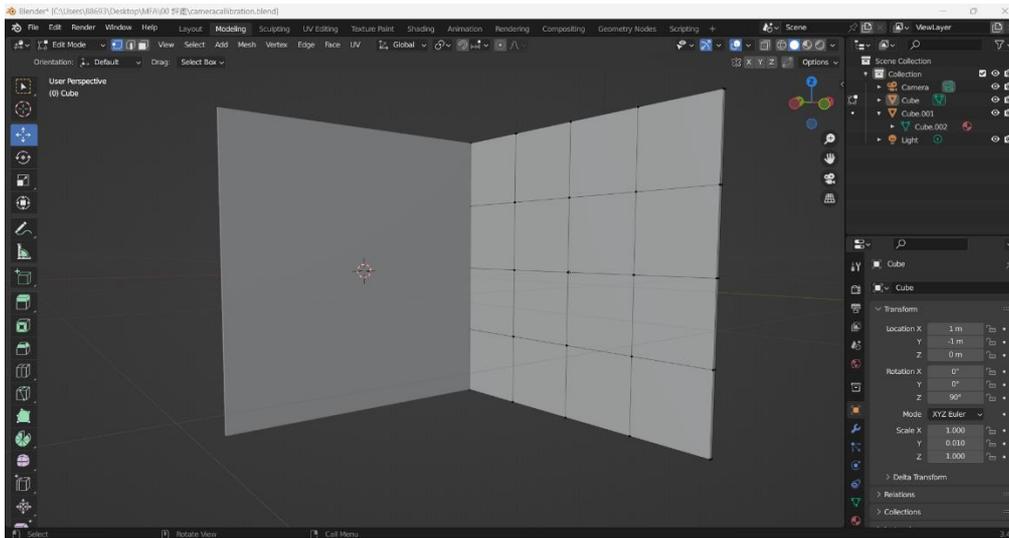


圖 14：Blender 內的模型

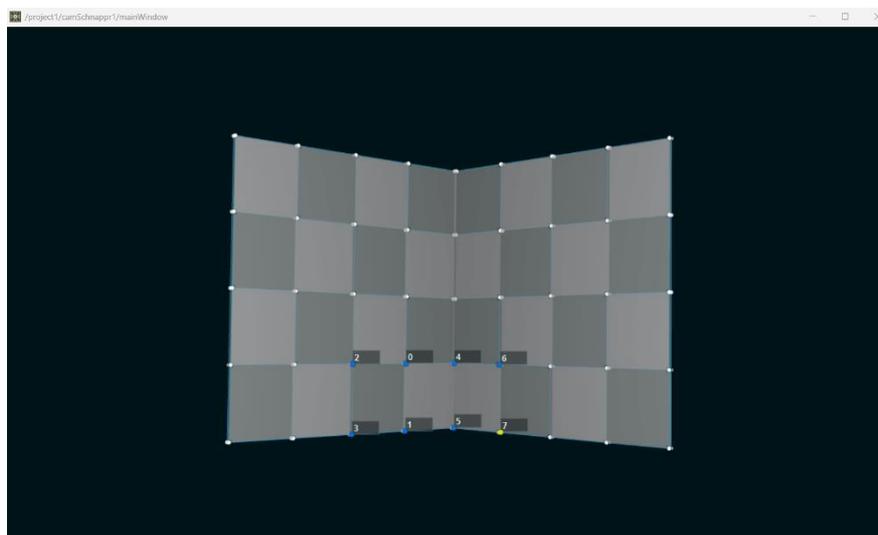


圖 15：Touch Designer 內的 Cam Schnappr 模擬校準結果

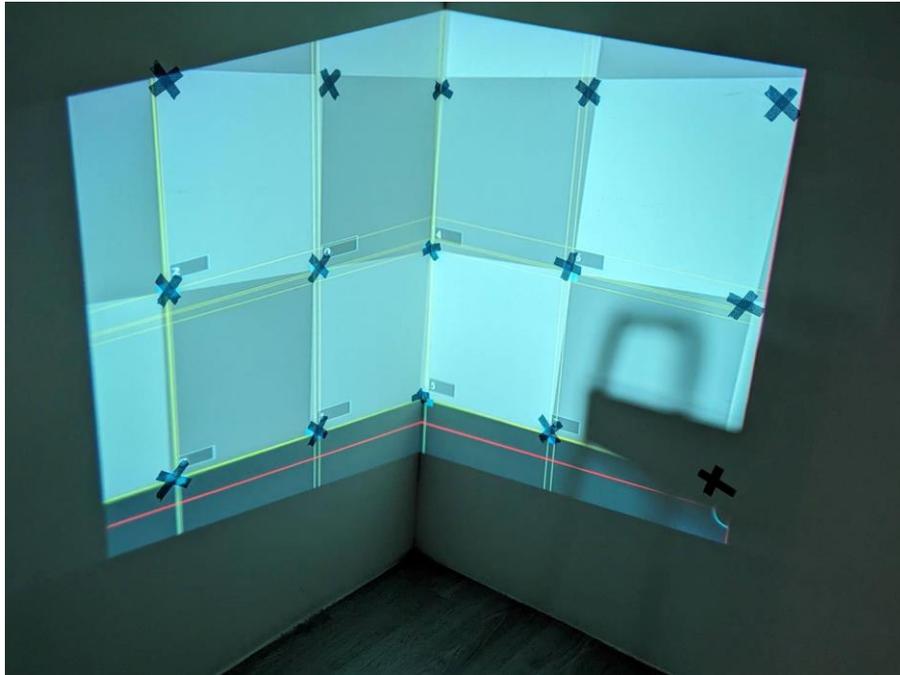


圖 16：實體投影的對位結果

### 3. 人體辨識：

利用 Touch Designer 匯入 OpenCV，並在 Touch Designer 內撰寫 Python 腳本，將人體辨識結果二值化，以取得黑白的影像，並計算出黑色區域中最小的 X 位置及最大的 X 位置，此 2 個數值則為人體辨識的最左側及最右側區域，最終將文字訊息的顯示位置設定為此 2 個數值，即完成跟隨人體移動的功能。

### 4. 流程圖：

本作品共分為 2 部分軟體整合(如圖 17)，第 1 個部分為透過 Python 去偵測麥克風是否被開啟，若被開啟後則會執行 Speech to Text 功能，並將辨識後的文字傳送給 Server，Server 會再將文字傳送給 Touch Designer。

第 2 個部分則先使用 Webcam 及 Touch Designer 依據環境做相機校正，取得 Webcam 和投影機之間的空間關係，讓 Webcam 的照射範圍和投影機的投影範圍相符合，以利後續能精準的將文字投影在人的兩旁，並在 Touch Designer 內執行 OpenCV 的人體辨識功能，並計算人體的 X 最大值及 X 最小

值作為投影位置，當 Client 接受到 Server 訊號傳送的文字時，則將文字投影在空間中。

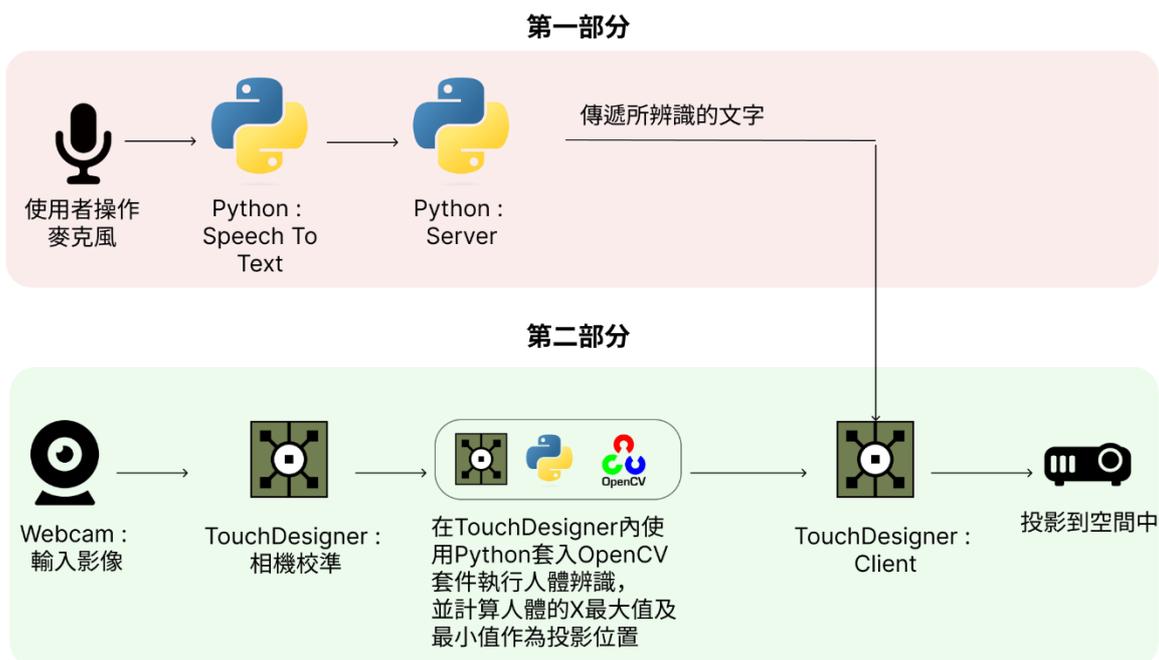


圖 17：〈高敏感的聽說讀寫練習〉軟體系統架構圖

在軟體流程圖可分為 Server 端(圖 18)及 Client 端(圖 19)，Server 端若偵測到麥克風有聲音訊號時，會將聲音存取下來，等待無訊號時，會將轉成音訊檔並儲存至本機端，執行 speech\_recognition 套件讀取該音訊檔，等待辨識完畢後會傳送到 Arduino UNO 旋轉伺服馬達按壓煙霧機開關，並將文字傳送至 Client 端。

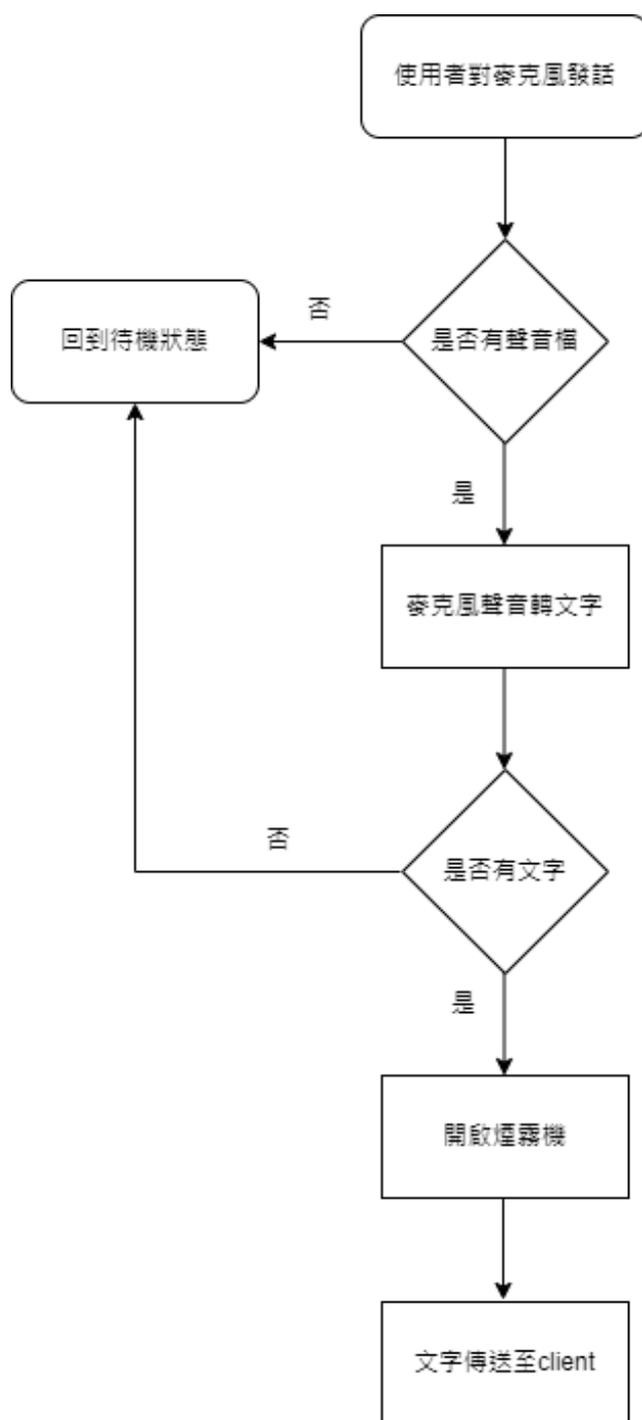


圖 18：〈高敏感的聽說讀寫練習〉Server 端軟體流程圖

而 Client 端(圖 19)與 Server 連線後，會偵測是否辨識到人體，若有則會確認是否有文字訊息，則會將文字訊息投影在人體位置，若人移開攝影機範圍後，則回到待機模式。

