

應用視覺化程式語言與人工智慧實現創新人機互動介面

Using Visual Programming Language and Artificial Intelligence to Develop an Innovative Human Machine Interaction Interface

黃立武¹、陳響亮²、李宜靜³、黃兆平³、董塘筠¹、陳品修¹

¹ 國立成功大學 製造資訊與系統研究所 研究生

² 國立成功大學 製造資訊與系統研究所 教授

³ 工研院機械所 控制核心技術組 機電控制整合部 副研究員

摘要：機台運動控制程式之編輯往往需要艱深的專業知識，且控制程式之程式語言需要專業背景才能閱讀。為了降低編輯運動控制程式之專業門檻，本文針對 PC-Based 控制器之六軸機械手臂與生產線周邊設備之運動控制進行研究，設計出一套視覺化程式語言 (Visual Programming Language, VPL)，並透過此 VPL 實作出一套人機互動介面，使用者可透過拖曳、連結不同功能方塊之方式，實現生產線控制流程之建立與調整。為更進一步降低系統之操作門檻，本系統導入深度學習、電腦視覺、語音識別、眼動追蹤 (eye tracking) 等技術，開發基於手勢、語音、視線等特徵之人機互動功能，實現創新人機互動 VPL 編程機制。藉由上述機制之輔助，使用者可透過較為直覺的方式完成控制程式之編輯。

Abstract : The editing of machine motion control programs often requires professional knowledge, and the programming language of the control program is difficult to read. To reduce the professional threshold for editing motion control programs, this article focuses on the design and development of a human machine interface by using the visual programming language (VPL) which can be applied for the performance testing of the performance of the developed system. The testing examples include the motion control of the six-axis robotic arms with PC-Based controllers and the control of production line peripherals. Through the VPL, a human-computer interaction system is developed, which allows users to establish and adjust the production line control flow by dragging and linking different functional icons. To further reduce the operating threshold of the system, deep learning, computer vision, voice recognition, eye tracking and other technologies are also introduced to develop the human-computer interaction mechanism based on gesture, voice, and vision features. With the help of the above mentioned mechanism, users can realize the editing of production line control programs in a more intuitive way.

關鍵詞：視覺化程式語言、運動控制、人工智慧、語音辨識、手勢辨識

Keywords : Visual programming language, Motion control, Artificial intelligence, Voice recognition, Gesture recognition

前言

在工業環境中，機械手臂或各機台之運動控制有著相當高的技術門檻，大多須由受過專業訓練之工程師才能負責此項工作。目前許多機器人大廠具有各自獨特的機器人控制語言，導致市面上運動控制程式語言的種類繁多，讓不具程式編輯經驗與非專業領域的使用者必須付出更多時間與心力，方能熟悉並運用不同的系統；Alexandrova 等人之研究 [1] 亦顯示，市售的機械手臂軟體環境仍過於封閉，針對不同的系統環境與日益複雜的應用情境，機械手臂的運動程式將更要求由專業的程式人員撰寫。為改善上述問題，本文將基於 VPL[2] 設計一視覺化人機互動介面，並於其中整合工業機械手臂與生產線周邊設備之控制功能，以期降低智慧製造領域使用者編輯運動控制程式之門檻。相較於文字形式之程式，VPL 能夠讓使用者不再受限於程式語法，只需排列、連結功能區塊即可實現程式執行流程的設計，將有助於降低程式編輯之難度並使程式更易於理解。且 VPL 各程式功能獨立呈現，新增功能時不必修改舊的程式方塊，僅須創建新方塊便能完成。該特性使 VPL 之擴展性高於一般文字程式語言，並有更高的自由度進行功能之擴展。

在傳統的製造環境裡，人機互動 (HMI) 多是透過教導器、按鈕或圖形化使用者界面完成，且大多只有受過專業訓練的工程師才知道該如何正確控制；而在未來的趨勢中，HMI 正朝著以人為本 (Person-Centered) 的複合控制模式發展，期望人與機器人可以使用更類似人與人之間的直覺方式互動 [3]。隨著感測技術的發展，人們與電子設備的交互方式正逐漸改變。除了傳統人機交互設備 (如鍵盤、滑鼠、觸控式介面等) 之外，手勢、語音、眼動追蹤、臉部表情、腦波、觸覺等資訊皆可做為 HMI 之輸入訊號。若將人與人溝通時經常採用的手勢、語言甚至眼神做為智慧製造場域中人機互動之媒介，將有助於降低初學者操作複雜設備時的負擔。藉由易於上手之人機互動功能，操作者可透過更為人性化的介面傳達需求給製造

