

二、研究計畫內容：

基於語音辨識與人機交互輔助以改善演講者報告之簡報系統 - Smartlide

(一) 摘要

隨著數位科技產品的迅速發展，簡報成為許多講者必備的輔助工具。依據教育部國民及學前教育署頒布之規定，每間普通教室皆需設置投影設備供師長及學生使用，藉此可知簡報對教育的重要性，更因其將複雜的概念、數據和統計等資訊，以有條理的方式呈現給學生及觀眾，使得其在各行各業扮演重要角色。

然而，講者在簡報時，經常需兼顧知識內容的傳達與簡報頁面的切換操作，這不僅增加講者的認知負荷，還可能因此影響簡報流暢度和觀眾注意力等問題。此外，講者也時常需要將當下所述的內容標記更為醒目，如畫線、框選或加入圖形等註記，讓觀眾快速閱讀，達到更佳的簡報效果。

本研究預期開發一款系統改善前述問題，運用語音辨識技術自動解析講者的語音指令，通過人機交互設計實現自動辨識及切換簡報頁面與標記簡報重點。此系統不僅減少了演講者對控制設備的依賴，還能提升簡報的互動性和輔助功能。除解決前述問題外，本系統亦提供簡報過程的文字記錄，供事後統整內容使用。藉此系統達到減少講者的操作負擔，提升簡報的互動性及專業度。

關鍵字：簡報輔助、語音辨識、自動化操作

(二) 研究動機與研究問題

2.1 簡報的重要性與挑戰

簡報是現代演講中不可或缺的工具，它不僅能傳遞訊息，更是強化演講者與觀眾之間互動的重要橋樑。對於演講者來說，簡報能幫助組織演講內容，讓複雜的概念更容易被理解。利用視覺元素的呈現方式，從而提升演講的效果 [1]。同時，良好的簡報設計亦可提高演講的專業性，一方面可將重點條理分明，避免疏漏。另一方面於演講過程中提供引導，幫助演講者保持焦點，避免演講偏題。由此可知，一份成功的簡報對演講者有著極大的幫助。

2.2 研究動機

隨著數位科技的快速發展，現代教育與專業領域工作會議中的簡報已成為傳遞知識和觀點的核心媒介之一。依據教育部國民及學前教育署頒布之《國民小學及國民中學設施設備基準》和《普通型高級中等學校設備基準》，規定每間普通教室皆需設置投影設備供師長及學生使用。同時，大學通識教育強調讓學生能思考、能表達、能反省並成為有責任的現代公民[2]，提供相關的課程以確保大學生在畢業前具備流暢的口語表達與溝通能力，這也將成為未來於學術、個人與事業成功的要件之一[3]。藉此可知簡報在不同求學階段皆佔有重要地位，更因其能將複雜的概念、數據和統計等資訊，以有條理的方式呈現給學生及觀眾，使得其在各行各業扮演重要角色。然而，儘管簡報帶來無數的便利和利益，但對講者而言，操作這些工具仍然存在許多挑戰。大多數演講者在使用投影設備時，需同時兼顧講解內容和簡報頁面的切換，這不僅增加了他們的認知負荷，甚至可能造成演講者過於緊張，降低觀眾對講者的信心[4]，還容易導致簡報內容的中斷，進而影響觀眾的注意力和整體理解。

此外，在簡報過程中，演講者往往需強調某些關鍵資訊，如透過畫線、標註、框選圖像或加入重點註記，這些操作通常需要手動進行，增加操作的繁瑣度和演講者的操作壓力。這不僅會分散演講者的注意力，降低簡報流暢度，影響整體學習效果。因此，如何減少演講者對設備的操作依賴，並讓簡報過程更加專業、提高互動性，是此研究欲解決的問題。