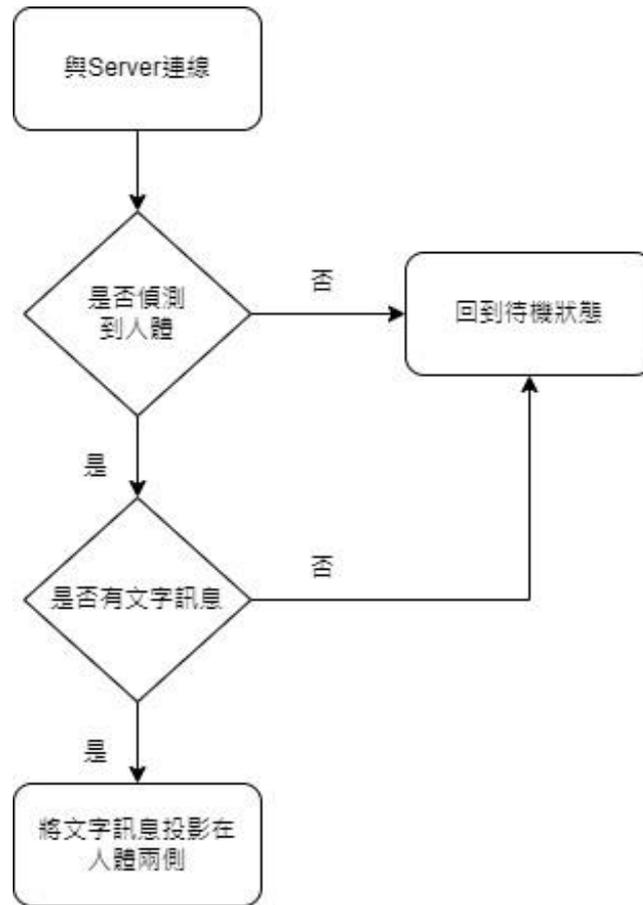


圖 19：〈高敏感的聽說讀寫練習〉Client 端 軟體流程圖

### (三)、互動設計

此裝置的運作機制主要為觀者 A 透過麥克風對話，該麥克風會連接至 1 台電腦(Windows 作業系統)，若當觀者 A 對麥克風發話後，裝置內的展架會開啟煙霧噴霧，會觸發此台電腦事先撰寫好的即時語音辨識功能，將觀者 A 所敘述的內容，轉換為文字訊息，該台電腦中會架設好 Server，並透過 TCP/IP 傳送到相同通訊網路的其他台電腦。

另一台電腦則利用 Touch Designer 作為 Client，接受到來自 Server 的文字訊息後，將文字刻意模糊化後利用投影機投射出來。另外，筆者先將另一台電腦連接 3 組 WebCam，並預先利用 Touch Designer 裡的校準工具 Cam Schnappr 進行相

機校準作業，讓實際場域的環境空間與投影機投射出來的畫面是相符的，即所謂為的光雕投影 Projection Mapping，讓其他的觀者可在限定的空間內自行移動，移動的過程中，電腦皆會利用 OpenCV 辨識 WebCam 所捕捉到的鏡頭畫面，並進行影像辨識，辨識觀者們在影像中的像素位置，將文字顯示於影像的像素上，於實際空間中，則會透過投影機投射至觀者們的四周，形成一堵又一堵的光牆，讓觀者們感受到因他人的言語所給予的壓迫。

筆者刻意利用 Touch Designer 將接受的文字模糊化，讓觀者們因無法輕易辨識文字訊息，筆者試想或許觀者們會利用手去捕捉文字，亦或是讓身體作為投射布幕，藉由將文字打在自己的身上去解讀、剖析愈傳遞的訊息，然而，在觀者的身體去捕捉的過程中，影像辨識功能又會不斷的更新觀者的最新位置，使得觀者一直在愈求得、愈求不得的狀態中。

### 三、呈現與紀錄

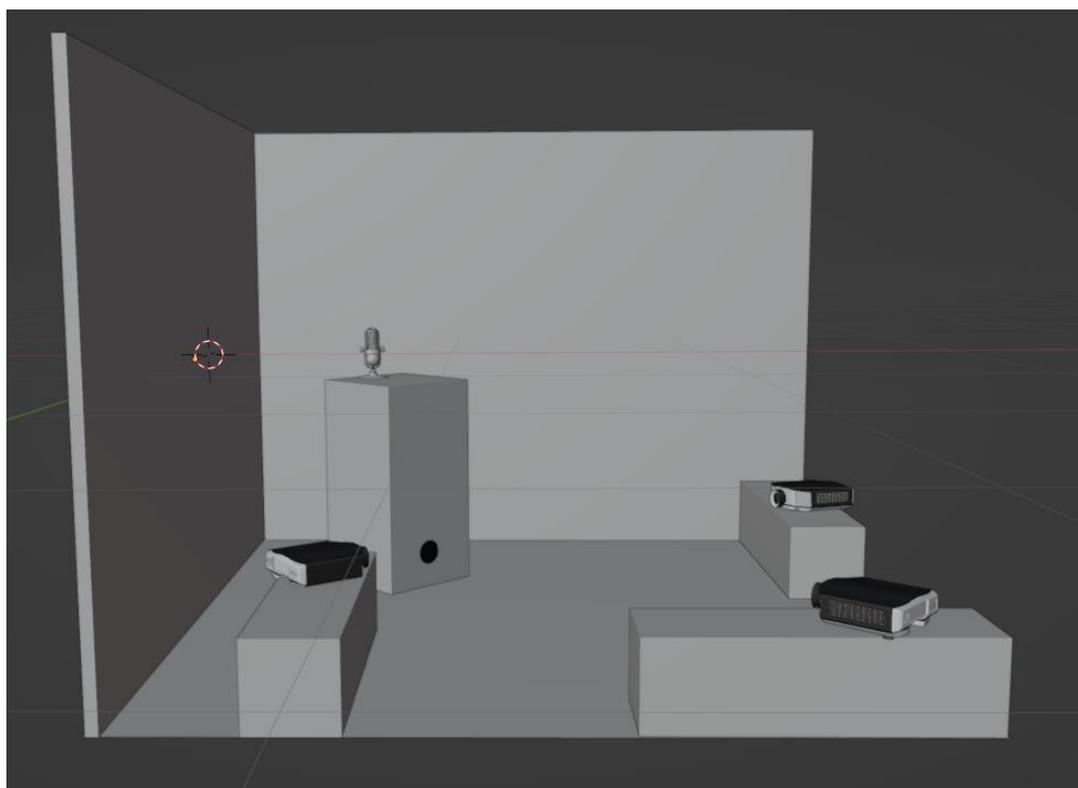


圖 20：〈高敏感的聽說讀寫練習〉空間配置圖

表 3：〈高敏感的聽說讀寫練習〉展出紀錄

步驟	紀錄
<p>當觀者走至展架 1，可看見展架上的麥克風，若對著麥克風說話時，即可開啟語音辨識功能，下方的煙霧機即會開啟噴霧。</p>	
<p>其他觀者走至展架 2 時，可看見從投影機打出的光、文字。</p>	
<p>其他觀者走至展架 3 時，可看見從投影機打出的光、文字。</p>	

## 第二節 Hello World

### 一、創作理念

「Hello World」這串文字已經成為了學習程式的傳統，常被用來程式輸出、測試時，嘗試印出的字串。這個簡單的語句在印出「Hello World」的同時，象徵著一個開始，宛如向世界宣告著「我來到這個世界了」，即將開啟探索世界的旅程。

在我們探究這個世界的同時，在無形中都會間接、直接的被社會影響(Social Influence)，進而產生各種想法、感受及行為，張淑美 2002[9]提到在認知心理學中，當我們接收到感官的刺激後，促發我們對於感受到的訊息作解釋，也就是所謂的知覺(Perception)，進一步再到認知(Cognition)，認知是將感觀輸入的訊息與其他訊息，與過去累積的知識交互作用，做一種主動的結合與解析，產生適切意識，象徵的解釋(曾啟雄、孫志誠，1999)[28]。而筆者在製作此作品時，發覺這段從感官刺激到知覺，再從知覺進行分析直到獲得認知經驗的歷程，與場景語意的辨識歷程有著相似的概念，場景語意是先從匯入影像後，在影像辨識上對影像中的物件及區塊有所認識後，透過訓練好的模型獲得辨識結果，再將辨識結果對影像逐一標籤化，並直接利用場景語意做為此作品的重要技術。

從眼睛到視覺受到環境影響的同時，環境中物件的位置、面積，你的意識專注的觀點，會主宰著你的世界。專注除了物理上的觀看外，還有著心理層面的投注，此作品希望藉由將兩者融合到作品之中，想要表達我們的專注點影響我們對世界的認知，像是恐懼、焦慮是否真的存在？也許存在，而且是我們讓它們存在，也許不存在，它們只是我們對未知事物的另一種詮釋。

有趣的是，即便我們的視覺接收到刺激，也可能發生「盲視」現象，例如觀察者無法察覺到視覺場景中顯著的變化。因此，筆者試圖用同樣的文字，放置在不同的位置，一來是挑戰觀者是否察覺到相同的文字，二來是意圖從視覺動線的角度來影響觀者對於語意的解讀。

而筆者有感於自身在製作第一件作品〈高敏感的聽說讀寫練習〉的歷程時，對於人際互動、事物的見解還是停留在人及物的本身，像是一池清澈的水常因外在環境的變化而映照不同的倒影，然而隨著創作的過程中，也許是自身的心態改變，開始退一步去看待這些外在影響，從原本的將自己的視野困在他人的目光及言語中，回頭去設想究竟是什麼影響了我們的思考模式，反而去思考若自身改變觀看的角度，會不會得到不同之見、不同之得，所以才以此期盼讓觀者感受到看著相同的景色，卻又有截然不同的解讀。

最後文字愈傳達出生命就像植物，而人生的起落就像天氣現象，時而晴天、時而雨天，我們唯一能做的是在天氣交替時，努力適應環境，並透過心態改變我們對外部環境的認知，進而影響環境本身。

## 二、創作方法與技巧

由於過去的 AR 體驗多在限定的 2D 圖案、實體物件及人物上，因此此作品是利用 ARCore Scene Semantics 機器學習辨識影像後，將符合辨識結果的像素加上標籤(如天空、建築物、樹、道路、人行道、車輛、人物等)，並且透過 Unity Shader 將特定的標籤結果添加其他影像，達到虛擬影像與現實疊合的效果。由於辨識的項目為天空、人物、水、樹等自然景色，因此 ARCore Scene Semantics 的 AR 效果在操作上並不需要另外準備 2D 圖案或實體物件，且當使用者在行走的過程中透過相機鏡頭隨時讀取影像，即時做 AR 效果的變化。

另外，本研究的 AR 體驗同時還利用 ARCore Geospatial，追蹤使用者的緯度、經度和高度，透過遊戲 NPC 引導使用者走到指定地點後，虛擬物件會在限定的現實空間中生成，讓使用者除了環境中的景色(如天空、人物、水、樹)外，對於空間也有不同的體驗。

互動過程中，為了讓使用者提高在遊戲世界中的自由度，且更接近真實體驗，使用者可自由與 NPC 對話，對話內容不受限制，同時為了讓 NPC 的回應能和使用者的對話呼應，因此 NPC 的對話是由 ChatGPT 即時生成，每次呼叫 ChatGPT 的

Text Generation[29] 功能時，本作品會以 System 的角色告知 ChatGPT 角色任務及扮演的 NPC 性格，讓 ChatGPT 能夠有效引導使用者，並給予正確提示。

最後，為了增加使用者在體驗 AR 的互動性及便利性，過程中使用者主要是以聽、說的方式互動，讓使用者的視覺專注在 AR 生成效果，降低視覺負擔、手部的操作。因此使用者的對話操作方式為錄音，並為了能將語音顯示在螢幕中，以及讓語音內容能傳送至 ChatGPT，則透過 OpenAI 的 Create Audio Transcription[30] 功能將使用者的錄音檔轉為文字(Speech to Text)，同時為了讓使用者無須閱讀 NPC 的回覆，可透過聽覺的方式對話，因此利用 Amazon Polly[31] 將 ChatGPT 生成的回應以文字轉語音(Text to Speech)的方式生成音檔並播放。