設備，不再需要牢記複雜的機台操作指令或設備控制方式。

本文提出之「創新人機互動介面」基於 PC-Based 控制器進行開發，並導入機械手臂、CNC、輸送帶等製程設備之控制功能，同時採用視覺化程式語言作為使用者編程之方式，讓使用者得以跳脫複雜且難懂的文字程式語言編輯，可以用更直觀的方式進行機台之控制與生產線之調整。同時為強化使用者與機械設備協作之實用性與便利性，本文之人機互動介面結合電腦視覺、深度學習、自動語音識別 (Automatic Speech Recognition, ASR)、眼動追蹤等技術，於系統中加入創新之手勢、語音及視線人機互動功能，讓操作者能以更直覺方式進行 VPL 程式之編輯，並能以更直觀之方式完成機台之運動操作或軌跡教導，實現降低機台操作門檻之目的。

人機互動介面總體設計

本文之人機互動介面之系統模組設計如圖 1 所示，此系統分為實體設備、系統前端與系統後端三部分。實體設備包含用於實現產線運作的工業機械手臂、輸送帶、I/O 控制卡，以及用於實現 HMI 之 IP 網路攝影機、指向性麥克風與眼動追蹤儀；系統前端負責實現視覺化程式之各項編輯機制，包含「VPL 控制介面模組」、「VPL 功能圖示模組」、「HMI 功能模組」；系統後端則透過混合式編譯器，解析使用者於前端設計之 VPL，並執行對應之機械設備控制功能或應用功能。

而系統前端之介面設計如圖 2 所示，可分為「工具列」、「工具箱」、「VPL 編輯頁面」、「功能視窗」四個主要區塊，於圖片中依序標示為數字 1 到 4，以下將對各區塊之功能進行詳細說明。

- (1) 工具列：主要用於儲存及開啟 VPL 程式及顯示各類功能子視窗。為使整體操作流程更加流暢，本文中將機器手臂的初始化流程整併進入人機互動介面中，透過「工具」列表的「手臂控制器」開啟機械手臂的設置頁面。「執行」子視窗負責 VPL 之編譯與執行，於 VPL

