結合眼動追蹤之 AI 教學影片實證研究：探討不同語音形式對學習成效之差異

An empirical study combining eye-tracking AI instructional videos: exploring the differences in learning effectiveness caused by different speech formats

林佳陞 [1]，林芝育 [1]，洪浚祐 [2*]
[1] 臺灣嘉義大學行銷與觀光管理學系暨研究所
[2] 臺灣陽明交通大學科技管理研究所
[*] gyo.cyh@gmail.com

【摘要】 隨著生成式 AI (Generative AI)技術的崛起，教學影片之製作已從傳統人力拍攝製作轉向 AI 自動生成之形式，然而其在教育情境中影片形式差異仍缺乏系統性實證。故本研究以多媒體學習認知理論為基礎，採用人聲、AI 類人聲、人聲與 AI 類人聲混合之靜態組間比較實驗設計，探討 AI 生成之知識傳遞型教學影片對學習者之影響。實驗預計透過眼動追蹤技術捕捉視覺注意分配與訊息處理之客觀歷程，探究對學習者學習成效之差異，同時以半結構式訪談輔助探討學習者之主觀感受。研究結果預期可補足多媒體學習認知理論於 AI 生成教學影片與語音設計情境下之實徵缺口。同時提供教學影片語音與字幕形式之參考，讓教育工作者、教材設計者與教育科技產業於生成式 AI 教學內容開發與應用能有所實證依據。

【關鍵字】 多媒體學習認知理論、AI 生成教學影片、AI 語音、眼動追蹤

Abstract: *With the rise of generative AI technology, the production of educational videos has shifted from traditional manual filming to multimodal forms automatically generated by AI. However, the differences in video formats within educational contexts still lack systematic empirical evidence. Therefore, this project, based on multimedia learning cognitive theory, employs a static inter-group comparative experimental design using human voice, AI-generated human voice, and a mixture of both to explore the impact of AI-generated knowledge-transmitting educational videos on learners. The experiment plans to use eye-tracking technology to capture the objective process of visual attention allocation and information processing, exploring the differences in learner learning outcomes. Semi-structured interviews will also be used to explore learners' subjective feelings. The research results are expected to fill the empirical gap in multimedia learning cognitive theory regarding AI-generated educational videos and voice design. Furthermore, it will provide a reference for the forms of audio and subtitles in educational videos, allowing educators, curriculum designers, and the educational technology industry to have empirical evidence for the development and application of generative AI teaching content.*

Keywords: Multimedia learning cognitive theory, AI-generated instructional videos, AI voice, Eye tracking

1. 前言

伴隨著 AI 生成的技術突破，的確大幅改變影片製作與應用情境的生態，教學影片當然也不例外。自 2024 年 Open AI 推出文字轉影片的生成工具 Sora 以來，AI 生成式影片因其高度擬真性與製作效率，快速引起學界與產業的廣泛關注 (Seo et al., 2025)。研究亦指出 AI 驅動之深度學習可顯著提升影像製作效率，降低時間與人力成本，簡化傳統影像製作中繁複的分割、渲染與合成的流程，使影片製作更加快速與普及 (Huang et al., 2023)。同時，相較傳統錄製影片，多模態教學影片能有效提升學習者的記憶保持效果，若能進一步結合眼動追蹤技術，可深入探討學習者於觀看 AI 教學影片時的內在認知歷程 (Xu et al., 2025; Coskun & Cagiltay,

2022)。Netland 等人 (2025)更以管理課程為研究情境，設計四種不同類型的教學影片進行實驗，研究結果顯示多模態教學影片在學習體驗評價上雖略低於傳統人工製作的影片，但差距有限。由此可知，AI 教學影片已具備與傳統影片相近的教育效能，凸顯出在課程中做為教學媒介之可行性與應用潛力。