#### (一)、環境與視覺規劃

此作品會辨識真實環境中的天空、人物、水及樹木，以及使用者必須到指定的地點 228 公園內才能完整體驗該作品。由於筆者在 228 公園的場域內注意到，當觀者不刻意注視特定環境的情況下，湖面的位置多在視覺的下方、樹木則是在視覺的上方(圖 21)。

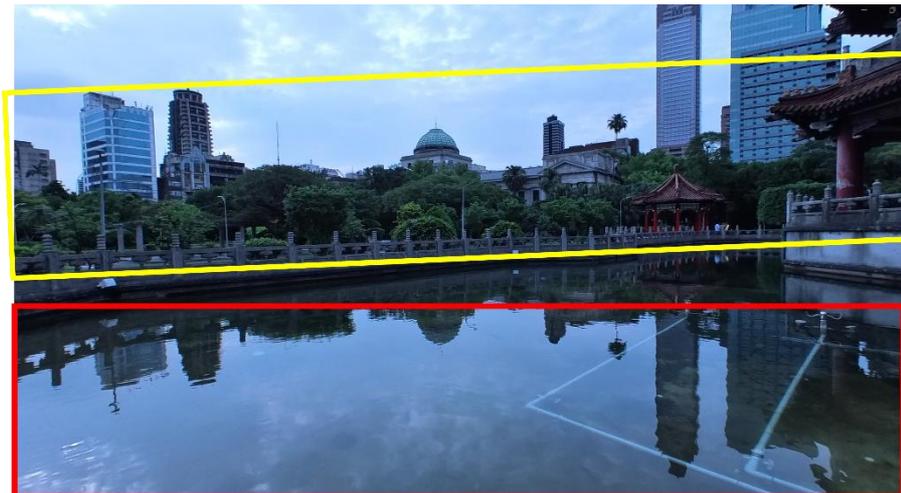


圖 21：紅色區域為湖面、黃色區域為樹木

因此筆者試圖將文字生成在特定物件上，在關卡 1 需要辨識的物件為湖面，預設使用者剛開始看到湖面時，湖面會佔據螢幕下方 1/3 之的位置(圖 22)，因此可看到的文字為「晨、幕、臨」，當使用者走近湖面後(圖 23)，湖面會佔據螢幕的全部，因此可看見的文字為「降、光、逝、消、夜、晨、幕、臨」，由於先看到的文

字為「晨、幕、臨」，所以預設使用者對字句的完整解讀為「晨光消逝，夜幕降臨。」。

當使用者來到關卡 2 時，辨識的物件改為樹木，由於樹木多為視覺的上方位置，因此預設當使用者站在遠處時，樹木只占據螢幕中間 1/3 位置，因此可看到的數字是「夜、迎、來」，當使用者走近樹木後(圖 23)，樹木佔據螢幕上方 2/3 位置，手機則可看到「夜幕降臨，迎來清晨」，筆者欲傳達的意涵為「夜幕降臨，迎來清晨」。

當使用者走到關卡 3 時，則脫離了文字與視覺的框架，而是會在 228 公園內的舞台生成一顆黑色的圓球，圓球內有一盞燈，走進球內會觸發多種不同的人臉面貌，但其實什麼也沒有，如同他人的目光所帶來的恐懼及意念，看似存在但也不存在。

關卡主要傳達的是即使在相同的文字下，因為使用者的視覺角度、所站的位置的不同，影響了可觀看的範圍，試圖影響使用者對於文字的理解。就此作品的文字為「晨光消逝，夜幕降臨」和「夜幕降臨，迎來清晨」，想要告訴使用者黑夜與白天的交替，是人生的如常，可以懷抱著害怕與恐懼去看待黑暗的到來，也可以持著希望去等待黎明的來臨，而人之所以踱步不前，正是因為無法面對未知的恐懼，但恐懼的背後是什麼？只是自己的意念罷了。



圖 22：當使用者站在較遠距離，樹木面積較小，可看到的文字範圍也較小



圖 23：當使用者走近樹後，隨著樹木在視覺上的面積變大，視覺可閱讀的文字範圍也越多

## (二)、硬體架構說明

本作品須使用 Android 系統手機為載體，且須具備以下規格：

表 4：〈Hello World〉硬體規格表

裝置型號	Android 版本	Min SDK 版本	Depth API
Any x86 、 x86_64 based AVD	Android 7.0 以上 版本	≥ 24	必須支援 Depth API

### (三)、軟體架構說明

#### 1. Unity

本作品以 Unity2021.3.10f1 版本進行開發。

#### 2. Google ARCore Scene Semantic

為了取得影像中的環境物件，透過 Google ARCore Scene Semantic 獲得機器學習辨識後的結果，取得目前影像上的資訊，得到每個像素的標籤，為了避免增加 CPU 的負載，將特定的標籤結果透過 Unity 的 Shader 做影像處理，讓影像的計算及替換由 GPU 執行，依照標籤不同，更換不同貼圖，此研究選用環境中常見的天空(圖 24a)、人物(圖 24b)、樹木(圖 24c)、水(圖 24d)作為 AR 效果，讓使用者能快速在現實環境中得到視覺回饋。

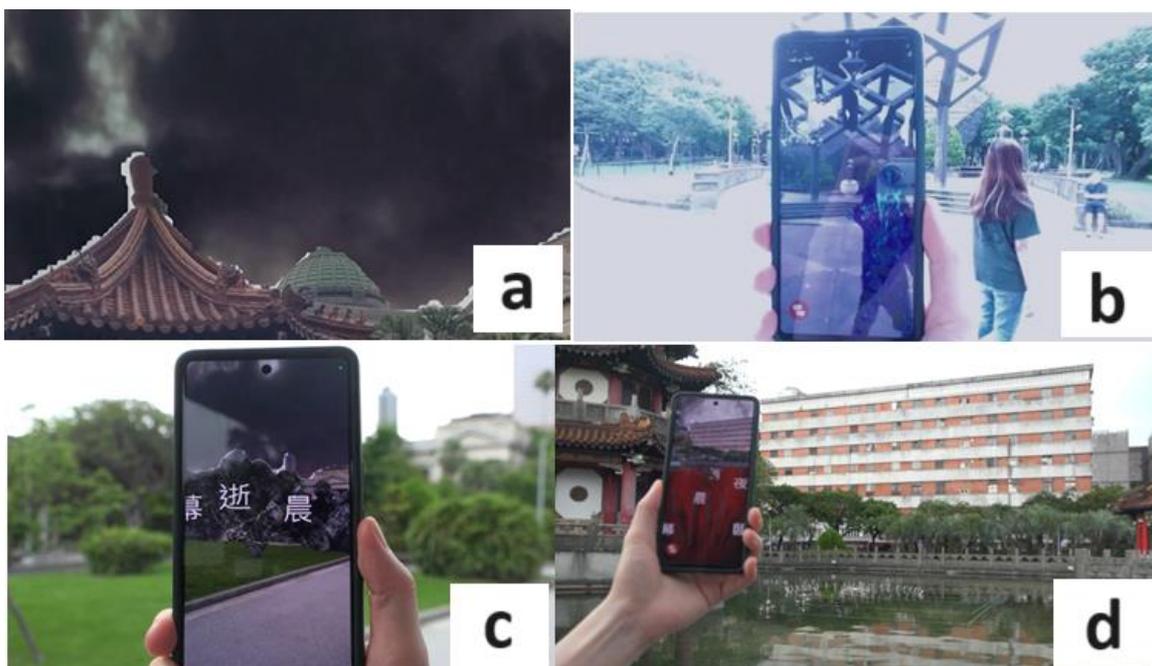


圖 24：ARCore Scene Semantic 效果

### 3. Google ARCore Geospatial

在 Unity 裡匯入 arccore-unity-extensions 插件，做為 Unity AR Foundation 跨平台擴增實境的 API，透過手機開啟 GPS 定位系統，將得知使用者位於筆者設定的地理位置，並開啟攝影機，將環境中可辨識的部分與 Google 視覺定位系統 (VPS) 提供的本地化模型進行比對，判斷使用者裝置的精確位置，以達到將 3D 模型放置於精確街景中。此外，此 API 也會將使用者的本機座標與 VPS 的地理座標合併，以便在單一座標系統中運作。

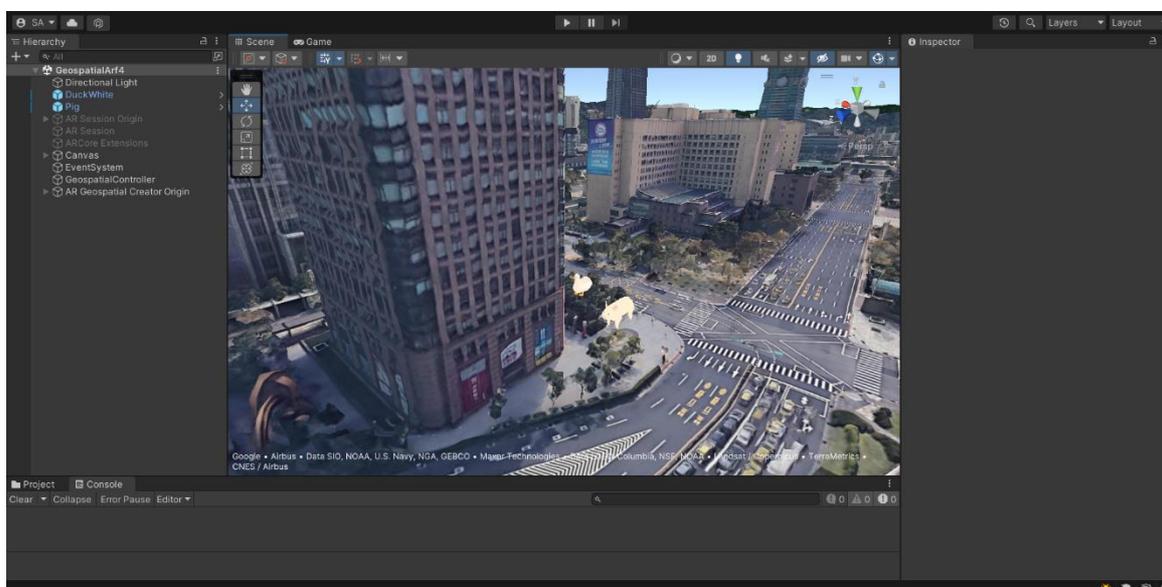


圖 25： ARCore 匯入 Unity 內的地理位置資訊



圖 26：ARCore 實際地理位置的 AR 結果

#### 4. Cesium for Unity

本作品使用 Cesium for Unity v1.7.1 插件，將 Cesium 和 3D Tiles 的 3D 地理空間功能與 Unity 結合，目的是在 3D Tiles 運行時，將大量高解析度真實世界攝影測量和 3D 地理空間模型可視化的顯示在 Unity 遊戲引擎中，筆者選擇 228 公園內的舞台，將生成的 3D 模型覆蓋在實際的舞台上。

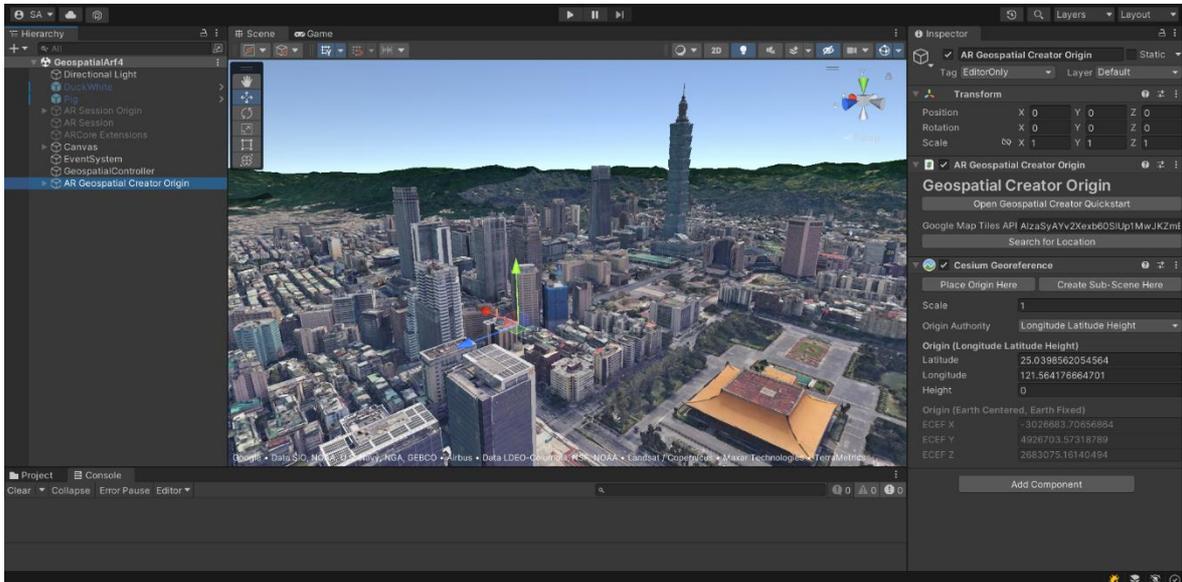


圖 27：Cesium 匯入 Unity 內的地理位置資訊

## 5. 語音互動

由於為了能讓使用者提高與真實世界的互動性，增加使用者探索環境的意願，因此加入故事、對話的方式，讓使用者加深扮演遊戲角色的認同，但同時考慮到使用者在操作 AR 時，必須要從手機觀看 AR 的視覺效果，同時注意移動時的周遭環境，因此互動方式以語音溝通的方式傳遞，降低使用者在視覺上的負擔。使用者可自由決定用錄音的方式與 NPC 的對話內容，且為了讓 NPC 回應能和玩家的對話有所呼應，因此 NPC 的對話是由 ChatGPT 即時生成(Text Generation)，其目的是為了能讓玩家提高在遊戲世界中的自由度，且更接近真實體驗，並且將 NPC 生成的文字以語音的方式播放，降低使用者的閱讀負擔，如圖 28。

