

一圖勝千言

指導老師：秦群立老師

學校名稱：中山醫學大學

張盛文 林嘉俊 陳俊任 陳素娟 張詠翔

競賽類別：資通類

摘要：

在這個以口語溝通為主的社會中，聽障人士由於**聽力受損**，必然會遇到許多不便的情況，如：面臨溝通障礙、學習障礙、社交障礙與就業困難等問題，進而**影響其生活品質**。雖然他們可以藉由手語、讀唇語、筆談或其他替代性溝通方式來理解他人的想法，但仍有許多限制，導致**聽障人士在日常生活中與家人、朋友和同事之間難以進行交流**。因此，為了輔助聽障人士更好地與他人溝通，本團隊開發出「一圖勝千言」系統，此系統是結合「語音轉文字」的技術與我們自行開發的「關鍵字抓取演算法」和「關鍵字搜圖演算法」，並搭配 AR 智慧眼鏡，將聽到的對話轉換成摘要式內容與相關的圖片即時顯示在 AR 智慧眼鏡上，讓**聽障人士可以實時「看到」對話**，進而輔助聽障人士能更容易理解他人想表達的內容，以減少溝通障礙，期望透過本系統能協助聽障人士**增進溝通的便利性**，增強其自信心和社交能力，並**改善生活品質**，進一步促進其**融入社會**和參與社會活動。

關鍵詞：聽障人士、溝通障礙、語音轉文字、AR 智慧眼鏡、摘要式內容。

一、前言

根據 2021 年衛生福利部統計處資料顯示，在台灣約有 12 萬人為聽覺機能障礙者，其中有 16.4% 的人為重度聽覺障礙[1]，這類族群常因自身聽覺的缺陷而對聲音的反應較為遲緩，在日常生活中必然會遇到許多不便的狀況，如：面臨溝通障礙、學習障礙、社交障礙與就業困難等問題，進而影響其生活品質，其中以溝通障礙最為不便，常讓聽障人士難以理解他人說話涵義而放棄彼此之間的溝通[2]。雖然社會上有許多協助解決聽障人士溝通困難的方法，諸如配戴聽力輔助裝置、使用手語或是識讀唇語等，但在這些多樣的方法中不乏缺點，像是配戴助聽器對於中重度聽覺障礙人士較無幫助；而使用手語則會因為與一般語言的架構不同，容

易引起誤會[3]，此外，唇語也是和聽障人士溝通的管道之一，但近年因疫情的關係，多數人在與人交流互動時都配戴著口罩，導致聽障人士無法透過口型識讀談話內容，增加了識讀唇語的難度[4]。為解決上述的問題，本團隊開發了發出「一圖勝千言」系統，透過 AR 眼鏡配合深度學習模型將與他人談話內容的文字呈現於使用者眼前，使其不必再透過配戴助聽器或看手語等方式來接收訊息，幫助聽障人士與他人溝通可以更加順利，並促進其融入社會和參與社會活動。

二、研究目的

為了使聽障人士能得知對話內容，本系統透過「語音轉文字」將使用者周遭對話的語音轉換成文字並顯示於眼鏡鏡片之上。並且，本系統會將語音轉出的文字透過「關鍵字抓取演算法」自動抓取關鍵字，以幫助使用者抓取他人說話的重點。接著，本系統透過「關鍵字搜圖演算法」查詢與關鍵字相關之圖片，一併呈現於眼鏡鏡片上，讓使用者更容易懂他人所想表達之事物。

本團隊運用「語音轉文字」、「關鍵字抓取演算法」及「關鍵字搜圖演算法」三項技術結合 AR 眼鏡開發出「一圖勝千言」系統來幫助聽障人士能夠在與人溝通時輕易了解他人想表達的內容，「一圖勝千言」系統示意圖如圖 1 所示。

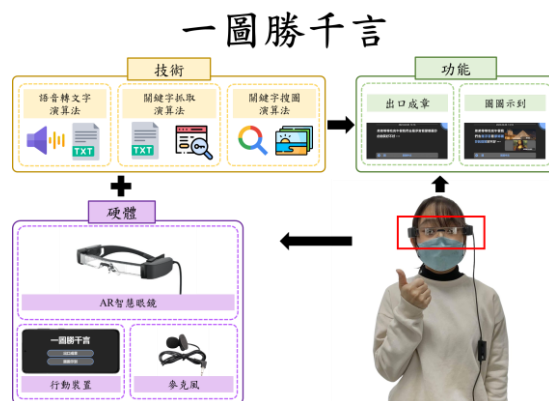


圖 1、「一圖勝千言」系統示意圖

三、原理與分析

為了使「一圖勝千言」系統得以運作，本團隊使用了三項技術，分別為**語音轉文字演算法**、**關鍵字抓取演算法**及**關鍵字搜圖演算法**。以下將詳細說明上述三項技術之原理。