綜上所述，本研究擬以多媒體學習認知理論作為理論基礎，聚焦於 AI 教學影片在人聲、AI 類人聲及人聲與 AI 類人聲混合，探討之主要研究問題為不同語音呈現方式之 AI 教學影片，是否會對學習者的學習成效產生顯著差異？相較於既有教學影片研究多以人類製作影片為主要研究對象或僅從技術效率使用層面探討 AI 生成內容，本研究進一步將 AI 生成之媒體特性納入學習理論脈絡中進行實徵驗證，並採用眼動追蹤技術作為客觀量測工具，透過眼動系統模型的建立設定視覺關注區域 (Areas of Interest, AOIs)，以數據深入分析學習者在觀看 AI 教學影片時的專注度分配與訊息處理歷程。

2. 文獻探討

Mayer (2002)所提出的多媒體學習認知理論 (Multimedia Learning Cognitive Theory, MLCT)是在認知心理學的基礎下，強調學習者在學習過程中如何處理多重訊息來源，並探討了文字、圖像、影片與語音等多媒體元素交互作用，從而促進或阻礙學習的理解和記憶。後續研究亦說明學習者會積極從多媒體內容中獲取新知的主動加工假設及表示學習者的記憶認知能力具有限度 (Mayer, 2024)。而生成式 AI (Generative AI)的普及也為多模態教材的發展提供了新的契機。生成式 AI 能透過大量資料訓練深度學習模型，產生出具意義的文字、圖像等內容，以回應使用者所給予的指令、提問或複雜提示 (Lim et al., 2023)，透過分析大量資料並深度學習，以力求創造更逼真且符合情境的多媒體內容 (Divya & Mirza, 2024)。

生成式 AI 使得圖像、音效甚至影像的生成技術日新月異，亦使教學影片在視覺與聽覺層面的設計彈性大幅提升，進而凸顯其於教育應用中對學習者認知歷程與注意分配影響的重要性 (Pataranutaporn et al., 2021)。但教材中的影像與聲音生成需高度仰賴專業的設計和縝密的思維，方能有效達成協同效應，在實際應用上也應避免單一通道負荷過重而影響學習歷程 (Shoufan, 2019)。生成式 AI 輔助多模態教材設計之勢興起，使得既有認知不斷被推翻與更新，尤其直接應用多媒體學習認知理論的原則來設計教材之研究也仍屬有限 (AlShaikh et al., 2024)。此一現象亦導致學者無法區分在教學中究竟是科技本身造成學習者的認知過載問題，還是受教材的設計不良所影響 (Candido & Cattaneo, 2025)。

聲音是教學影片的重要元素之一，甚至會直接影響學習者的學習成效 (Lawson & Mayer, 2022)。而隨著生成式 AI 技術的發展，合成聲音已然進入教育，並廣泛應用於教學影片與數位學習內容中，用以取代或輔助傳統的人聲旁白，相關研究亦逐漸關注 AI 聲音於教育情境中對學習的影響 (Xu et al., 2025)。而生成 AI 聲音的門檻大幅下降，使得教材製作時間也能隨之縮減，教育工作者能大規模且快速地生成系列化的聲音內容。然而，目前關於 AI 聲音之研究多集中於人聲與 AI 類人聲的比較。Xu 等人 (2025)比較真人教師錄製影片與具備 AI 形象與聲音之 AI 影片，結果顯示 AI 生成之形象與聲音組合於記憶保留表現上甚至優於真人影片，但在學習成效遷移層面無差異。

3. 研究方法

3.1. 研究架構

本研究採用實驗法作為主要的研究方法，並以 3*1 靜態組間比較之實驗設計，比較在知識傳遞型 AI 教學影片中，採用人聲、AI 類人聲及人聲與 AI 類人聲混合之不同語音形式呈