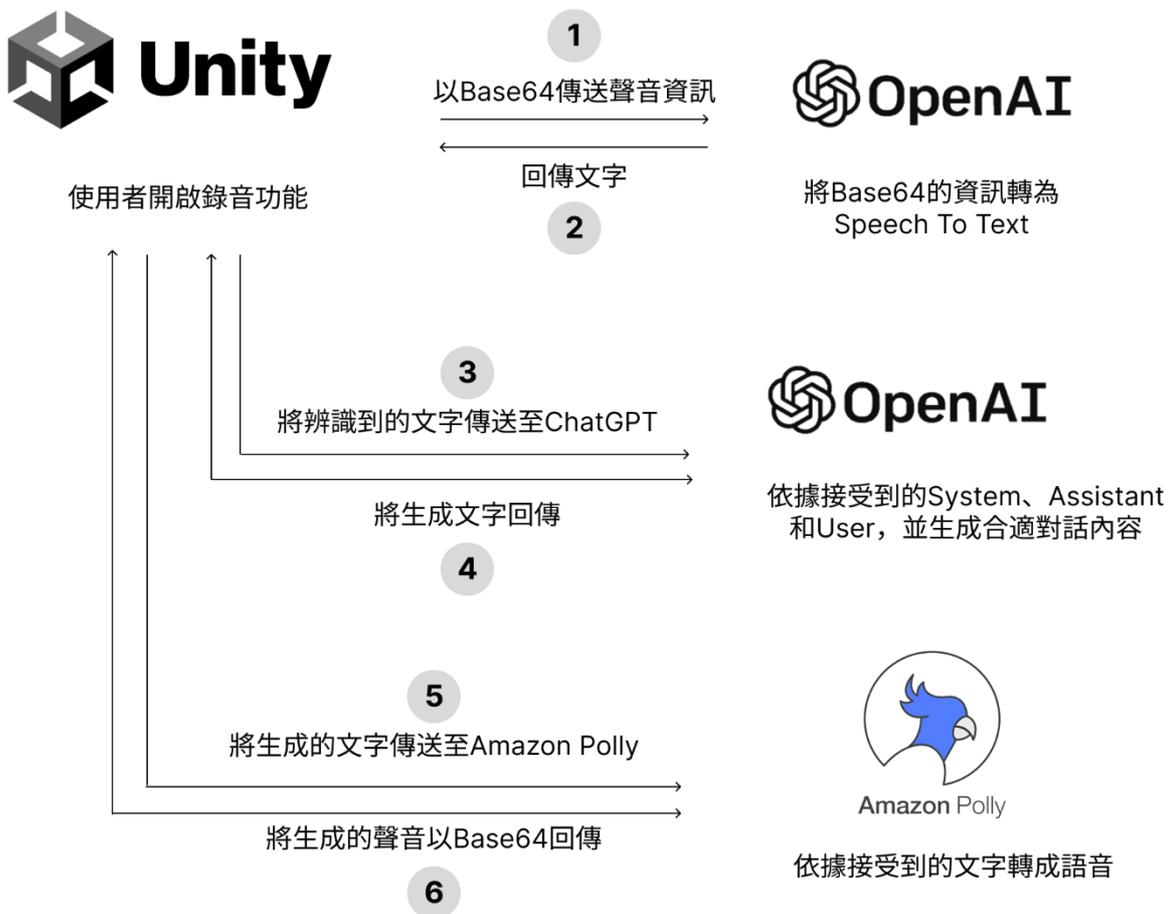


圖 28：對話生成架構

### 5.1 生成式語言模型 ChatGPT

基於轉換器的生成式預訓練模型 (Generative pre-trained transformer, GPT) 是由 OpenAI 公司於 2018 年所推出的模型，為一種革命性機器學習模型，利用監督和非監督學習技術，專注於理解和生成類似人類語言的模型 [32]，名稱中的「pre-trained」是指模型先在大量非標籤的文本數據上進行預先訓練[33]，而「Transformer」是指 Transformer 模型，是由 Bengio 等人於 2017 年提出的一種深度學習模型架構，其功能是在處理自然語言(Natural Language Processing, NLP 等順序輸入資料，可應用於翻譯、文字摘要等任務的方法[34]。該方法包括兩個階段：無監督的生成式「預訓練」階段，使用目標函式來設定初始參數；以及有監督的判別式「微調」階段，將這些參數在目標任務上進行微調[35]。

由於此作品希望能讓使用者宛如在另一個真實時空，能夠不受到限制的互動，根據 Deep Learning 的 ChatGPT Prompt Engineering for Developers 教學[36]，提示詞一共有 3 種角色作為設定依據，分別是 System、User 和 Assistant(圖 29)。

System 為高階角色，是用來提示模型，意即在模型接收訊息前，會先根據開發者以 System 角色所下的指令來決定怎麼處理接收到的訊息。以此應用程式為例，筆者傳送訊息至 ChatGPT 前，會先以 System 角色，提示 ChatGPT 扮演遊戲世界的 NPC，目的要將關卡提示詞告知使用者。

而 User 則為使用者輸入的訊息，意指為 ChatGPT 需要接收的訊息。以此應用程式為例，則會將使用者錄音後所辨識的文字(即使用者的對話內容)作為傳送的文字。當 ChatGPT 接受到使用者所傳遞的訊息時，會以所設定的角色與使用者對話，順利扮演 NPC 角色。

最後，會將過去的對話內容以 Assistant 的角色傳送，讓 ChatGPT 承接過去的對話內容，以避免每一次傳送訊息時，ChatGPT 只能單就該次訊息做回應，而是能延續對話訊息，有助於對話溝通的順暢度。

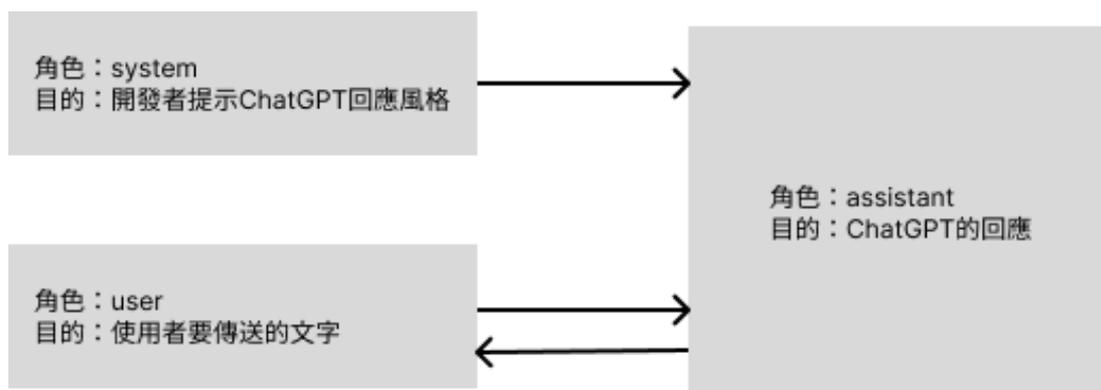


圖 29：對話生成架構

## 5.2 語音轉文字 (Speech-To-Text, STT)

語音轉換文字為透過收聽音訊，辨識其音訊內的語言和將辨識結果翻譯成文字，也稱為語音辨識或電腦語音辨識。

OpenAI 公司在 2022 年 9 月時推出 Whisper 模型[37]，Whisper 是一種自動語音識別（ASR）系統，可用來將語音轉換成文字，其特點為不同的口音或背景噪音吵雜的狀態下，依然能有穩定的辨識結果。[38]

由於此應用程式的預設環境為戶外，因此 Whisper 的高穩定性，可降低使用者操作時的錯誤率，避免判斷結果不佳，造成使用者重複錄製的問題。當使用者開始錄音功能時，程式會請使用者允許開起麥克風，當使用者點選確認按鈕後，應用程式會將聲音以 Unity Audio Clip 儲存並寫入手機內部，再將聲音以 Base64 格式傳送至 OpenAI 的 Create Audio Transcription 功能，透過 Whisper 模型進行處理，使用者的錄音檔轉為文字(Speech to Text)，待文字回傳後顯示在使用者畫面上。

### 5.3 文字轉語音(Text-To-Speech, TTS)

文字轉語音又稱為語音合成，是以人工方法產生語音訊號的技術，以文字作為輸入端，並生成仿造人類的語音[39]。

Amazon Polly 是 Amazon 子公司 Amazon Web Services 提供的一項雲端服務，使用深度學習技術來合成聽起來自然的人類語音，讓使用者可以將文字轉換為語音。

由於當使用者完成錄音後，會透過語音轉文字，並將文字以 User 的角色傳送給 ChatGPT，等到 ChatGPT 生成對話文字後，再將對話文字傳送至 Amazon Polly 轉換為語音，待 Amazon Polly 完成生成語音後，將語音檔以 Audio Stream 儲存至手機內部，並轉成 Unity 內建的 Audio Clip 播放。

## 6. 流程圖

使用者從首頁進入到遊戲頁面時，會先看到故事前情提要，接著開啟相機則會看到 AR 效果的環境畫面，當使用者欲和 NPC 對話時，則點選錄音鈕後，程式會將錄音內容轉為文字並傳送給 ChatGPT 生成回應，待對話完後則會開啟關卡 1 的 AR 效果作為第 1 關的題目，當使用者從相機內看到題目時，則可以錄音的方式完成回答，若答案正確即可進入下一關。

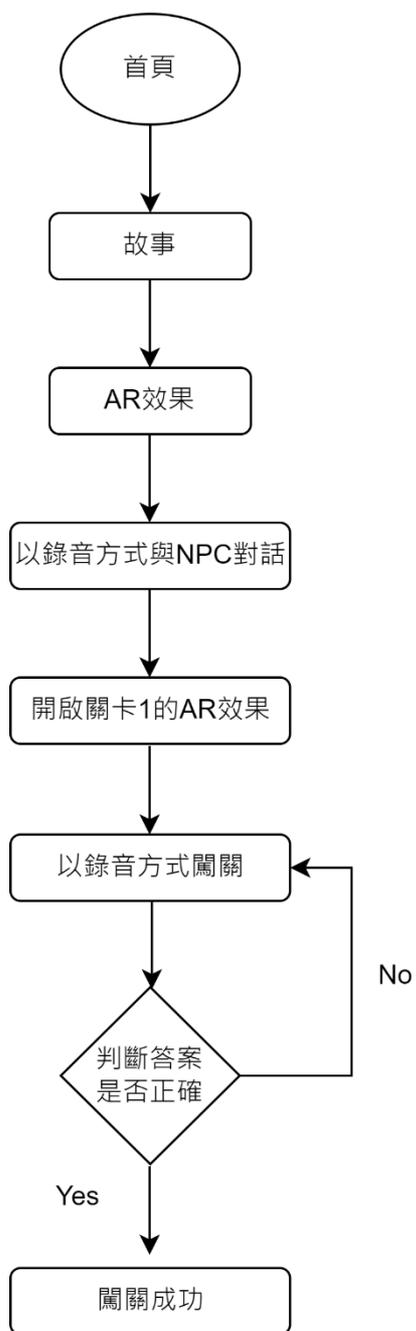


圖 30：〈Hello World〉作品關卡 1 的軟體架構圖

等到第 1 關闖關完後，則再以錄音方式與 NPC 對話，NPC 則給予第 2 關卡的提示，結束對話後，會開啟關卡 2 的 AR 效果作為第 2 關的題目，當使用者從相機內看到題目時，則可以錄音的方式完成回答，若答案正確即可進入下一關。