2.3 研究目的

本研究旨在透過語音辨識技術，減輕演講者在進行簡報時的操作負擔，使其能更專注於內容的講解與觀眾的互動，而無需頻繁依賴滑鼠、簡報筆或其他輔助工具來控制簡報頁面。研究的首要目標是開發一套語音控制系統，讓演講者能夠透過自然的語音指令來實現簡報的自動辨識與頁面切換，使簡報過程更加順暢且直覺，從而減少因外在操作因素所帶來的認知負荷[5]。此設計不僅能提升演講者的操作便利性，也有助於維持其專注力，使其能更精準地表達內容，提升演講的整體流暢度與專業度。

此外，在許多學術演講、商業簡報或教育教學的場合，講者往往需要即時標記簡報中的關鍵資訊，以幫助觀眾快速掌握重點內容。然而，這類標記通常需要透過滑鼠、觸控筆或其他手動方式操作，這不僅影響演講的流暢度，甚至會分散講者的注意力。因此，本研究亦將設計一套語音指令觸發的標記功能，使講者能夠透過簡單的語音指令，如「畫一條紅線」或「標記這個數字」，即時在簡報上標註，進一步提升簡報說明的清晰度與說服力。

除了頁面切換與標記功能外，本研究亦考慮到許多簡報場合會涉及資訊的即時記錄與後續整理，因此，系統將加入語音轉文字的即時紀錄功能，讓演講內容能夠自動轉錄成文字，提供觀眾或講者作為事後回顧與學習的參考。這項功能特別適用於教育領域的課堂講授、研討會記錄，或商業會議的決策紀錄，不僅能提高資訊的可獲取性，也能有效減少手動筆記的需求，提升學習與知識管理的效率。

整體而言，本研究期望透過語音辨識技術，結合自動簡報控制、標記強調與即時紀錄三大核心功能，來打造一個更智能且便利的簡報輔助系統，使講者能更加專注於內容傳達，提升簡報的互動性與影響力。

(三) 文獻回顧與探討

3.1 多媒體簡報

美國訓練與發展協會 ASTD(American Society for Training and Development)將多媒體定義為：包含互動性文字、圖片、聲音及色彩。一頁簡單的簡報到複雜的互動性模擬皆為多媒體簡報的範疇[19][20]。本研究以 Google 公司提供之簡報系統 Google Slide 做為目標操控對象，透過講者講述的內容，自動切換簡報至對應頁面。

3.2 簡報對講者的重要性

專業簡報是教育、交流和分享最新資訊的重要方式。根據文獻指出，非職業演講者因少有時間和機會進行訓練、準備或設備不足，將導致無法讓觀眾充分吸收寶貴的資訊，反而讓他們感到無聊和分心[6][7]。而非專業演講者通常依賴朗讀，但語音節奏、語調變化與自然口語相比可能較為僵硬，進而影響觀眾的理解與專注度。因此，適當的演講技巧訓練、聲音與語速的掌握、視聽輔助工具的使用、吸引觀眾參與和提升演講娛樂性，能使演講更具吸引力與影響力[8]。

3.3 認知負荷

在認知心理學中，認知負荷(Cognitive load)是指工作記憶資源的使用量。根據認知負荷理論，人類的學習可以分為生物學上基礎知識(primary knowledge)與次級知識(secondary knowledge)。基礎知識是人類天生的能力，如語言與社交互動，而次級知識則需要藉由教育與學習。由於演講技巧屬於次級知識，因此學習者在面對複雜的演講時，若沒有足夠的長期記憶，可能會遇到高認知負荷的挑戰。為提升學習效果，教學設計應考慮減少外在認知負荷[9]。

3.4 自動語音辨識(Automatic Speech Recognition, ASR)

自動語音辨識(ASR)核心目標是將語音訊號轉換為對應的文字，進而提升人機互動的便利性與效率。典型的自動語音辨識系統架構中，包含訊號處理與特徵擷取(Signal Processing and Feature Extraction)、聲學模型(Acoustic Model, AM)、語言模型(Language Model, LM)與假設搜尋(Hypothesis Search)[10]。

然而，自動語音辨識仍存在許多問題，如地域性差異、社會語言學因素與環境與個人特徵等特殊問題[11]，使其在特定場景下的辨識準確度下降。為了解決這些問題，自動語音辨識系統逐步引入深度學習技術持續進化，朝向更高準確率與更廣泛應用的方向發展。