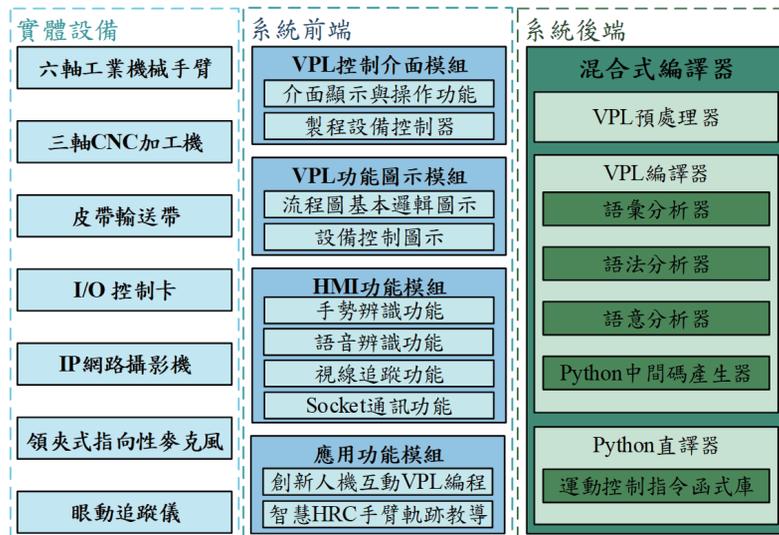


圖 1 人機互動介面系統模組設計圖

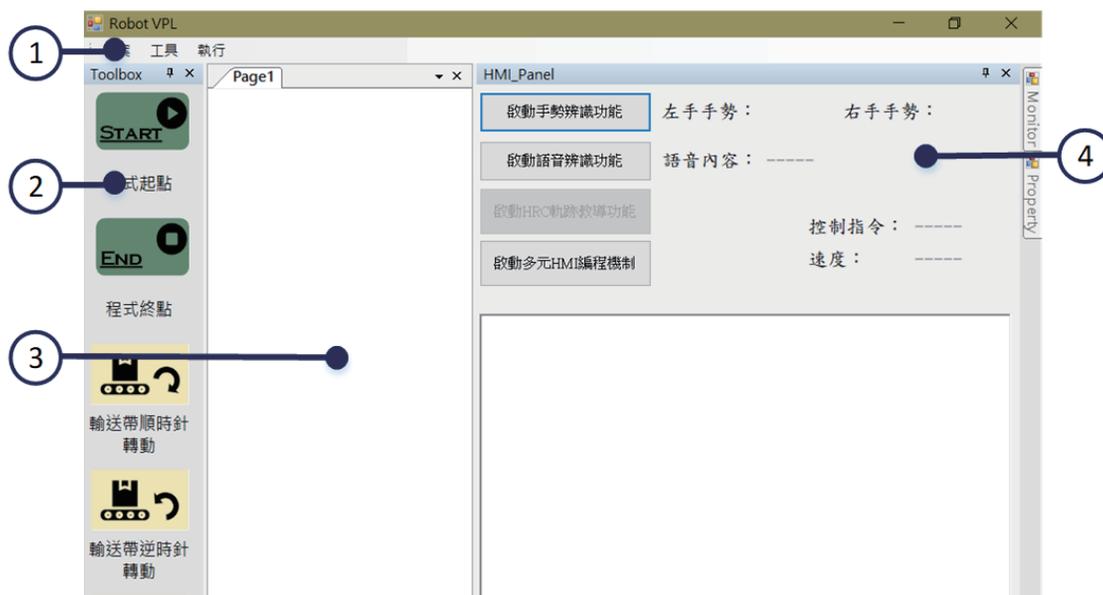


圖 2 VPL 之使用者介面設計

程式編輯完成後，可透過該功能完成運動程式之編譯並開始執行，如圖 3 所示。

- (2) 工具箱：此視窗以圖案搭配文字說明之形式，顯示系統中所有可用的功能圖示供使用者選擇。使用者可透過拖曳方塊之方式，將 VPL 之功能方塊拖曳至圖 2 中③之 VPL 編輯頁面，以進行 VPL 程式之撰寫。目前本文所開發之

VPL 程式包含圖 4 所示之功能圖示可供使用。

- (3) VPL 編輯頁面：從工具箱中拖曳並創建之功能圖示將顯示於此編輯頁面，使用者可對其進行排列、連結等操作。於此處點擊滑鼠右鍵便可叫出操作選單，可透過操作選單完成功能方塊之連結、清除頁面或儲存 VPL 等操作，如圖 5 所示。