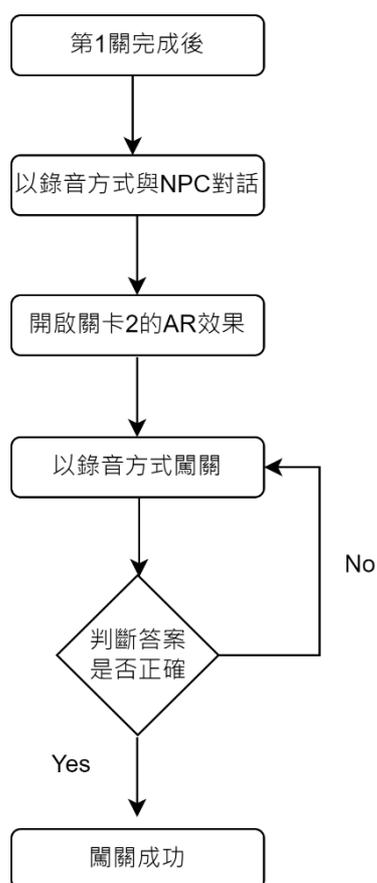


圖 31：〈Hello World〉作品關卡 2 的軟體架構圖

等待第 2 關完成後，則可與 NPC 對話取得第 3 關線索，並在結束對話後，開啟第 3 關的 MR 效果，MR 空間為一顆黑色圓球，靠近後則有一盞燈打在正中間，引導使用者前進，當使用者走向燈底下時，四周出現詭異的臉孔圍繞著使用者，當使用者選擇穿過臉孔，離開 MR 環境後，則可成功結束第 3 關卡，並開啟不同色調的 AR 效果，結束此遊戲。

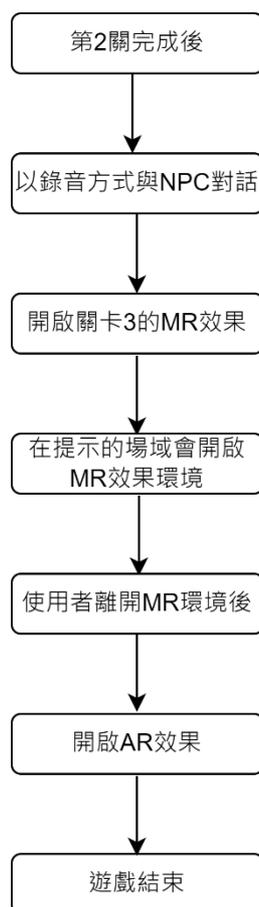


圖 32：〈Hello World〉作品關卡 3 的軟體架構圖

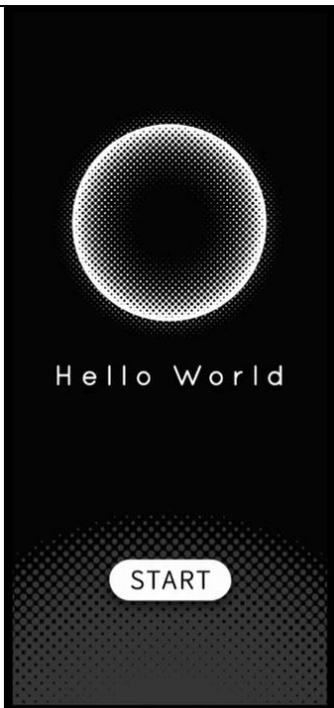
#### (四)、互動設計

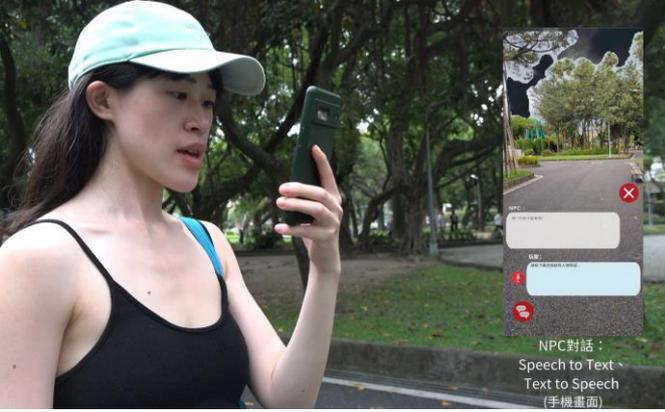
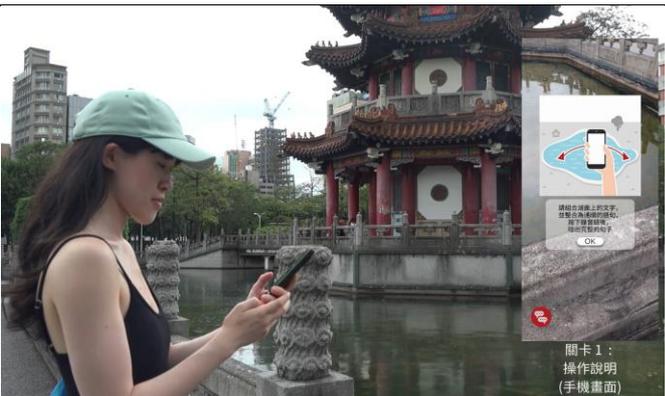
這個作品是參考注意力及自動化思考學理後發展的互動應用程式。使用者將使用一部 Android 手機並打開相應的應用程式。此應用程式將引導使用者開啟 GPS 和相機鏡頭，接著作品會透過故事的方式讓使用者進入遊戲情境，在這過程中，ChatGPT 將扮演 NPC 的角色，與使用者對話，並且引導使用者觀看增強實境 (AR) 效果。

當故事前言結束後，使用者開啟相機能觀看到的天空、人、樹木、水會依照故事情境而有不同的視覺效果，並透過與 NPC 的對話，獲取關卡的提示，並依照關卡指令將文字排列後以語音的方式唸出來。

最後來到指定關卡，讓觀者走進筆者設定的虛擬空間中，體驗筆者欲傳遞的訊息。

表 5：〈Hello World〉作品互動流程圖

序號	說明	示意圖
1	<p>歡迎頁面：點選「Start」按鈕後，會跳至下一頁，並開啟遊戲故事。</p>	
2	<p>AR 特效：從手機看到的現實環境中天空、人、樹木、水會依照故事情境而有不同的視覺效果。</p> <p>其目的是為了讓使用者能更好融入故事氛圍。</p>	

<p>3</p>	<p>操作說明頁面：使用者透過語音的方式與 ChatGPT 對話。</p> <p>其目的為降低使用者操作上的不便，且透過 ChatGPT 可以讓使用者在互動中自由對話。</p>	
<p>4</p>	<p>操作說明頁面：與 ChatGPT 對話後取得線索，並抵達關卡 1。</p> <p>湖面的影像為散落的文字，文字的擺放方式依照預設的使用者視覺所見位置而擺放。</p>	
<p>5</p>	<p>操作說明頁面：指示使用者將文字排列組合後，並用語音的方式回答。</p>	
<p>6</p>	<p>操作說明頁面：與 ChatGPT 對話後取得線索，抵達關卡 2，並依指示使用者將文字排列組合後，且用語音的方式回答。</p>	

7	<p>操作說明頁面：與 ChatGPT 對話後取得線索，抵達關卡 3，並走進虛擬的空間中，感受創作者欲傳達的理念。</p>	
---	---	--

### 三、呈現與紀錄

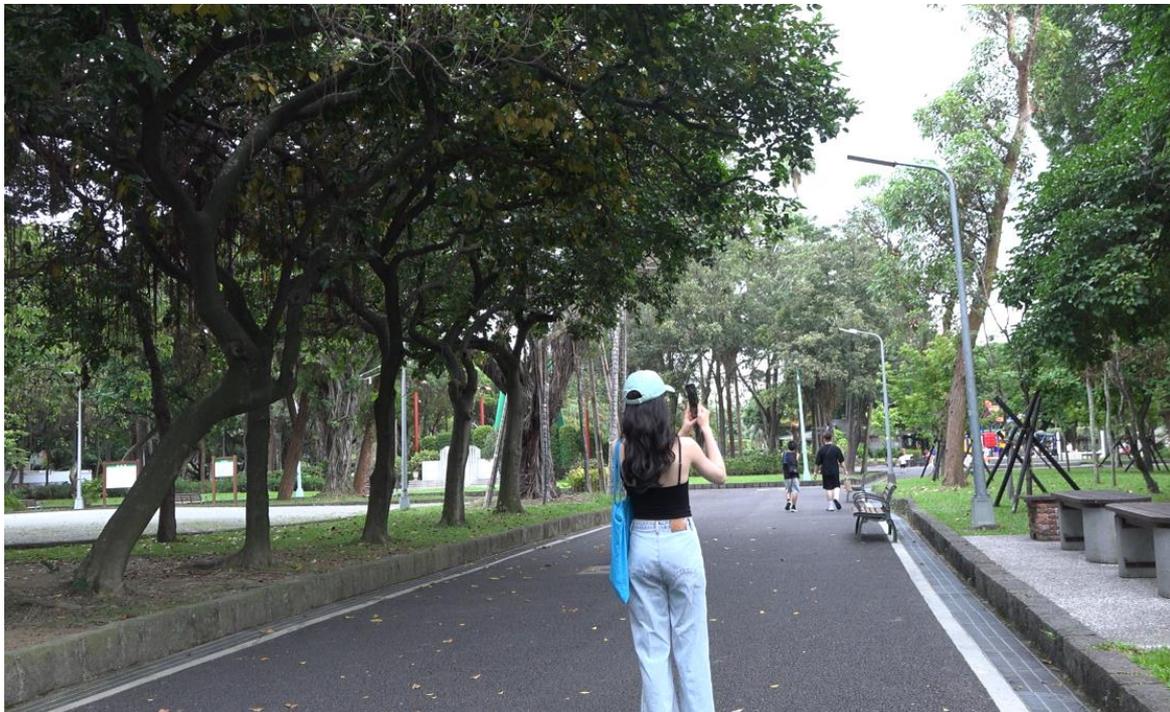


圖 33：〈Hello World〉實際展示畫面 1



圖 34：〈Hello World〉實際展示畫面 2

### 第三節 鏡中境

#### 一、創作理念

面對著無以名狀的情緒，該如何從中解脫？在前兩年筆者曾從事類似研發類型的工作項目，但可惜的是自身的專業知識、能力並不足以克服工作中的困境，所以常常在大家下了班後，獨自坐在公司的研究室裡不斷地 Try Error，試圖釐清原理與實作的關係，想要整理出頭緒來，然而這中間過程中多半是既無力且困頓的時刻，加上對於軟體工具的不熟悉，時常導致一個低級操作錯誤就讓先前 100 次的測試結果等同報廢，甚至因此下了錯誤的決策，讓之後的階段性的工作都往錯誤的方向前進，而時間依然在走，Deadline 一天一天靠近，面對著一籌莫展的進度，加速了自己焦慮不安的狀態。

然而回首過往，那段時間的 Try Error 乍看無用，彷彿浪費時間、精力，但現在想想，其實也不是毫無累積，因為從每一次的測試結果，開始思考自己的實驗步驟、流程哪裡需要改進，到最後嚴格控管每次實驗的變因，讓釐清問題的思考脈絡更加清晰，而身處在原本環境的我，充滿著焦慮的情緒，和時常覺得自己很失敗的念頭，有段時間讓自己卡在不願面對的狀態，離開了那個環境後，來到相對適合自己發展的地方後，發現當時的自己像是越級打怪，還沒學會走就想要飛，雖然不代表不能破關，但過程極其艱辛，現在回頭來看那段回憶，反而坦然接受，並心懷感激當時一步一步修正的過程，成為日後面對問題的 Mindset。

作品是以樂高組裝成顯微鏡的外觀呈現，之所以選用樂高作為媒材是因為樂高為一款拼砌玩具，拼砌過程像是遊戲，但又是需要思考及建構的過程，就像是走過那段日子後的我對於人生的感受，想起第一個學習的程式語言是行動裝置課程的 Swift，每當啟動一個新的專案時，都可以開啟名為 Playground 遊樂場的空專案，想傳達出最好的學習方法是探索、盡情發揮想像力，以及嘗試挑戰、解謎。