A. 語音轉文字演算法

為了實現「出口成章」之功能，本團隊開發語音轉文字演算法，將麥克風作為系統之輸入裝置，使用自行開發的 speech2text 模型將使用者透過麥克風擷取之語音轉換成文字，並顯示於「一圖勝千言」AR 眼鏡的鏡片之上。

B. 關鍵字抓取演算法

為了擷取使用者身邊談話內容的關鍵字，本團隊設計了關鍵字抓取演算法深度學習於抓取關鍵字，在模型訓練時使用 Chinese-BERT 作為模型預訓練權重，並以 Cross Entropy 作為損失函數，最後將訓練完成的模型部署在使用 Django 建立之網站伺服器中，將**語音轉文字**之輸出結果作為模型輸入，透過**關鍵字抓取演算法**自動抓取文字中的關鍵字並輸出。

C. 關鍵字搜圖演算法

為了使用關鍵字查詢圖片，本團隊設計了關鍵字查詢圖片演算法，將**關鍵字抓取演算法**輸出之關鍵字結果作為此演算法之輸入，自動於網路上進行圖片搜索，最後將結果透過智慧眼鏡呈現給使用者。

四、軟體系統

「一圖勝千言」系統是由軟體與硬體所組成，硬體系統結合了智慧眼鏡、行動裝置及麥克風等硬體，而軟體系統使用 WebView 技術進行開發，並運用本團隊設計之**語音轉文字**、**關鍵字抓取演算法**及**關鍵字搜圖演算法**三項技術，實現本系統的各項功能，以下將對軟體及硬體做詳細的分析與介紹。「一圖勝千言」系統架構圖如圖 2 所示。

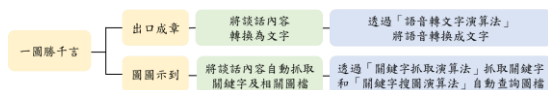


圖 2、「一圖勝千言」系統架構圖

A. 軟體

軟體功能共分為兩個部分，分別為「出口成章」以及「圖圖示到」，以下將針對這兩大功能進行介紹。

I. 出口成章

為了讓使用者能夠看到與他人談話之內容，本團隊設計了「出口成章」功能，麥克風

會將聲音收取並傳送至手機，並透過本團隊設計的**語音轉文字**，將收取到的聲音轉換為文字，並呈現給使用者。「出口成章」功能流程圖如圖 3 所示。

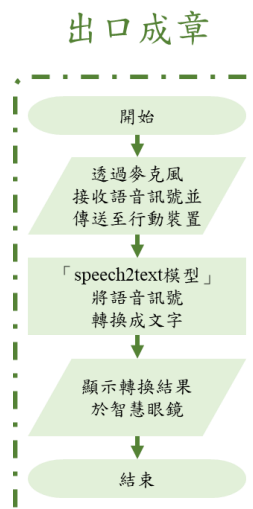


圖 3、「出口成章」功能流程圖

II. 圖圖示到

為了讓使用者更了解與他人的談話內容，本團隊設計了「圖圖示到」功能，將「出口成章」功能得出的文字透過本團隊設計之**關鍵字抓取演算法**技術，將重要的關鍵字擷取，並運用**關鍵字搜圖演算法**技術，將圖片呈現給使用者。「圖圖示到」功能流程圖如圖 4 所示。