現，對學習者學習成效之差異。研究對象為對於知識傳遞型影片有興趣之自願學習者。此外，實驗同時也將透過眼動儀蒐集學習者視覺關注區域，了解並比較各組間學習者的視覺在不同的聽覺刺激下之影響。最後亦會設計半結構式問卷訪談。期望以量化與質性資料並行之質量混合法設計，補足單一方法之限制。下圖 1 為本研究之研究架構圖。

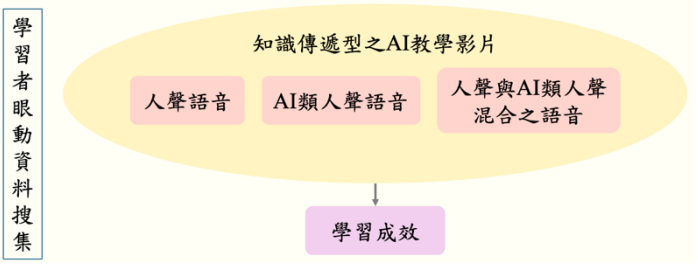

圖 1、研究架構圖

3.2. 實驗流程

為確保本研究實驗設計的嚴謹性與可重複性，實驗組別皆遵循標準化且一致的實驗流程，全程約預計耗時 50 至 60 分鐘。實驗流程共分為準備階段、正式施測階段、質性訪談階段。首先，在準備階段，研究團隊人員將說明實驗目的並請其簽署受試者同意書並進行前測問卷填寫，隨後進入眼動設備校準階段，受試者將依隨機分配之組別就座，並由研究人員引導進行眼動追蹤系統的九點校準，確保視覺關注區域數據採集的精確度。

接著，進入正式施測階段，學習者將開始觀看其所屬組別之教學影片。此過程中系統將同步啟動眼動追蹤採集，持續記錄受試者的注視次數、注視時間與視線移動軌跡等關鍵指標，以客觀捕捉學習者在觀看教學影片時的視覺注意力分布並填寫實驗後測問卷。最後進入質性訪談階段，將針對有意願接受訪談之學習者，了解其在實驗過程中之主觀感受。綜上所述，本研究透過標準化之實驗程序，結合量化指標、眼動數據與質性回饋，以期全面解構不同語音之 AI 教學影片之學習成效差異，下圖 2 為本研究之實驗流程圖。