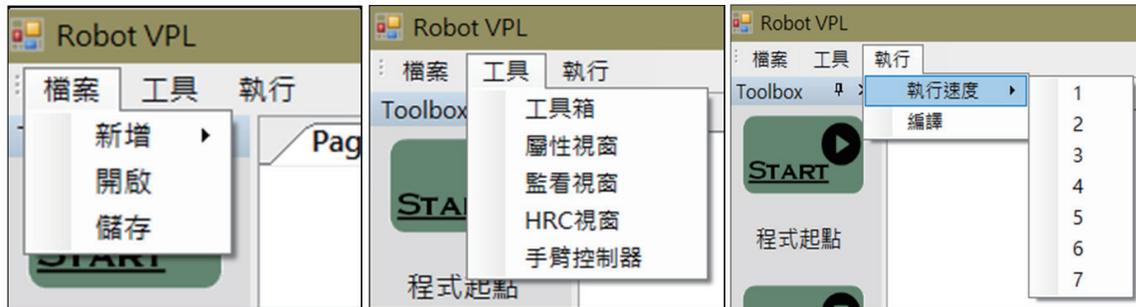


圖 3 工具列之詳細選單內容



圖 4 VPL 功能圖示一覽

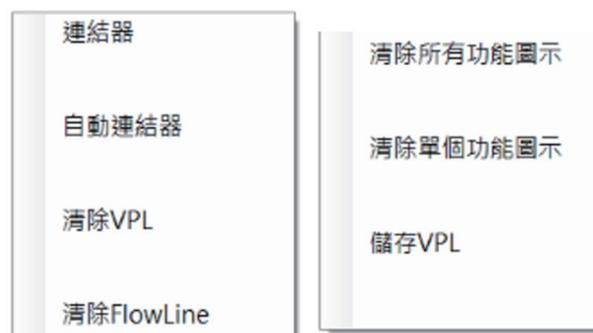


圖 5 VPL 編輯頁面之右鍵選單詳細內容

(4) 功能視窗：此視窗可用於切換顯示「人機協作視窗」、「功能圖示屬性視窗」及「VPL 監看視窗」。「人機協作視窗」中包含開啟人機協作功能之按鈕，並於此處顯示手勢辨識之結果以及操作提示資訊。「功能圖示屬性視窗」供使用者確認當前 VPL 功能圖示之詳細屬性及數值設定。「VPL 監看視窗」則

以樹狀方式呈現當前 VPL 程式之架構，供使用者快速檢查 VPL 程式，如圖 6 所示。

開啟人機協作功能後整體畫面將如圖 7 所示，使用者對網路攝影機比出手勢，控制左下圖中之機器手臂，同時手勢辨識結果與對應機器手臂動作將顯示於「人機協作視窗」(圖 7 右)中。



圖 6 (a) 功能圖示屬性視窗 (b) VPL 監看視窗



圖 7 實際操作圖

人機互動功能之實現

本文之人機互動介面採用多項人工智慧與深度學習之技術以實現智慧人機互動功能，主要包含「語音辨識」、「手勢辨識」及「眼動追蹤」，各技術之詳細敘述如下：

- (1) 語音辨識：由於市面上自動語音辨識技術數量繁多，本文考量製造場域實際狀況，以辨識率及導入系統難易度作為選擇之主要評估因素。依據 Kěpuska 等人 [4] 之研究，其針對 CMU Sphinx-4、Microsoft Speech API 以及 Google API 等主流語音辨識技術進行測試，得出之字錯率 (Word Error Rate, WER) 分別為 37%、18% 和 9%，與商業標準 WER 之 20% 相比 [5]，可發現後兩者之字錯率皆處於可接受之範圍。但由於 Google API 需仰賴網路連接且按次收費，考慮到製造場域往往為網路封閉之空間，本文最後採用可於本地端執行之 Microsoft Speech API。同時利用 Microsoft Speech API 之自定義語音詞彙庫功能，事先將會使用到的語音指令先行輸入詞彙庫，進而限定語音辨識之搜尋範圍，雖會降低語音辨識之靈活性，但可有效地提高整體辨識正確率，且能避免讓系統接收到非預期之指令內容。
- (2) 手勢辨識：將語音指令用來控制家用環境中的機器人已十分普遍，但在工業製造環境中採用語音控制機台仍有其使用上的限制。由於下達語音指令所需時間較長，導致語音指

令僅適合進行個別指令的呼叫、不適合用於連續的遙控。因此，許多學者都指出採取手勢下達指令在機器人的連續控制中將是較佳的做法 [6-7]。

手勢辨識方法大多可分類為兩類，分別為使用穿戴式裝置，及使用電腦視覺相關技術。相較於使用穿戴式設備，基於電腦視覺的方式可避免複雜、耗時的傳感器穿戴步驟，更換使用者時亦不需重新進行校準，可減少人機互動過程中的不適感，因此本文最終決定以基於電腦視覺之方法實現手勢辨識功能。

為了實現接近即時 (Real-Time) 的影像檢測，檢測速度是本文選擇深度學習模型時重要的考量因素。而在各種辨識模型中，YOLOv4 不僅有良好的辨識準確度，同時還具備十分出色的影像處理速度，因此本文選擇物件檢測模型 YOLOv4 作為辨識人機手勢互動之模型。