另外，由於觀者須透過目鏡觀看透明片上的影像，就像在操作顯微鏡時，觀察載玻片上的微生物一樣。影像內容則是筆者自己的模型，筆者自身的人物模型困在顯微鏡內，逃不出去，其情緒從困頓到無能為力，再到焦慮不安，而在此之前觀者都像是一個人類與微生物之間的關係，然而當影像上的人抬頭望向觀者發出求救

訊號時，觀者瞬間不再只是觀察者與被觀察生物之間的關係，而是從一個觀看的角度，變成劇情中的一部分，成為了被觀察者生命的一塊。最後影像上的人跳離了透明片的牢籠，順著實體的目鏡脫困，

而觀者原本的視覺專注點停留在透明片上的成像，在視覺影響下，影像成為了「主體物體」，切割了影像和目鏡之間的關係，也降低了實體目鏡在視覺中的重要性，但當影像的人物攀爬至目鏡逃離時，給予影像和目鏡之間新的關聯性，目鏡從裝置變成了影像的一部份，也讓觀者的視覺角度回歸到「整體」去思考，重新回到主體與周遭物件的關係。

這件作品是筆者回顧過去前兩件作品後的心情抒發，從第一件作品〈高敏感的聽說讀寫練習〉，心境及思緒映照著外界的人、事、物，常因一些風吹草動而在心裡起了波紋，而第二件作品〈Hello World〉反而選擇以主動的方式改變自身的視角，重新解讀那些映照在心境中的人、事、物，最後這件作品選擇用淡然的心態回頭看待過去的一切。而在學理的脈絡中，從原本的因為高敏感的性格，使得自己受困在他人的反應之下，退而去思索在認知心理學中，人類在自動化思考下，若以物理的方式改變感官刺激，嘗試影響知覺的歷程，驗證是否能獲得不同的結論，讓觀者在不同的角度下去觀看自我，是否能夠改變自己的解讀，而最後這件作品則是帶入認知心理學中不同的視覺刺激，對於知覺的思考脈絡有所轉換的概念。

## 二、創作方法與技巧

### (一)、硬體架構說明

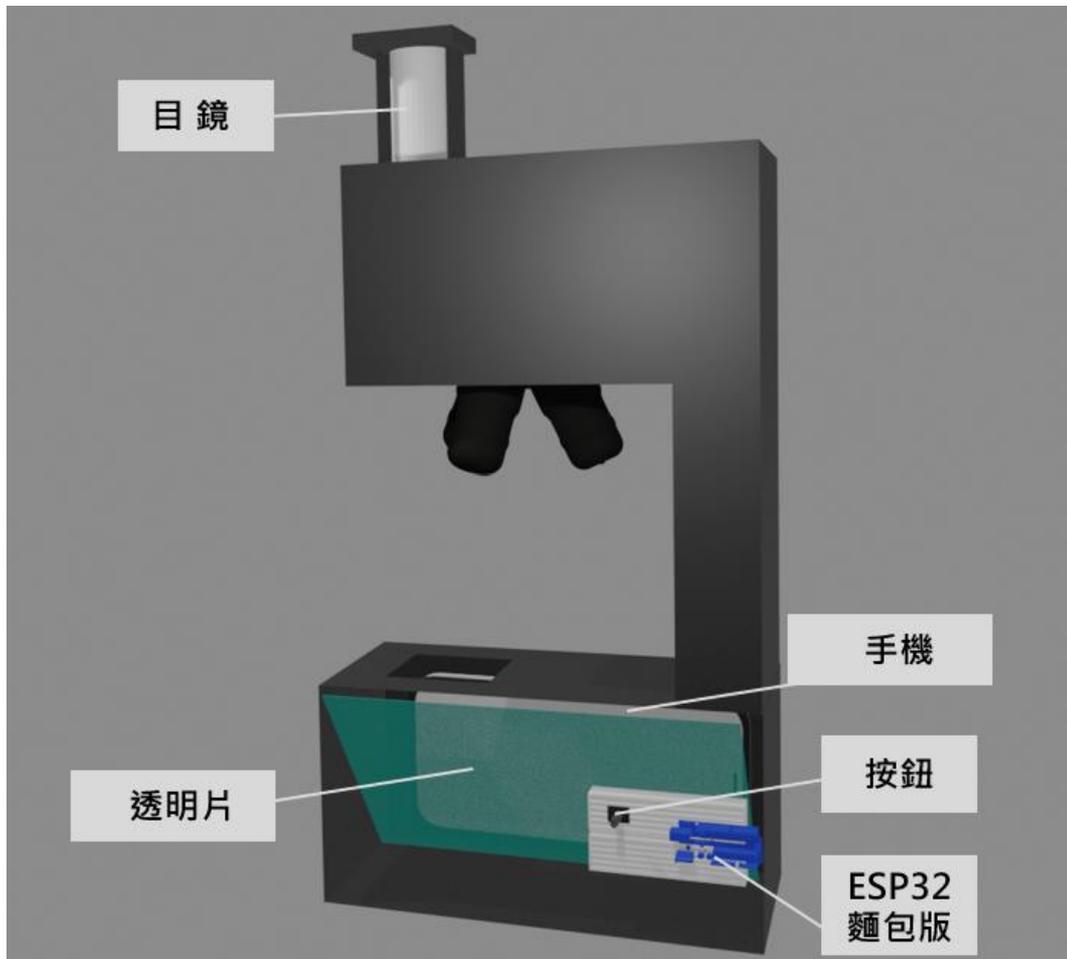


圖 35：〈鏡中境〉作品結構

作品的結構如圖(35)，利用樂高組裝類似顯微鏡的模型結構，並將目鏡放置於模型內，當使用者愈觀看影像時，則需透過目鏡觀看，且目鏡可讓使用者將目光聚焦於影像上，並具放大影像的功能，同時將 ESP32、麵包板及按鈕零件放置在模型內部，當使用者轉動模型的物鏡時，物鏡內的樂高模型會按壓按鈕開關，並透過 ESP32 作為 Client 發送訊號至 Server。

當手機接受到 Server 的訊號時，手機會播放相對應的畫面，透明片則會放置手機前方，以利將手機的畫面以 Hologram 全息投影的方式呈現在透明片上，讓使用者可從目鏡看到透明片上的影像。

表 6：〈顯微鏡〉硬體規格表

硬體名稱	數量	英文名稱與型號
伺服器運算用電腦	1 台	PC

單晶片微控制器	1 組	ESP32
安卓手機	1 台	Android smartphone
樂高	1 組	Lego
透明片	1 片	Transparency

## (二)、軟體架構說明

當使用者按下按鈕時，ESP32 會以 Client 的方式透過 Http 與 Server 溝通，若 Server 收到來自 ESP32 的請求，則會修改代表當前狀態的變數。而手機端則會每一秒持續向 Server 發送請求以取得當前狀態的變數，並依照回傳的變數判斷為待機模式或播放影片。

### 1. Node.js

Node.js 是跨平台、開源的 JavaScript 執行環境，可在 Windows、Linux、macOS 等作業系統上執行，本作品之伺服器(Server)是以 Node.js 架設，並透過 Http 接收 Client 端的請求。

### 2. Arduino、ESP32 Library

單晶片微控制器(ESP32)是以 Arduino 的 WiFi Library 及 ESP32 的 Http Library 開發。

### 3. Unity

本作品的安卓手機 App 是以 Unity2021.3.10f1 版本進行開發。

### 4. 系統架構圖

本作品軟硬體整合共分為 2 個部分(圖 36)，第 1 個部分為 Server 與 ESP32 的溝通及 Server 與手機端的溝通。此作品之所以選用 ESP32 是由於 ESP32 為低功耗的單晶片微控制器，並且本身整合了 WiFi 和藍牙功能，不須另外外接模組，ESP32 可透過 Arduino 的 WiFi Library 連接現場場域的無線存取點

(Wireless Access Point)，並透過 ESP32 的 Http Library 以通訊協定方式和 Server 溝通。

而 Server 則是由 Node.js 架設，並依據 ESP32 的 Post Request 而更改裝置狀態，並將狀態以 Response 的方式傳送給 Unity App，當 Unity App 接受到 Server 的回應後會依據 Response 內容而選擇該播放的影像。



圖 36：〈鏡中境〉作品軟硬體系統架構圖 1

Server 與 ESP32 的溝通流程為當使用者按下按鈕時，按鈕會透過腳位將訊號傳送至 ESP32 以觸發開關，若 ESP32 判斷結果為是則會透過 Http 通訊協定以 Post 的方式將變數傳送至 Server，當 Server 接受到來自 ESP32 的訊號時，便會儲存其變數作為最新的裝置狀態；若無使用者觸發開關或 Server 沒有接收到 ESP32 的訊號時，則回到預設的待機狀態，等待下一次 Client 發送的訊號。

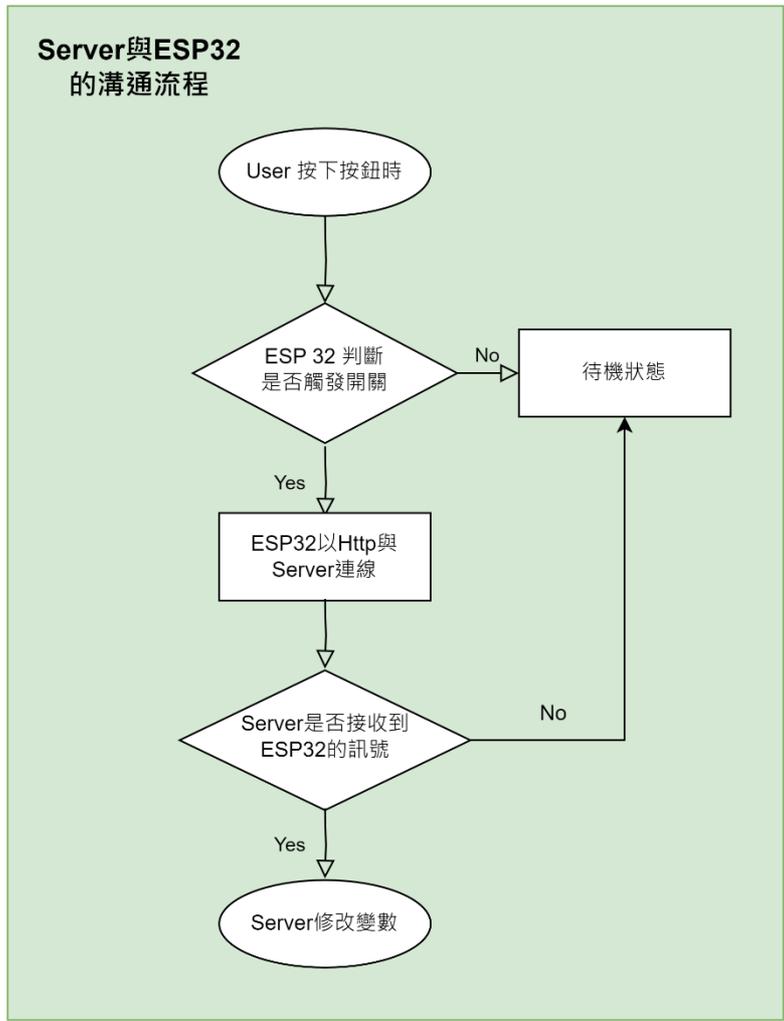


圖 37：〈鏡中境〉作品 Server 與 ESP32 之流程圖

而手機端在開啟 App 後會與 Server 建立連線，並以每 1 秒為頻率持續以 Http Get 的方式向 Server 取得變數作為最新的裝置狀態，若 Server 回傳的變數不同時，則會依據變數來選擇播放的動畫影片；若動畫影片尚未播放完畢，且 Get 所取得的變數依然為播放動畫影片時，則程式會視為新的使用者觀賞其裝置，影片會重新播放；若影片播放完畢且最新的裝置狀態為待機模式時，則會播放黑畫面，等待下一次播放動畫影片。