3.5 檢索增強生成 (Retrieval-Augmented Generation, RAG)

檢索增強生成(RAG)是將檢索和生成結合的自然語言處理 (Natural Language Processing, NLP)，其賦予生成式人工智慧模型資訊檢索能力的技術[12]。檢索增強生成最佳化大型語言模型的互動方式，讓模型根據指定的檔案回應使用者的查詢問題，並使用這些資訊增強模型從自身的靜態訓練資料中提取的資訊，並檢索增強生成技術促使大型語言模型能使用特定領域的資[13]。

3.6 人機互動 (Human-Computer Interaction, HCI)

現今科技發達，電腦及相關設備已無處不在，因此對於能支援高效人機互動(HCI)的介面需求日益增加。人機互動其宗旨為於任務與工作情境中，為人類使用而設計、評估和實作互動式計算系統，並研究其相關的主要現象[14]，被視為一門專注於研發理想的使用者介面設計的學科。使用者與機器之間資訊相互往來的介面與環境稱為使用者介面(User-Interface, UI)。使用者透過系統的介面操作與其互動，機器將對於人的各種輸入做出對應動作[21]。

3.7 系統易用性量表

系統可用性量表(System Usability Scale, SUS)是由 John Brooke 於 1996 年發表之量表[15]。其為一種廣泛使用的標準化問卷，用於評估使用者感知的可用性[16]。其標準形式中包含 10 項五級量表題目，並使用正面和反面的題目敘述，如圖一所示。受測者可根據自身使用情況填答，且透過正面和負面的敘述防止受測者未經思考填答，提高量表結果的準確性。

The System Usability Scale Standard Version		Strongly Disagree	Strongly Agree			
		1	2	3	4	5
1	I think that I would like to use this system frequently.		0	0	0	0
2	I found the system unnecessarily complex.		0	0	0	0
3	I thought the system was easy to use.		0	0	0	0
4	I think that I would need the support of a technical person to be able to use this system.		0	0	0	0
5	I found the various functions in this system were well integrated.		0	0	0	0
6	I thought there was too much inconsistency in this system.		0	0	0	0
7	I would imagine that most people would learn to use this system very quickly.		0	0	0	0
8	I found the system very awkward to use.		0	0	0	0
9	I felt very confident using the system.		0	0	0	0
10	I needed to learn a lot of things before I could get going with this system.		0	0	0	0

圖一、標準版系統易用性量表

3.8 現有的相似系統

手持式無線簡報遙控器

手持式無線簡報遙控器為一款讓簡報者在走動時，也能夠操作簡報內容的控制器，大多使用 2.4GHz 無線技術，使得在 10 至 20 公尺範圍內控制連接的電腦和投影幕，並可利用遙控器提供之激光指示器在簡報中指向重要的資訊、圖表和圖形，方便於觀眾閱讀講述內容。雖有幫助講者有效的簡報並高效使用有限的時間[17]，但手持式裝置可能導致簡報時分心或有誤觸的可能性，影響演講的內容及整體效果。