實現自定義手勢辨識 YOLOv4 模型需完成四個重要環節，分別為蒐集數據、標註數據、訓練模型以及評估模型。本文所定義之手勢包含下列幾項，由於現有的開源數據庫中並無適合本文應用需求之手勢數據集，為了收集足量的目標手勢圖片以進行訓練，本文錄製研究人員比出目標手勢之影片，過程中並透過不時移動或旋轉手腕來改變手的位置，以增加資料集之變異性。影片錄製結束後以每秒 30 幀的頻率將影像轉為圖像，最終獲得每種手勢類別約 3,750 張、總數約 52,500 張的圖像數據集，如圖 8 所示。

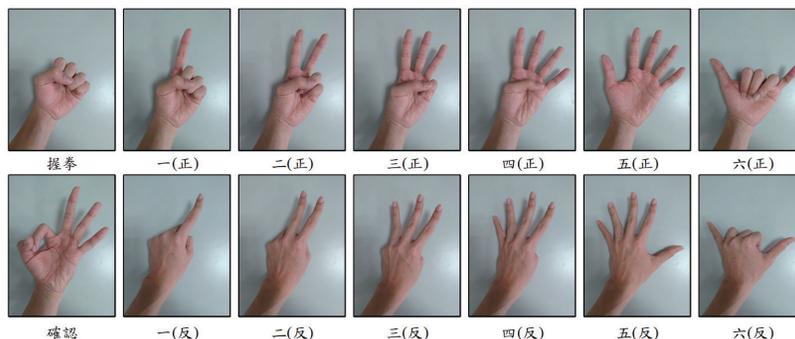


圖 8 手勢一覽圖

```

<?xml version="1.0" encoding="UTF-8"?>
<annotation>
  <filename>101_00001</filename>
  <size>
    <width>640</width>
    <height>480</height>
  </size>
  <object>
    <id>obj_000001</id>
    <type>gesture0</type>
    <segmentation>None</segmentation>
    <behavior/>
    <bndbox>
      <xmin>190</xmin>
      <ymin>120</ymin>
      <xmax>274</xmax>
      <ymax>201</ymax>
    </bndbox>
  </object>
</annotation>

```

圖 9 手勢資料集 XML 檔案之詳細內容

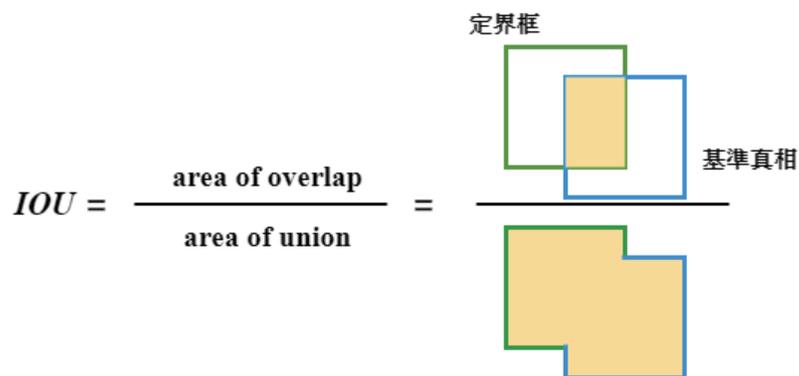


圖 10 IOU 之定義示範圖



圖 11 手勢辨識模型之測試圖片

在標註數據之階段，本文採用線上影片高效標註工具 ezLabel [8] 進行快速的手勢資料集標註 (annotation)。ezLabel 讓使用者只需標註影片中的第一幀和最後一幀，即可自動完成剩餘影像中該物件之框選及類別選取，取代傳統方式中須逐一標註每張圖片之繁瑣過程。完成資料集標註後，輸出之 XML 文件如圖 9 所示。於該標註文件中包含訓練模型時所需之各項資訊，如影像名稱、影像尺寸、物件所屬類別名稱以及物件定界框 (bounding box) 座標等，完成標註之 XML 資料集即可投入 YOLOv4 模型進行訓練。