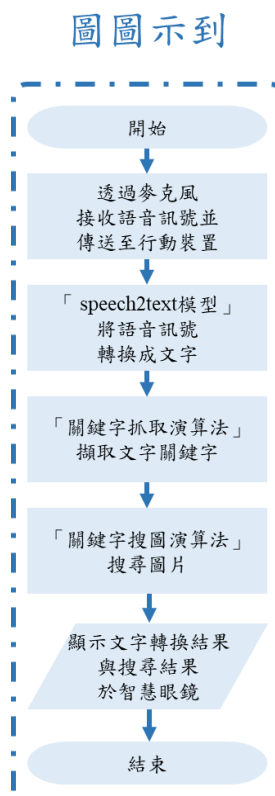


圖 4、「圖圖示到」功能流程圖

B. 硬體

本團隊將智慧眼鏡、行動裝置及麥克風組成「一圖勝千言」系統之硬體部分。「一圖勝千言」硬體裝置示意圖如圖 5 所示。



圖 5、「一圖勝千言」硬體裝置示意圖

麥克風連接著智慧型手機，使麥克風接收音訊回傳至手機，進行軟體部分之運算。手機連接著智慧眼鏡，使手機之畫面傳至智慧眼鏡的投影區，供使用者觀看。以下將詳細介紹「一圖勝千言」系統中使用的硬體設備，如表 1 所示。

表 1、「一圖勝千言」系統硬體元件
應用說明

 智慧眼鏡	提供擴增實境的穿戴式裝置，可直接連結行動裝置並觀看其中的內容，用於將圖片顯示於使用者眼前。
 行動裝置	作為使用者連接系統服務之媒介，並提供智慧眼鏡電力來源與輸出訊號。
 麥克風	收音裝置，用於收取使用者周遭的聲音。

五、實驗結果與比較

A. 實驗結果

「一圖勝千言」系統透過智慧眼鏡、行動裝置及麥克風等硬體配合「出口成章」及「圖圖示到」兩項功能，使聽障人士能夠透過本系統輕鬆的與他人溝通。為使本系統更符合使用者的需求，本團隊針對實際使用情況和問卷調

查進行探討，以下將詳細說明本系統之開發環境、使用者介面、實際使用情況、問卷調查結果以及使用者體驗。

I. 系統開發環境

在硬體部分，本系統使用行動裝置作為核心，並透過麥克風接收音訊，再連接智慧眼鏡，將行動裝置畫面投影於智慧眼鏡。

在軟體部分，本團隊使用 WebView 技術開發，並將其建置於 Windows Server 的作業系統環境下，其網頁伺服器為 Django，資料庫為 MySQL。本系統開發環境如表 2 所示。

表 2、系統開發環境

開發需求	開發環境
伺服器作業系統環境	Windows Server 2019
主要開發程式語言	HTML5
	Python3.8
	JavaScript 1.8.5
資料庫	MySQL 5.7.31
網頁伺服器	Django 3.2.18 (LTS)

II. 使用者介面

使用者可於「一圖勝千言」手機 App 的首頁中，選擇使用「出口成章」及「圖圖示到」兩項功能，手機首頁介面圖如圖 6 所示，眼鏡首頁介面圖如圖 7 所示。

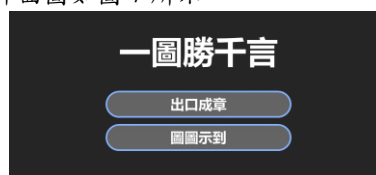


圖 6、「一圖勝千言」手機首頁介面圖



圖 7、「一圖勝千言」眼鏡首頁介面圖

為了讓使用者在與他人溝通時能夠得知其內容，因此本團隊設計了「出口成章」功能，系統會將與他人的對話語音即時轉為文字並顯示於智慧眼鏡之鏡片上。手機功能介面圖如圖 8 所示，眼鏡功能介面圖如圖 9 所示。

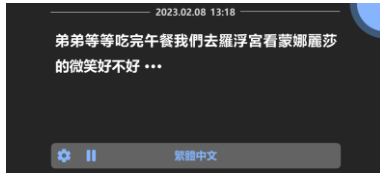


圖 8、「出口成章」手機功能介面圖



圖 9、「出口成章」眼鏡功能介面圖

為了文字以圖片方式呈現給使用者，本團隊設計了「圖圖示到」功能，系統會擷取「出口成章」功能中的關鍵字，並將擷取之結果以圖片的方式於智慧眼鏡上。手機功能介面圖如圖 10 所示，眼鏡功能介面圖如圖 11 所示。



圖 10、「圖圖示到」手機功能介面圖



圖 11、「圖圖示到」眼鏡功能介面圖

III. 實際使用情況

本團隊以呈現文字與圖片於眼鏡上的方式，幫助聽障人士更加輕易聽懂並理解他人說話之內容，為了驗證此方法的可行性，本團隊以聽障人士作為實驗對象，進行使用情況與效果之測試，經過實驗結果得知，在語音轉文字的方面，轉換的正確率約達 93%，從接收到語音到呈現圖片，所需時間大約為 0.8 到 1.5 秒，更重要的是，聽障人士也能與他人正常的溝通。