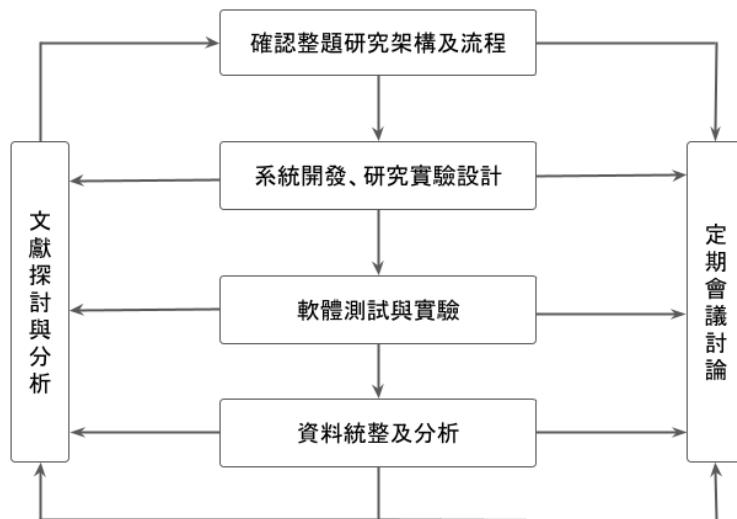
語音聲控簡報錄影系統

語音聲控簡報錄影系統為一款 MSR 語音辨識控制系統的關聯子產品，結合有線與無線麥克風、電腦等語音載具，執行開啟簡報檔、執行程式與簡報換頁放映等特定指令。其雖可取代傳統遙控器、簡報器等設備，但根據本次計劃研究發現，該產品僅能執行特定指令，如開(關)電腦、切換下一頁等固定指令，並無法依照講者當下講述的內容，自動切換至相對應的投影片，並且該系統也無法對簡報進行加注編輯，無法讓觀眾更快速了解講述內容[22]。

(四) 研究方法及步驟

4.1 研究步驟

本研究旨在開發一款結合語音辨識之簡報輔助系統，透過演講者於簡報過程中之講述內容，自動辨識並切換簡報頁面，並可在過程中加入標記註解，讓觀眾更易於閱讀。此計畫之研究步驟規劃和預計時程如下圖二及表一所示。於系統開發前，須先將整體系統之架構設計完善。接著，進行系統開發，並設計可量化的實驗研究，以確保系統最終的準確性與完整性。當系統開發接近完成時，開始測試各項功能，並讓使用者參與簡報測試，量化使用數據，作為後續統整與分析準確性及完整性的依據。此外，在各階段過程中，應安排定期會議，以討論進度與研究方向，並在必要時重新檢視與分析相關文獻與資料。



圖二、研究步驟計畫

執行內容	七月	八月	九月	十月	十一月	十二月	一月	二月
程式架構設計								
系統開發								
軟體測試								
使用者測試及實驗								
資料統整及分析								

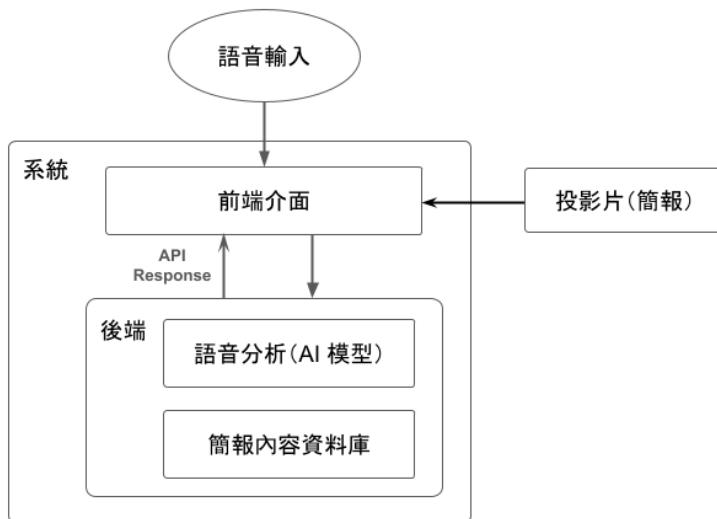
表一、研究計畫甘特圖

4.2 研究方法

本研究預計以 OpenAI 提供的開源的自動語音辨識(Automatic Speech Recognition, ASR)神經網路模型 Whisper 執行語音辨識功能。其優勢可支援多種語言且表現優異，能準確識別語音減少指令錯誤執行的風險。

語音分析將實作邊緣小語言模型(Small Language Models, SLM)，以 LLaMA 3.1 8B 為模型基底，並運用 Google Colab 進行 Unslloth 、 LoRA(Low-Rank Adaptation) 和 QLoRA(Quantized LoRA)訓練和調整參數，將此做為系統的大腦，分析使用者的輸入，並做出準確度極高的決策並調用對應 API 完成回應。