訓練完成之模型需進行驗證，以確保其實用性及正確性。由於物件偵測技術不僅需要分析影像中是否有出現目標物件，同時亦需要判斷物件出現之位置，因此無法僅使用模型之準確率 (accuracy) 來評估該模型之好壞。本文透過測量定界框和基準真相 (ground truth) 方式，計算預測結果 Intersection over union (IOU)，以作為預測正確率之判斷依據。圖 10 中左上方框代表定界框，即模型之預測結果；右下方框代表基準真相，即真實情況中手的所在位置，IOU 計算兩者的重疊面積除以聯集面積，若 IOU 超過所設定之閾值，則該測量結果為 TP，即預測結果與真實結果一致。

本文於模型完成訓練後，以手勢影像資料集中的測試集進行模型之驗證。IOU 值設定為 0.5。

圖 11 為一次測量結果之示例，其定界框與基準真相分別如圖中所示。於此張示意圖中，IOU 之計算結果為 93.54%，大於設定之閾值 0.5，因此判定本手勢模型對此張圖之辨識結果為 TP，模型驗證之預測結果大致可與現實情況相符。

- (3) 眼動追蹤：本文將眼動追蹤儀設置於控制系統之伺服端螢幕下方，用於偵測系統使用者之凝視點 (Gaze point) 資訊，並以此取代傳統編程過程中必要的滑鼠操控過程。不同於基於頭戴式裝置的方式，本文設計之眼動追蹤功能採用非穿戴、遙測式之眼動追蹤儀 Tobii Eye Tracker 5 實現。此眼動儀系統內建用於 2D 螢幕之校準程式，並擁有頭部追蹤功能，

使眼動追蹤之效果更為穩定，適合導入基於螢幕視窗的交互應用程式之中。

測試案例 —— 創新人機互動 VPL 編程機制

創新人機互動 VPL 編程機制作為使用者編程 VPL 之輔助機制，將眼動追蹤技術及手勢、語音辨識技術融入 VPL 介面之中，讓使用者不須操作滑鼠及鍵盤，即可編輯控制生產線設備之 VPL。使用者啟動此創新編程機制後，程式將透過眼動追蹤功能，獲取使用者凝視點數據以移動滑鼠游標。當游標移動至使用者欲操作之物件上時，使用者可朝向螢幕上方之網路攝影機比出手勢，以觸發該手勢對應之滑鼠事件；此外，本機制採用語音指令實現自動設定 VPL 屬性之功能，使用者以視線搭配手勢選擇欲設定的功能圖示後，可透過預設的語音指令完成特定功能圖示之設定，手勢一覽與對應滑鼠功能如表 1。

測試案例中，首先進行 VPL 功能圖示之建立。透過視線將游標移動至所需的圖示上之後，以右手手勢零按下滑鼠左鍵，而後將視線移動至 VPL 編輯視窗內，轉換手勢為右手手勢五以放開滑鼠左鍵，完成 VPL 功能圖示之拖放與建立，同時也可藉由該方式移動 VPL 編輯視窗內的功能圖示，讓功能圖示可重新排列。完成排列後，即可以右手手勢二呼叫出滑鼠右鍵選單，以「連結器」或「自動連結器」功能完成圖示之間的連線，如圖 12 所示。

若程式中包含「選擇結構 (if-else)」或「循環結構 (while, for)」之功能圖示，則需針對該功

表 1 創新人機互動 VPL 編程機制支援手勢與滑鼠事件對照表

手勢指令類別	對應滑鼠事件
右手手勢零	滑鼠左鍵長按
右手手勢一	滑鼠左鍵點擊
右手手勢二	滑鼠右鍵點擊
右手手勢三	滑鼠滾輪前進
右手手勢四	滑鼠滾輪後退
右手手勢五	滑鼠左鍵放開