IV. 問卷調查

為了讓本系統更符合使用者的期待，本團隊採用隨機抽樣的方式進行問卷調查，針對方便性、流暢度和實用性等方面進行調查，共發放 153 份問卷，有效問卷有 148 份，其中團隊發現有 92% 的使用者認為本系統使用過程流暢，且可透過擴增實境的方式，使聽障人士在與人溝通時能夠輕鬆地理解他人所想表達的內容。

V. 使用者體驗

92% 的使用者認為本系統能夠以簡易的方式，讓聽障人士在時輕鬆地理解他人溝通時所想表達的內容。其中有部分使用者提供關於圖片數量呈現太多以及關鍵字擷取不準確的問題。因此，基於上述建議，本系統改善了圖片呈現的數量，針對數量的上限做調整，並重新檢視擷取關鍵字之方法，以解決關鍵字擷取不準確的問題。

B. 與聽障人士相關之產品比較




I. AR 助聽眼鏡

AR 助聽眼鏡雖可以即時將周遭的對話轉換為文字並顯示於眼鏡鏡片之上，使聽障人士能知道他人說話的內容，但是因為其內容不會如「一圖勝千言」，會將說話的內容抓取關鍵字，且也沒呈現關鍵之相關圖片，若語句較長聽障人士則難以了解與他人談話之內容。

II. Google 即時轉錄 App

Google 即時轉錄 App 雖可讓聽障人士在與他人溝通時清楚地得知其內容，但在溝通時若他人語句較長且只靠著語音轉文字的協助，並無法讓聽障人士理解完整內容的涵義。

表 3、與聽障人士相關之產品比較表

相關產品	 一圖勝千言	 AR 助聽眼鏡	 Google 即時轉錄 App
產品特色			
即時呈現與人交談之內容	O	O	O
呈現談話內容之關鍵圖片	O	X	X
以擴增實境的方式呈現文字與圖片	O	X	X
產品使用族群廣	O	X	X

六、結論

溝通是日常生活中基本的需求，但對於大多數聽障人士而言，因聽力的損失，不僅無法聽清他人說話的內容，且在語言學習方面也比一般人艱難[6-7]，雖然他們可以藉由替代性溝通方式來理解他人的想法，但聽障人士仍無法清楚了解他人所表達的涵義，因此本團隊為了幫助聽障人士能順利與他人溝通，將 AR 智慧眼鏡等硬體結合深度學習，開發了「一圖勝千言」系統。而為了實現本系統的功能，本團隊設計了**語音轉文字、關鍵字抓取演算法及關鍵字搜圖演算法**三項技術，以擷取其重點及相關圖片，讓使用者能透過智慧眼鏡中擴增實境之功能呈現與他人談話內容之文字。此外，本團隊經實際使用情況及問卷調查，證實本系統能讓聽障人士理解他人所想表達之內容，且順利與他人溝通。期望透過本系統能協助聽障人士在與人溝通時，不再因溝通困難而卻步，能更有自信地與他人溝通，進一步促進其融入社會和參與社會活動。

七、參考文獻

- [1] Li-Mei Lin, “The Current 語音轉文字tus of Deaf Tour Guides in Taiwan,” vol. 2, pp. 12, 2022.
- [2] Wan-Zhen Chen, “An Ethnography of Deaf Communication: The Distinctiveness of Visible Conversation,” vol. 3, pp. 27, 2013.
- [3] Valentin Belissen, Annelies Braffort and Michèle Gouiffès, “Experimenting the Automatic Recognition of Non-Conventionalized Units in Sign Language,” vol. 13, no. 12, pp. 1-2, 2020.
- [4] M Ángeles Martínez Sánchez 1, Antonio Muñoz-García, and Cristina Ros Gil. “Perception of the Impact of COVID-19 on a Sample of Spaniards with Hearing Disabilities,” vol. 20, no. 2, pp. 1-2, 2023.
- [5] 「關於臺灣手語的傳承與學習：一位學習手語者的感想」，社團法人中華民國晴天社會福利協會。檢自 <https://reurl.cc/WqeOjZ> (2023/2/1)。
- [6] Castellanos, Irina, Kronenberger, William G., Pisoni and David B. “Psychosocial Outcomes in Long-Term Cochlear Implant Users,” vol. 39, no. 3, pp. 527-539, 2018.