Unslloth 為一款開源的深度學習模型微調工具，其透過使用 OpenAI 的 Triton 重新實施模型計算過程，同時降低記憶體使用量。LoRA 為一款針對模型微調的輕量化方法，它通過將權重矩陣分解為低秩矩陣減少參數更新。QLoRA 為 LoRA 的擴展版本，主要通過將 LoRA 模型的參數量化，使得每個參數的表示所需的數量減少，使得在受限的情況下能處理更大的模型[18]。本次研究將以上述方法實作邊緣小語言模型，提升指令判斷精準度。



圖三、系統架構圖

4.3 系統設計

使用者開啟系統後，系統將自動啟動語音辨識功能，持續監聽並解析使用者的語音輸入內容，如圖四所示。當使用者講述簡報內容或提及特定關鍵字時，系統將即時分析語音語意，匹配相應的投影片頁面，並自動切換至對應的簡報畫面，確保簡報流程的流暢性。此外，系統可支援多種語音指令，如「回到上一頁」、「跳至結論」、「標示記號」等，以提供更直覺且高效的互動體驗，提升簡報的便利性與專業度。



圖四、系統使用示意圖

系統功能說明

1. 簡報及教材上傳

使用者可將演講簡報及該次簡報相關教材上傳系統，做為系統語音分析模型的素材，使演講過程中的指令判斷更精確。

2. 語音識別切換簡報

使用者在演講過程中，透過語音輸入讓系統判別目前最有可能講述的簡報頁面，並根據使用者的語音輸入自動辨識並切換頁面。

3. 自動加入標記功能

使用者在演講過程中，透過語音輸入讓系統自動辨別並標記螢光筆記號與圖形標示等標記功能，讓簡報過程更為生動。

4. 彙整演講摘要

使用者在演講過後，可借助演講時語音辨識紀錄，匯出演講摘要，方便使用者彙整內容。

4.4 研究設計

可量化量表及系統使用過程分析

本計畫將選定一定數量有演講需求的演講者，分別以五種型態的簡報情境實驗，分別為競賽型簡報、演講型簡報、提案型簡報、教學型簡報與會議型簡報，透過不同簡報型態，可呈現不同使用情境中的效果。進行兩次相同內容的演講，第一次演講由講者自行操控簡報，第二次演講使用本系統功能，自動切換簡報頁面，並於過程中紀錄系統判斷頁面準確性及判斷速度，作為後續分析的依據。在兩次演講後，請講者填答使用體驗表單和 SUS(System Usability Scale) 易用性量表，藉助可量化數據，分析對使用者的影響程度和效果。

(五) 預期結果

本計畫旨在開發一款結合語音辨識之簡報輔助系統，透過自動辨識演講者的講述內容切換簡報頁面，並在過程中加入標記註解，使觀眾能更容易閱讀與理解簡報內容。預期研究結果將從系統準確性及指令執行正確性、演講效率與使用體驗三個面向進行探討。

1. 系統準確性及指令執行正確性

在系統準確性方面，本研究將透過對比演講者自行切換簡報頁面與系統自動切換的準確率來評估。本研究預期在語音辨識技術下，系統能夠正確識別多數語音指令，並準確對應至簡報的正確頁面。

本研究將透過測試收集錯誤切換頁面的案例，並分析其可能原因，作為後續改善模型與演算法的參考依據。此外，系統在切換頁面的反應時間亦是重要指標，預期從語音指令發出到頁面切換的時間應控制在 5 秒內，以確保簡報的流暢度。

2. 演講效率提升分析

相較於傳統手動切換簡報，本系統預期能提升演講的整體效率，減少演講者因切換簡報而中斷思路的情形。透過對比演講者手動切換與系統自動切換的演講時間，並提升演講流暢度。

3. 使用者體驗評估

使用者體驗將以問卷調查與訪談評估，並使用 SUS(System Usability Scale)量表衡量系統的可用性與易用性。預計研究結果將顯示多數使用者將能有效減少手動切換簡報的操作負擔，使其更專注於演講內容與聽眾互動。

