會議語音即時轉錄

一、研究動機

因中鋼每年例行之股東常會與會者眾多，於各項環節上皆可能有股東提問，主題縱貫資金調度、勞資關係、社會責任乃至公司決策之規劃，主席與紀錄人員一時之間難以逐一分項完整記錄，並安排對應職務人員進行回覆。且因提問者的語言可能包含國語、臺語與英語多種語系，甚至同一時間語句夾雜其中兩者或三者，並帶有各地之腔調或方言，主席需要重新組織文句，方能理解問題的核心。因此，需要一套語音轉錄系統將說話的人（這裏簡稱為『語者』）其發言的內容即時轉換成文字紀錄，供回覆者針對問題內的細節逐一詳實答覆。再者，轉為文字紀錄後，可經由關鍵詞頻與主題模型提取問題之關鍵字，以節省雙方溝通理解之時間。

二、研究目的

1. 英語、國語與台語的語音轉錄：當語者在一句話當中參雜不同語言以及方言時，需要能正確地進行辨識，並且顯示。
2. 語者變更之識別：當發言的人切換的時候，必須要能夠進行辨識，並且當有同時多人說話時，必須要能辨識出主要發言的幾位語者。
3. 音訊聲紋之去識別化：用來辨識語者的聲紋資訊必須要去識別話，不可以跟語者的個人識別資訊進行連結，因此辨識時所使用的語音資料需要進行轉換，再進行辨識。
4. 基於歷史文檔抽取關鍵字或摘要：可以將語者的語音資訊經過轉換成文字內容之後，進一步的透過關鍵字以及語意的分析，來摘要說話的內容以及發言的語意。

三、研究方法

1. 語音轉文字
   i. 詞序分析

      相異於異常分析與金融分析中的數值型時序，自然語言分析中的詞序類型更傾向狀態型時序，故考量 RNN 與 LSTM 等具序列性質之神經網路。

   ii. 狀態模型

      於自然語言處理中，字句的構成多半有既定之詞性規範，故需將聲符佐以詞性建立狀態模型，並加以詞序分析，方可於眾多拼音組合中匹配最適之選項。

   iii. 建立慣用語與專用術語之字典
2. 語者識別

i. 聲紋頻域訊息

ii. 音訊之去識別化

3. 語意分析

i. 詞頻

通過會議慣用詞，可以快速排序修正詞之備選名單

ii. 主題模型

透過字句內主題字詞之分布，可判斷該字句欲表達之領域傾向

4. 語者驗證機制

i. 語音數位簽章與預設的文字相依的複雜密碼的綁定機制設計

ii. 驗證用的語音數位簽章可以與環境聲音結合，用來辨識驗證過後的語音數位簽章，可以用來查驗後續特定語者的發言語音資料。

四、問題與挑戰

1. 語音轉錄之準確率

現行雖已有多家軟體公司致力於開發語音轉錄系統，諸如 Google、IBM 以及 Microsoft 等知名的軟體大廠，也在音訊至聲符的轉換上達到相當驚人的準確度，然在聲符至字符的轉換上，仍存在不少同音、近音錯別字，致使轉錄出的字句或文章前後語義不通，造成閱讀者理解上的障礙，甚至產生誤解。故我們將透過會議中常見的慣用語與語句中的排序，來對此類技術難點進行挑戰。

2. 語言模型之建立

由於當前雙方並未有任何音訊開發的經驗，若要重新架設一個全新的語言模型系統，可能會耗費大量時間用於音訊、語料與字典等基礎資料庫。若欲減輕開發之複雜度，可引入第三方之開源 API，協助音訊自聲符之轉換，再接續針對聲符至字符之轉換進行完善，達到高準確度的語音轉錄；然這亦會衍生一新問題，第三方 API 無法保證是否會將語者的個人隱私或特徵泄漏，存在安全性之疑慮，這同時尚待考量的問題。

3. 語音轉錄之效率

因中鋼股東常會參與者眾多，若當期發言踴躍，主席與提問者的溝通效率便會影響大會的進程，而即時性的轉錄語句生成效率，方能改善主席與提問股東雙方的溝通品質，也是本案的必需克服的關鍵要點。

4. 語者變更之識別

在尚未保有中鋼內部人員聲紋資料庫前，是無法通過聲紋識別語者的具體對象，然可通過音訊間聲紋的相對差異嘗試辨別語者對象的轉換。

5. 多語言的同時辨識，需要引入機器學習演算法，先針對語音資料進行段落的切割，在各別根據各種語言的訓練模型進行語言的辨識，而後才能

針對語音內容進行轉換成文字內容。這過程當中，牽涉多種類別的機器學習演算法，辨識的正確性以及即時性，將是一大考驗。

6. 如何能事先的辨識出幾個主要的語者，以及透過場控進行快速的調整（操作必須簡單易用，正確率高），使得語音辨識的過程，能夠鎖定特定的語者，並且在股東發言時，能夠透過司儀或場控的發言，快速定位未知的語者（新發言的股東），成為當下特定的發言語者。方法的設計需要同時考慮到高度的可用性並且同時需要達到可快速簡易操作，是一大挑戰。另外，需要在一個具有眾多人同時說話的環境當中，進行聲音的辨識，也是一大考驗。
7. 如何在股東會當中，將在場語者音訊收集過後，可進行去識別化的動作，並且不儲存原始語音檔案的情況底下進行辨識，確保在場與會主管、股東們的私密聲音檔案不會被未經授權使用，也可符合隱私性保護法規的規範。
8. 如何能根據開會的場景（如股東會）以及語者發言的內容，進行快速的摘要，並且分析出語者發言所要表達的意思大綱，會需要發展出新的機器學習辨識模組。因需要考慮語者發言的口誤，需要可以進行模糊比對與分析。分析出的關鍵字與摘要的正確性，也是一大挑戰。
9. 語者的驗證與識別需要考慮到產生誤判或是偽造語者發言的狀況，因此具有高度安全性的驗證機制，使得一些較為關鍵的語者（其發言具有決定性影響），可以有高安全性的驗證機制，確保辨識系統不會產生誤判或是無法辨識是否為偽造語音的情況。

五、預期結果

1. 即時性的語音轉錄系統
2. 可識別英語、國語以及台語夾雜之音訊
3. 模糊化音訊之個人特徵
4. 條列音訊轉換之字句
5. 允許人工修正錯別字
6. 提供推薦校正選項
7. 自動標記關鍵字
8. 生成會議文檔