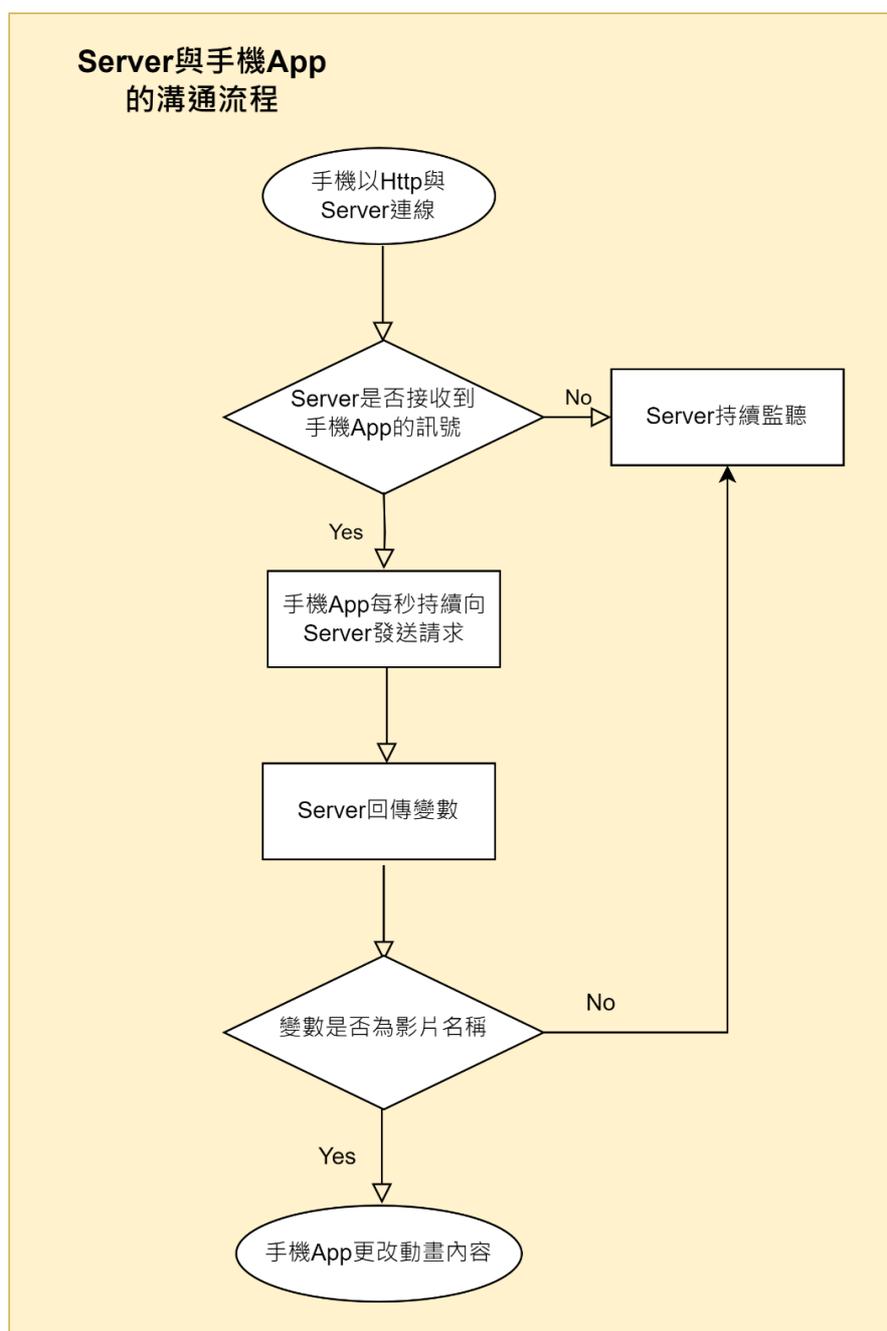


圖 38：〈鏡中境〉作品 Server 與 App 之流程圖

### (三)、影像設計

影像設計的製作共可分為建模、骨架兩個部份，流程為先透過 in3D App 建立筆者自己的模型，再透過 Rokoko 錄製真實動作，並生成動作骨架後，利用 3D 軟體 Blender 將模型綁定動作骨架，最後渲染為影片，如圖 39。

1. in3D：in3D 是一款專門掃描真實人物的應用程式，透過安裝 in3D App 的智慧型手機相機分別拍攝真實人物的臉、頭及身體，並透過電腦視覺、深度學習將 2D 影像生成為 3D 模型。
2. Rokoko Vision：Rokoko 是丹麥專注於動作捕捉和動畫技術產品品牌，其推出利用單一攝影機錄製影片，即可將影片人物的動作生成動作骨架，如圖 40。



圖 39：影片的製作流程

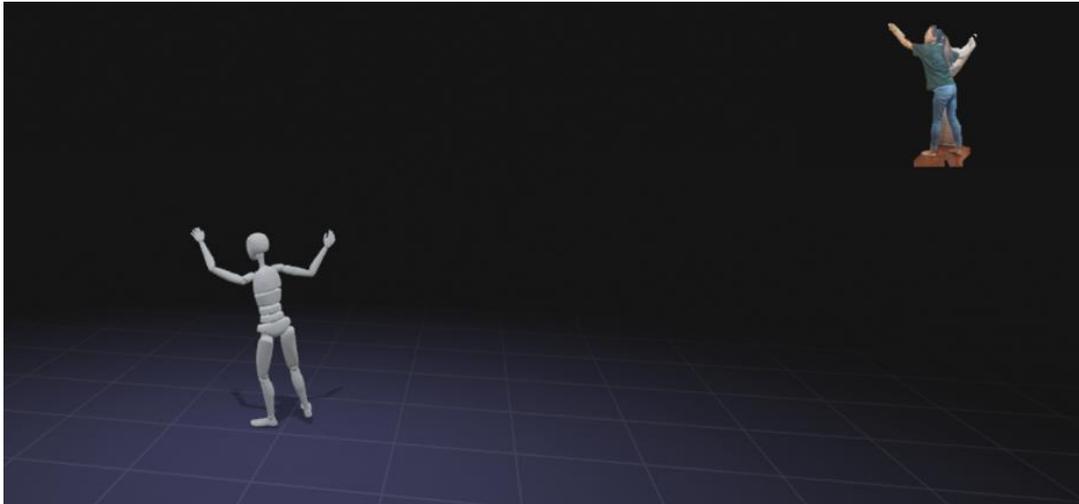
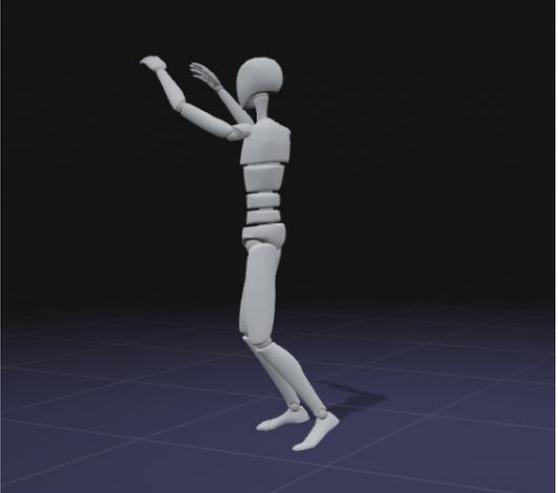


圖 40：筆者的動態影片與 Rokoko Vision 結果畫面

3. 動作設計：

表 7：〈顯微鏡〉動作設計圖

序號	說明	示意圖
1	人物被困在一堵圓形的牆內，不斷地拍打牆面。	
2	人物低著頭回到中間。	

3	突然下起雨了，人物下意識用手遮住頭。	
4	因為圓形的牆為一座密閉空間，腳下也感覺水正漸漸地淹上來了，不斷地踩踏腳下的水。	
5	突然抬起頭向上望，看到正在看著一切的顯微鏡的觀察者，並向上呼叫、求救。	

6 奮力跳躍，盤爬著顯微鏡的目鏡逃脫。



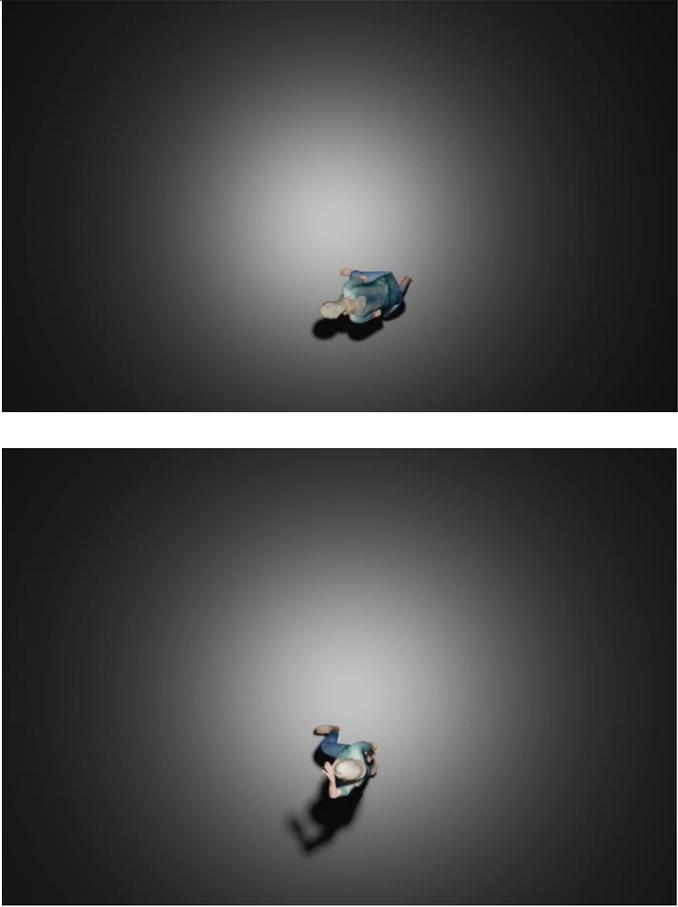
(四)、互動設計

表 8：〈顯微鏡〉作品互動流程圖

序號	說明	示意圖
1	使用者站立於裝置前方	
2	使用者按壓側邊按鈕時，手機會播放動畫，動畫會以 Hologram 全息影像的形式呈現。	

<p>3</p>	<p>使用者低頭看向目鏡</p>	
<p>4</p>	<p>目鏡內所看到的 Hologram 全息影像的 動畫。</p>	
<p>5</p>	<p>影像為筆者對於困境的 反應。(此為畫面為影 片畫面，非從目鏡所看 到的畫面)</p>	

6	<p>人物被困在一睹圓形的牆內，不斷地拍打牆面，希望能得到回應。 (此為畫面為影片畫面，非從目鏡所看到的畫面)</p>	
7	<p>人物低著頭回到中間，象徵著困在牆內的無力感。(此為畫面為影片畫面，非從目鏡所看到的畫面)</p>	
8	<p>人物因下雨而造成的淹水感到焦慮不安，不斷地踩踏腳下的水。象徵著時間的逼近，加速了焦慮的心情。(此為畫面為影片畫面，非從目鏡所看到的畫面)</p>	
9	<p>突然抬起頭向上望，看到正在看著一切的顯微鏡的觀察者，並向上呼叫、求救。(此為畫面為影片畫面，非從目鏡所看到的畫面)</p>	

<p>10</p>	<p>可惜觀察者也愛莫能助，但人物依照觀察者的目光，選擇奮力跳躍，盤爬著顯微鏡的目鏡逃脫。(此為畫面為影片畫面，非從目鏡所看到的畫面)</p> <p>此時使用者就像實驗的觀察者，在觀察著被觀察生物的樣貌，像是人生在某些時刻回首過往，反而已經不再糾結當下的心情，而能夠更宏觀的角度看待這段經歷對。</p>	
-----------	---	---

### 三、呈現與紀錄

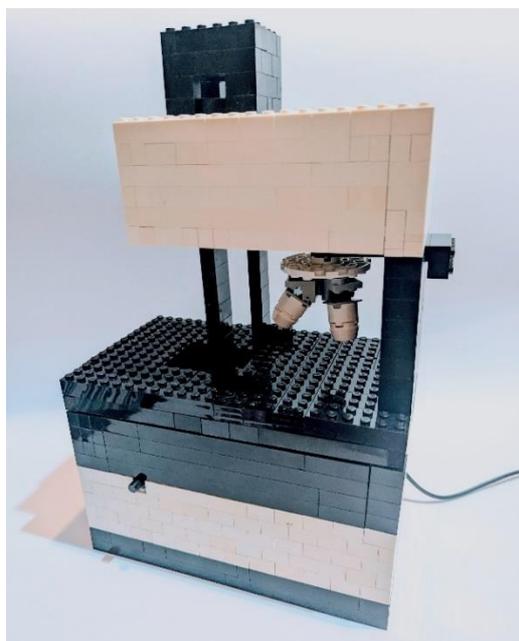


圖 41：〈鏡中境〉實際展示畫面

## 第四章 結語

本論文是一個屬於自我的學習、摸索、掙扎、重生的狀態，這三件作品呼應了創作者本身的體悟、感懷，在創作的過程中去挖掘內心的樣貌，甚至對自我的批判，不知道該怎麼處置這些意念時，只能從生活、書籍中尋找，不知不覺這樣的狀態也療癒了自己不安的心境，由於筆者自身的性格難以自我揭露，所以試圖將自我覺察的歷程以心理學的角度出發，讓作品與筆者自身保持一些距離，而不做到完全的揭露自我，同時透過文字、物件去改變觀者對於場景語意的認知，試圖傳遞筆者在創作時自身的內心變化。