(六) 需要指導教授指導內容

1. 資料存取整合：系統收集簡報和相關資料整合處理。
2. 語音與文字的轉換整合：將輸入的音訊轉換為文字應用。
3. 簡報附加功能設計：根據需求新增簡報輔助功能。

(七) 參考文獻

國外文獻

- [1] Green, E. P., & Green, E. P. (2021). The basics of slide design. *Healthy Presentations: How to Craft Exceptional Lectures in Medicine, the Health Professions, and the Biomedical Sciences*, 37-62.
- [2] Chou, F. Y.(2019). Action Research on General Education Curriculum for Implementing the University Social Responsibility. *Journal of General Education: Concept & Practice,* ****7*(1), 1-36. [https://doi.org/10.6427/jgecp.201903_7\(1\).0001](https://doi.org/10.6427/jgecp.201903_7(1).0001)
- [3] Morreale, S. P., & Pearson, J. C. (2008). Why communication education is important: The centrality of the discipline in the 21st century. *Communication Education*, *57*(2), 224-240.
- [4] Aziz, K. (1998). The key to perfect presentations. *Industrial and Commercial Training*, *30*(6), 214-217.

- [5] Sweller, J., Van Merriënboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational psychology review*, *10*, 251-296.
- [6] Ballotta, D. (2009). *Public Speaking and Presentations a Critical Review “The Caring Speaker”* (Master's thesis, ISCTE-Instituto Universitario de Lisboa (Portugal)).
- [7] Hoffman, M., & Mittelman, M. (2004). Presentations at professional meetings: notes, suggestions and tips for speakers. *European Journal of Internal Medicine*, *15*(6), 358-363.
- [8] Gelula, M. H. (1997). Effective lecture presentation skills. *Surgical neurology*, *47*(2), 201-204.
- [9] Sweller, J. (2011). Cognitive load theory.
- [10] Yu, D., & Deng, L. (2016). *Automatic speech recognition* (Vol. 1). Berlin: Springer.
- [11] Benzeghiba, M., De Mori, R., Deroo, O., Dupont, S., Erbes, T., Jouvet, D., ... & Wellekens, C. (2007). Automatic speech recognition and speech variability: A review. *Speech communication*, *49*(10-11), 763-786.
- [12] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, *33*, 9459-9474.
- [13] Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., ... & Wang, H. (2023). Retrieval-augmented generation for large language models: A survey. arXiv preprint arXiv:2312.10997.
- [14] Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., & Carey, T. (1994). *Human-computer interaction*. Addison-Wesley Longman Ltd..
- [15] Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry*, *189*(194), 4-7.
- [16] Lewis, J. R. (2018). The system usability scale: past, present, and future. *International Journal of Human–Computer Interaction*, *34*(7), 577-590.
- [17] ERCAN, T., ŞAHİN, Y. G., & BALTA, S. (2009, May). An Indispensable Contributor To Effective Time Management In Class: Remote Presenter. In *Proceedings of 9 th International Educational Technology Conference*.
- [18] Dettmers, T., Pagnoni, A., Holtzman, A., & Zettlemoyer, L. (2024). Qlora: Efficient finetuning of quantized llms. *Advances in Neural Information Processing Systems*, *36*.

國內文獻

- [19] 李世忠、葉盈秀 (2006)。認知負荷與多媒體教材設計之探討。視聽教育雙月刊，47:6=282 民 95.06 ， 1-19。
- [20] 陳湘茹 (2017)。以多媒體簡報融入面積概念教學對四年級學生學習成就與認知負荷之研究。〔碩士論文。國立臺南大學〕臺灣博碩士論文知識加值系統。<https://hdl.handle.net/11296/82q5e2>。
- [21] 林詩曼 (2020)。結合浮力人機互動對話式線上評量之科學探究學習成效探討。〔碩士論文。國立臺中教育大學〕臺灣博碩士論文知識加值系統。<https://hdl.handle.net/11296/egq7r5>。

參考網站

- [22] 語音聲控簡報錄影系統 - 創鴻科技股份有限公司
<http://www.secotek.com.tw/>