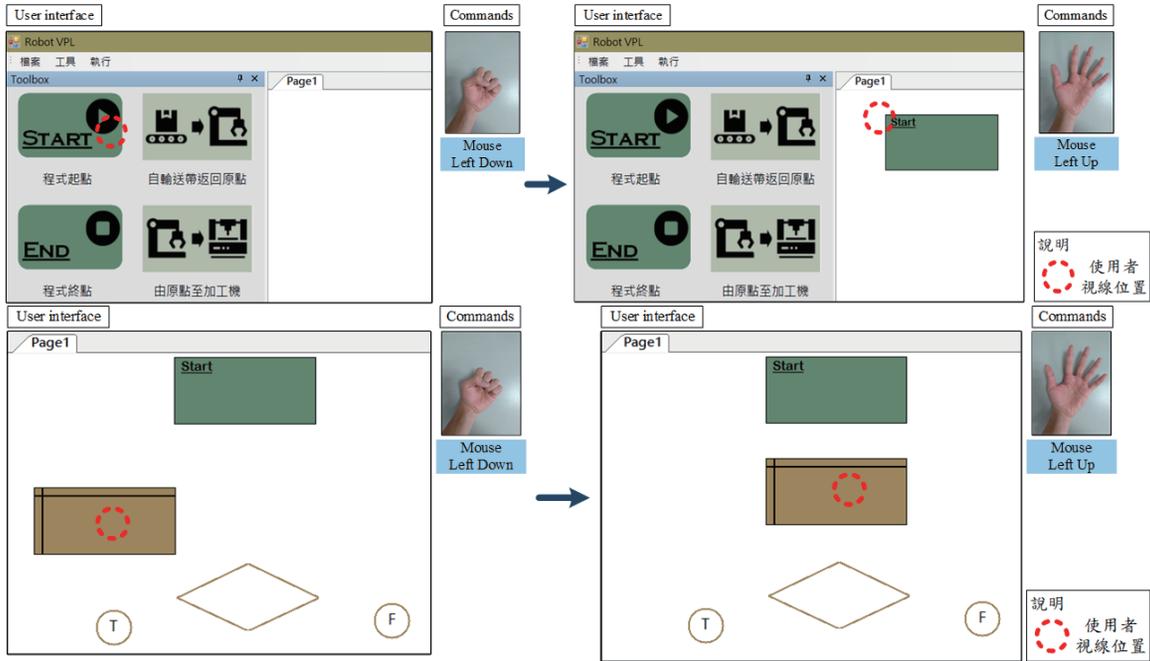


圖 12 以手勢實現 VPL 功能圖示之創建與拖放

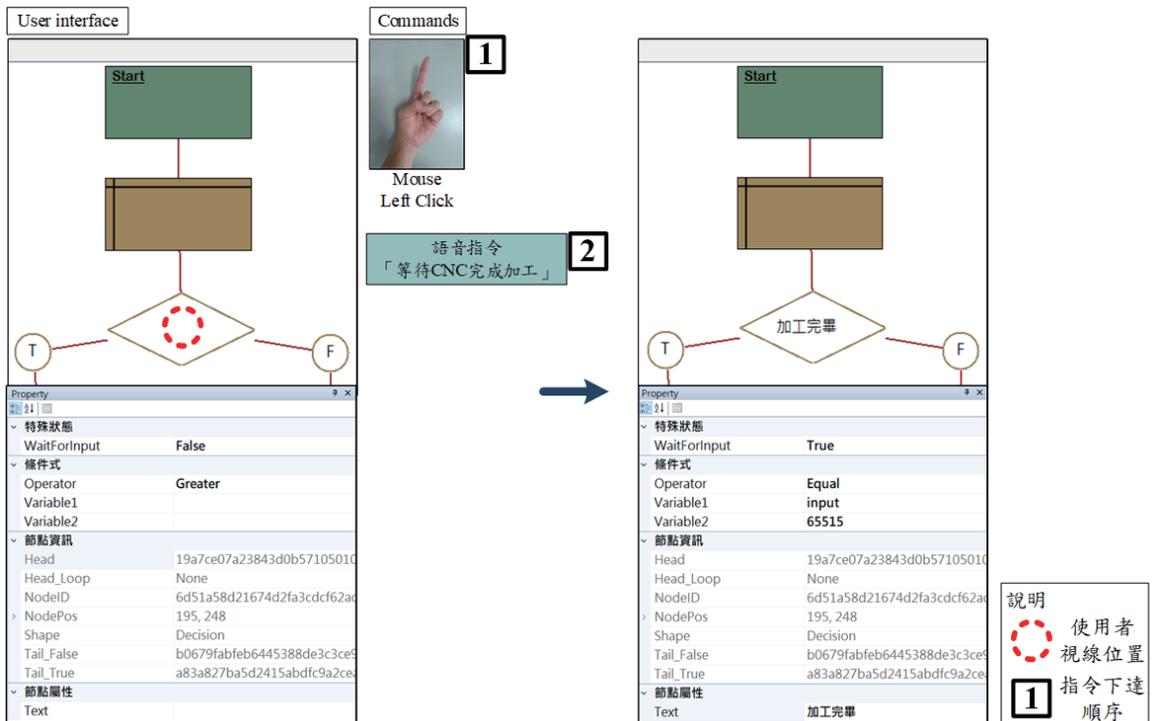


圖 13 以語音指令設定 VPL 功能圖示之屬性

能圖示進行屬性設定，使用者可透過語音指令完成此步驟。於操作案例中，目標 VPL 程式需等待 CNC 加工完成才能執行後續指令。使用者須將視線移動至欲設定屬性之功能圖示上方，並以右手手勢一進行該圖示之選取，接著以口頭下達「等待 CNC 完成加工」之指令，即可完成該圖示屬性之設定，如圖 13 所示。

至此便可完成一套 VPL 程式之編輯，接著可透過工具列之編譯功能，完成運動控制程式之編譯並執行，或將該 VPL 儲存以供後續使用。

結論

本文所提出之視覺化人機互動介面採用 VPL 作為運動程式編輯之基礎，讓使用者可跳脫困難的文字程式編輯，降低非專業領域使用者編輯運動程式之技術要求。同時採用人工智慧之技術，將語音辨識及手勢辨識等技術融入工業情境中，使用者能空出雙手且無須配帶感測器的狀況下，以直覺的手勢、語音搭配視線進行 VPL 編程，達到降低系統 VPL 編輯難度之目的。未來之發展方向可透過新增 VPL 之方塊圖示種類，擴大本系統可應對之製造情境，亦可針對手勢及語音指令之種類進行調整，使本系統能更加符合所需之情境需求。

誌謝

本計畫承工研院機械與機電系統研究所的支持（計畫編號 B109-F007、B110-F011），並承行政院科技部提供研究經費（計畫編號 MOST109-2221-E006-098、MOST110-2221-E006-174），計畫得以順利進行，特此致謝。

參考文獻

- [1] S. Alexandrova, Z. Tatlock, and M. Cakmak, "RoboFlow: A flow-based visual programming language for mobile manipulation tasks," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, 5537-5544, 2015.
- [2] M. Erwig, K. Smeltzer, and X. Wang, "What is a visual language?," *Journal of Visual Languages & Computing* 38(6), 9-17, 2016.
- [3] Ilias El Makrini, et al. "Working with walt: how a cobot was developed and inserted on an auto assembly line," *IEEE Robotics & Automation Magazine* 25(2), 51-58, 2018.
- [4] V. Kěpuska and G. Bohouta, "Comparing speech recognition systems (Microsoft API, Google API and CMU Sphinx)," *Int. J. Eng. Res. Appl* 7(3), 20-24, 2017.
- [5] Microsoft Docs, "評估和改善自訂語音精確度," 2022. [Online]. Available: <https://docs.microsoft.com/zh-tw/azure/cognitive-services/speech-service/how-to-custom-speech-evaluate-data>
- [6] Y. R. Oh, J. S. Yoon, J. H. Park, M. Kim, and H. K. Kim, "A name recognition based call-and-come service for home robots," *IEEE Transactions on Consumer Electronics* 54(2), 247-253, 2008.
- [7] M. Jain, M. Sehgal, Y. Kalra, P. Jain, N. Aggarwal, R. Singh, D. Kumar, and T. K. Bera, "Object detection and gesture control of four-wheel mobile robot," in *2019 International Conference on Communication and Electronics Systems (ICCES)*, 303-308, 2019.
- [8] ezLabel, <https://www.aicreda.com>