本論文的第一件作品<高敏感的聽說讀寫練習>是筆者長期一直有的困擾，在就讀研究所時期身旁的同學們也有不少人在關注高敏感的議題，所以筆者同時身處在這樣的心理狀態時，又像一個抽離的第三人觀察著其他有相同困擾的同學，發現這樣的狀態會造成自我內耗，把自己關進一個隱形的牢籠裡，而筆者自身也不知道該怎麼面對的樣的心境，只能將這樣的感受紀錄在作品中，並且透過文字具象呈現在空間中，讓互動過程改變場景中的語意傳達。

然而，不過在創作、美學上還是有諸多要改進的地方，逃避創作、美學相關的研究太久了，不管是從什麼時候看這件作品時，自己都深知作品的問題及不足，畢竟所有事物都是如此，時間花在哪，成就就在哪，沒有花時間打磨，自然不可能可以一步登天。

在技術層面中，慶幸的是從就讀研究所開始到回職場工作的這幾年，都持續學習程式開發的能力，所以比起更早期的創作，在技術實驗及掌握的困難度也降低了，所以對筆者而言像是在技術積累後的成果。另外，就技術層面，過往相關的互動科技作品多半透過物理性的移動攝影機或投影機的位置，甚至是直接改變影像內容去達成 Projection Mapping 的效果，而此論文紀錄創作過程中如何透過相機校準的方法，讓受限於物理空間位置條件、影像內容下，能夠做到攝影機成像和投影機成像範圍一致。

延續第一件作品是針對高敏感的心理角度出發，在創作第二件作品<Hello World>在創作發想時，筆者也在自身反思，思考著究竟是什麼讓我們對於外界的刺

激下，不同的人有著截然不同的理解、感受，因此選擇延續心理學脈絡，再退一步探究人類在認知前期，有哪些歸因可以改變最終的認知理解。

探究原理的過程中，也發覺人類的認知心理學的歷程和影像辨識中場景語意的模型開發歷程其時有高度的相似，因此將兩者的概念結合，並嘗試在作品中直接使用場景語意的技術去表現認知心理學的狀態。同時在開發與探究學理的過程中，不知不覺也影響者筆者對於人生觀的反思，從原本的受困於他人言語、行為後，再回歸自我，試圖找到改變自我解讀的方法。

在技術層面中，整合了許多不同的技術，在 AR 的部分則加入了 Scene Semantic 場景語意的實作和 Geospatial，讓 AR 互動在 2D 影像及 3D 虛擬物件上都有不同的視覺效果。此外，在遊戲互動上透過許多 AI 模型工具做到完整的 NPC 互動體驗，並且提出能降低使用者操作負擔的互動流程，讓使用者只需透過聽、說的方式就能和遊戲 NPC 互動。

然而，在創作的過程中，對於關卡文字的限縮，使得使用者在闖關中只能回答一種答案才算完成，未來在作品上會多加思量，若文字的設計可以以不同的排列方式，或是留一扇門給觀眾，讓觀眾有更開放的空間，而這樣的修正方向同時也更加深筆者原先愈傳達的理念，因此在未來創作上也會思索在文字的表達上如何才能發揮更多可能性。

延續前面兩件作品，從訊息分析到改變認知，再更退一步追尋那麼又是什麼改變我們的感官刺激，因此嘗試往視覺看到物件的區域的分析方式不同出發，從原本的以物理性的方式改變感官刺激，進而影響認知，更進一步嘗試在不改變感官刺激下，試圖影響認知理解。

第三件作品<鏡中境>，筆者從原本的試圖透過心理學學理，讓作品像是一個旁人在觀察後的結論，變成改變原本的初衷，加入了自身的過往經驗，並利用顯微件的物件意涵，改變空間中的場景語意，讓原先只是「觀者」觀看影像，變成「觀察家」在觀察微生物的狀態，因為顯微鏡的象徵意義，讓觀者、影像有了不同的屬性，而原有單純的「觀看」的行為訊息也變成了「觀察」。在創作的歷程中，筆者的創作心境，也從最初的從由他人行為反應下的自我，再到思考如何選擇用不

同的視角改變自我解讀，再到坦然接受過往的一切，這樣的狀態，讓筆者想起了蘇軾的定風波「回首向來蕭瑟處，歸去、也無風雨也無晴。」過往的得與失、成與敗，早已置之度外，未來的風雨與天晴也都是人生的必然罷了。

而此作品原本僅是讓使用者按下按鈕完成機具操作的行為，但未來可朝是否有更好的呈現方式改進，讓使用者的行為既能有機具操作的目的、情境帶入，又能傳遞該行為所象徵的意涵，例如也許透過放置載玻片、轉動目鏡等，讓互動過程更加具說服力。

這三件作品從「跟隨他人目光」到「嘗試改變自我的認知」，再到「專注自我在觀看的視角」，嘗試透過心理學中的「高敏感」、「認知階段」，最後是「感官刺激到知覺」的概念作為作品發展的方向，並結合場景語意中，從被動的等待文字改變空間中意涵，到加入場景語意的實作，最後主動透過物件去改變場景中的語意，這個過程宛如是在剖析自己的心理狀態，最後意外地過程中療癒自己。

最後，回頭檢視這一路的轉變，在技術層面上，進行了許多的嘗試與整合，力求在互動過程上有更好的體驗。而在概念層面，起初以學理為基礎，使作品與筆者之間保持一定的距離，想以客觀的角度來進行創作；隨著時間推移，開始從自我經驗出發，讓作品更加貼近內心的感受與體驗，每一個階段的嘗試與轉變，不僅豐富了筆者的創作歷程，也讓作品和自我之間的關係有了更深的理解，每一步的探索和變化，都是形塑自我覺察的重要里程碑。

## 參考文獻

- [1] John C. Bogle 著，吳莉君譯，《觀看的方式》，臺北市，麥田出版，2010。
- [2] 《Wikipedia•彼得潘症候群》  
(<https://zh.wikipedia.org/wiki/%E5%BD%BC%E5%BE%97%E6%BD%98%E7%97%87%E5%80%99%E7%BE%A4>，2019/12/23 檢索)
- [3] Ilse Sand 著，呂盈璇譯，《高敏感是種天賦：肯定自己的獨特，感受更多、想像更多、創造更多》，台北市，三采出版，2017。
- [4] 《容易想太多、受他人影響？這可能是高敏感人特徵！心理師教你如何面對自己的「天賦」》  
(<https://www.healingdaily.com.tw/articles/%E9%AB%98%E6%95%8F%E6%84%9F%E4%BA%BA%E7%89%B9%E5%BE%B5-%E5%BF%83%E9%9D%88%E7%B4%93%E5%A3%93>，2024/02/25 檢索)
- [5] Pashler, H. (1998). *The Psychology of attention*. Cambridge, MA: MIT Press.
- [6] 楊桃蹊、包燕(2013)。視覺信息加工中的空間注意與特徵注意。北京大學心理學系及機器感知與智能教育部重點實驗室。中國大陸北京市。
- [7] Peterson, M. A., & Rhodes, G. (Eds.). (2003). *Perception of faces, objects, and scenes: Analytic and holistic processes*. Oxford University Press.
- [8] Adam Benforado 著，堯嘉寧譯，《不平等的審判：心理學與神經科學告訴你，為何司法判決還是這麼不公平》，台北市，臉譜出版，2016。
- [9] 張淑美(2002)。手機呈現角度之觀測閾值及其影響因素研究。國立成功大學工業設計學系碩博士論文。未出版。台南市。
- [10] Solso, R. L. (1998). *Cognitive psychology* (5th ed.). Boston, MA: Allyn and Bacon.
- [11] 張春興（1998）。教育心理學：三化取向的理論與實踐（2版）。臺北：臺灣東華。
- [12] 《Wikipedia•認知發展論》(<https://zh.wikipedia.org/zh-tw/%E8%AA%8D%E7%9F%A5%E7%99%BC%E5%B1%95%E8%AB%96>，2024/03/03 檢索)
- [13] Reed, S. K. (1988). *Cognition: Theory and application*. Pacific Grove, CA: Brooks/Cole.

- [14] 《Wikipedia•訊息處理論》 (<https://zh.wikipedia.org/zh-tw/%E8%A8%8A%E6%81%AF%E8%99%95%E7%90%86%E7%90%86%E8%AB%96>，2024/03/03 檢索)
- [15] Elliot Aronson, Timothy D. Wilson, Robin M. Akert 著，黃光國校閱，余伯泉、陳舜文、危芷芬、李茂興譯，《社會心理學》，新北市，揚智文化，2011。
- [16] Bargh JA, Ferguson MJ “Beyond behaviorism: On the automaticity of higher mental processes”. *Psychological Bulletin*, 126(6), 925–945.
- [17] 賴韋伶(2020)。從《道德經》看 James Turrell 光藝術創作。臺北市立大學，臺北市。
- [18] Krakow Witkin Gallery. Pearls, truisms & survival. Retrieved (<https://www.krakowwitkingallery.com/artwork/pearls-truisms-survival/>，2024/06/08 檢索)
- [19] Adrian Kaehler, Gary Bradski 著，賴屹民譯，《OpenCV 3 學習手冊 (Learning OpenCV 3: Computer Vision in C++ with the OpenCV Library)》，臺北市，歐萊禮，2017。
- [20] Rafael Lozano-Hemmer ([https://www.lozano-hemmer.com/artworks/tape\\_recorders.php](https://www.lozano-hemmer.com/artworks/tape_recorders.php)，2024/06/07 檢索)
- [21] 王暉博(2019)。利用領域調適技術轉移抽象場景結構資訊以提升語意式影像切割與深度估計。國立交通大學，新竹市。
- [22] 龔艾瑤(2015)。根據語意式切割與物體危險等級預測的道路場景理解。國立清華大學，新竹市。
- [23] ARCore - Scene Semantic (<https://developers.google.com/ar/develop/scene-semantic>，2024/06/07 檢索)
- [24] Crawford, K., & Paglen, T. (2019). Excavating AI: The politics of images in machine learning training sets. (<https://excavating.ai>，2024/06/07 檢索)
- [25] ARCore (<https://developers.google.com/ar/develop/geospatial?hl=zh-tw>，2024/02/25 檢索)
- [26] 《OpenCV CustomPattern》 ([https://docs.opencv.org/4.x/d3/d3e/classcv\\_1\\_1ccalib\\_1\\_1CustomPattern.html#a3829a895b1530f961fdd8a41f5419fff](https://docs.opencv.org/4.x/d3/d3e/classcv_1_1ccalib_1_1CustomPattern.html#a3829a895b1530f961fdd8a41f5419fff)，2024/03/03 檢索)
- [27] 《TouchDesigner Documentation-Palette:camSchnapp》 (<https://docs.derivative.ca/Palette:camSchnapp>，2024/02/28 檢索)
- [28] 曾啟雄、孫志誠(1999)。視知覺中的律動感與空間感之探究-以歐普藝術為

例。臺灣美術。

- [29] OpenAI. Text generation models. (<https://platform.openai.com/docs/guides/text-generation> , 2024/05/31 檢索)。
- [30] OpenAI. Speech to text. (<https://platform.openai.com/docs/guides/speech-to-text> , 2024/05/31 檢索)。
- [31] AWS. Amazon Polly. (<https://aws.amazon.com/tw/polly/> , 2024/05/31 檢索)。
- [32] 賴郁仁(2024)。生成式人工智慧工具介入學生執行設計流程的初探-以 ChatGPT 與 Midjourney 為例。國立成功大學，臺南市。
- [33] 賴益楚(2024)。探索台灣中高齡族群使用 ChatGPT 之意圖與接受程度。輔仁大學，新北市。
- [34] 余昌翰(2022)。基於 Transformer 及姿態辨識之即時手語翻譯系統。國立中央大學，桃園市。
- [35] Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training. ([https://cdn.openai.com/research-covers/language-unsupervised/language\\_understanding\\_paper.pdf](https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf))
- [36] Deep Learning AI (<https://www.deeplearning.ai/short-courses/chatgpt-prompt-engineering-for-developers/> , 2024/06/01 檢索)
- [37] 呂可名(2024)。以 Whisper 模型多語言語音辨識與分段之即時語者分段標記系統。長庚大學，桃園市。
- [38] OpenAI Whisper (<https://openai.com/index/whisper/> , 2024/06/01 檢索)。
- [39] 蔡孟庭(2024)。語音合成技術應用於電腦輔助發音訓練之研究。國立臺灣師範大學，台北市。