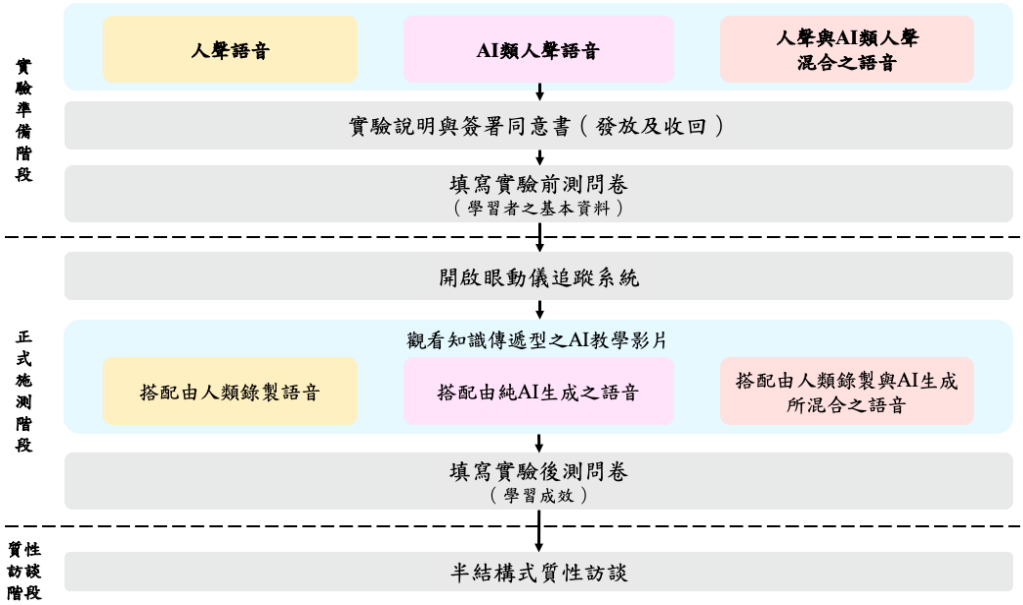

圖 2、實驗流程圖

4. 本研究小結

本研究尚在持續進行中，期望透過不同語音形式融合眼動追蹤技術的介入，解構學習者在雙通道資訊處理中的注意力分配規律，最終結合質性訪談詮釋學習者之主觀感受，進而建置具實證基礎的 AI 教學影片評估基準與設計指引。

參考文獻

AlShaikh, R., AlMalki, N., & Almasre, M. (2024). The implementation of the cognitive theory of multimedia learning in the design and evaluation of an AI educational video assistant utilizing large language models. *Heliyon, 10*(3). https://doi.org/10.1016/j.heliyon.2024.e25361

Candido, V., & Cattaneo, A. (2025). Applying cognitive theory of multimedia learning principles to augmented reality and its effects on cognitive load and learning outcomes. *Computers in Human Behavior Reports, 18*, 100678. https://doi.org/10.1016/j.chbr.2025.100678

Coskun, A., & Cagiltay, K. (2022). A systematic review of eye-tracking-based research on animated multimedia learning. *Journal of Computer Assisted Learning*, *38*(2), 581-598. https://doi.org/10.1111/jcal.12629

Divya, V., & Mirza, A. U. (2024). Transforming content creation: The influence of generative AI on a new frontier. *In Exploring the frontiers of artificial intelligence and machine learning technologies* (p. 143). Blue Hill Publications. https://doi.org/10.59646/efaimltC8/133

Huang, Y., Lv, S., Tseng, K. K., Tseng, P. J., Xie, X., & Lin, R. F. Y. (2023). Recent advances in artificial intelligence for video production system. *Enterprise Information Systems*, *17*(11), 2246188. https://doi.org/10.1080/17517575.2023.2246188

Lawson, A. P., & Mayer, R. E. (2022). The power of voice to convey emotion in multimedia instructional messages. *International Journal of Artificial Intelligence in Education, 32*(4), 971-990. https://doi.org/10.1007/s40593-021-00282-y

Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., & Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *The international journal of management education*, *21*(2), 100790. https://doi.org/10.1016/j.ijme.2023.100790

Mayer, R. E. (2002). Multimedia learning. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 41, pp. 85–139). Academic Press. https://doi.org/10.1016/S0079-7421(02)80005-6

Mayer, R. E. (2024). The past, present, and future of the cognitive theory of multimedia learning. *Educational Psychology Review*, *36*(1), 8. https://doi.org/10.1007/s10648-023-09842-1

Netland, T., von Dzengelevski, O., Tesch, K., & Kwasnitschka, D. (2025). Comparing human-made and AI-generated teaching videos: An experimental study on learning effects. *Computers & Education*, *224*, 105164. https://doi.org/10.1016/j.compedu.2024.105164

Pataranutaporn, P., Danry, V., Leong, J., Punpongsanon, P., Novy, D., Maes, P., & Sra, M. (2021). AI-generated characters for supporting personalized learning and well-being. *Nature Machine Intelligence, 3*(12), 1013–1022. https://doi.org/10.1038/s42256-021-00417-9

Shoufan, A. (2019). Estimating the cognitive value of YouTube's educational videos: A learning analytics approach. *Computers in Human Behavior*, *92*, 450-458. https://doi.org/10.1016/j.chb.2018.03.036

Xu, T., Liu, Y., Jin, Y., Qu, Y., Bai, J., Zhang, W., & Zhou, Y. (2025). From recorded to AI-generated instructional videos: A comparison of learning performance and experience. *British Journal of Educational Technology*, *56*(4), 1463-1487. https://doi.org/10.1111/bjet.13530