語音專題第二階段專題題目選擇

1. Segmentation-Based Speech Processing 以分段為基礎的語音處理
   由於人工標記和訓練音素模型(如HMM)須使用大量資源、時間與資料庫，所以我們希望在不使用前述方法下，仍然能直接利用輸入語音訊號本身的資訊特性，以非監督式(unsupervised)的方式，自動將訊號切割成一個個類似於音素的片段，而分段後的結果就非常適合當作各式複雜語音處理的基礎，如口說詞偵測(spoken term detection)等。為此我們將介紹HAC(Hierarchical Agglomerative Clustering)演算法來做語音訊號的分段，讓同學學習、熟悉且實作。
   助教: 蔡政昱 r02942067@ntu.edu.tw

2. Blind Source Separation
   大家也許直覺認為，音訊的疊加就像是水彩的調和，一旦混合就難以再區分，然而觀察混合訊號的頻譜，我們可以透過一些拆解與重組，試圖重建原始未混合的訊號，這樣的task叫做 Blind Source Separation。本專題主要使用PLSA模型套用在頻譜上，做些類似以下實驗: 音樂樂器分軌、音樂去人聲、人聲去噪音、多人交談分軌。
   助教: 熊信寬 b98901024@ntu.edu.tw

3. 語音資訊檢索辭典外(Out-of-Vocabulary)查詢詞問題
   語音資訊檢索最基本的方法，就是先把聲音辨識成文字，然後套用文字檢索的方式來搜尋這些辨識結果。然而，因為使用者輸入的查詢詞(query)大多為專有名詞(人名、地名等)、專用術語或新詞(如「八八水災」)，但這些詞往往沒有被辨識系統的辭典(lexicon)所涵蓋，導致語音辨識系統完全辨識不出這些辭彙，也就完全沒辦法直接套用文字搜尋的方式來檢索了。本專題將探討如何運用最新的技術來解決辭典外(Out-of-Vocabulary，OOV)查詢詞的問題。
   助教:周伯威 botonchou@gmail.com

4. 語音文件摘要(Spoken Document Summarization)
   與純文字資料相比，語音文件不易閱覽且耗時，在基本的摘要系統中，以選擇段落或語句來構成摘要(summary)，在限制字數下選出的摘要，能夠盡量涵蓋原本語音文章中的資訊。利用語意與詞彙重要性、關鍵詞擷取、語調資訊，並同時考慮摘要冗餘分數，為每一個語句評分，並擷取適當語句作為摘要。本專題將會利用「數位語音處理概論」課程影音以及「公視新聞」兩個資料庫作為素材，實作各種廣泛使用之摘要系統，引導同學進行自動化語音文件摘要之研究。
   助教:向思蓉 r01921050@ntu.edu.tw

5. 概念比對(Concept matching)之資訊搜尋
   常見的搜尋引擎大多以字面比對(literal matching)的搜尋方式為主，也就是檢索系統只會回傳包含使用者查詢辭的文件，但這樣的作法顯然無法完全滿足使用者的需求，比方說，當使用者輸入查詢辭「歐巴馬」時，通常希望找到的是與美國總統相關的資訊，即使沒有包含「歐巴馬」這三個字的也希望能找到。本專題將從文件檢索的標準方法著手建構具概念比對(concept matching)功能的語音資訊檢索技術，進而利用語音處理技術來提

升效能。
助教:李昀樵 r01942062@ntu.edu.tw

6. 使用動態時間校正(Dynamic time warping)之語音搜尋:
語音資訊搜尋中最基礎的功能就是語音詞彙偵測(Spoken term detection)-- 找到語音文件中的查詢詞。雖然主流的做法是利用語音辨識，但是過多的辨識錯誤常會大幅降低搜尋正確率。在這個研究當中，我們探討並設法實作一個近年新興的做法:利用動態時間校正(Dynamic time warping)直接找語音文件中聲音類似查詢詞的片段。這個方法可以直接比對聲學特徵的相似度省去了辨識的步驟。而類似的方法也可以應用在音樂搜尋上。在這個專題中，助教會教同學用C++或python寫出Dynamic time warping相關程式，抽取各式各樣不同的特徵參數，使用不同的距離量測，並研究分析怎麼作比較好。
助教:李昀樵 r01942062@ntu.edu.tw

7. 非督導式用模型為基礎的語音處理(Model based Unsupervised Speech Processing)
今日語音辨識基本上是用有人標註好(註明每一段聲音在說什麼)的語料去訓練模型，但要人標註語料總是麻煩的。近年有人思考不要標註語料的方法，假設在沒有文字標註的情形下，我們希望可以使用 HMM 模型，直接從聲音中學出語言中一些類似音素(phoneme)的基本單位，再根據這些基本單位回頭把原本的聲音辨識成類似音素的單位，再根據這些單位訓練出更準的基本單位，不斷的循環。就像小孩子在學習語言時所學的音素(phoneme)及音標一樣，不斷反覆的練習聆聽自己朗讀的聲音，自動的讓機器學習組成語音的基礎單元。再將這些單元和語言學家所設定的音素比較重覆性，看看我們在沒有標註的情況下學習出的語言基本結構具有什麼性質。在這個專題中，助教會教同學用 HTK 及(python或C++)來進行實作。同學可以提供自己較感興趣的語料，或是某種隨時間變化的訊號進行實驗。
助教:鍾承道 f01921031@ntu.edu.tw

8. 類神經網路語者調適 (Neural Network for Speaker Adaptation)
語者調適 (Speaker Adatpation) 的目標是利用少量新的數據將現有的系統快速調整至適合新的語者或資料。傳統HMM-GMM系統上有許多泛用的辦法包含MLLR以及MAP等等。近年來因為機器學習 (Machine Learning) 的發展， 類神經網路 (Neural Network) 已經被廣泛用於語音系統， 特別在聲學模型上取得了非常大的成就， 但許多傳統用於HMM-GMM的方法並不能用直接使用在類神經網路上， 因此在類神經網路上的語者調適就成了重要的研究課題。 本專題第一階段會實作HMM-GMM上的MLLR以及MAP實驗，同時建立類神經網路的基礎實驗結果。 第二階段則是在類神經網路上開發調適演算法。
基礎需求: 基礎線性代數， Linux操作， 機器學習基礎理論
額外需求(Optional): 自有nVIDIA GTX 560以上等級GPU， 聲學模型相關知識
助教: 葉青峰 d00942013@ntu.edu.tw

9. 行動裝置語音辨識系統開發
隨著行動裝置 (手機、平板) 的快速普及， 行動裝置上的語音辨識系統顯得越來越重要. 不同於現有供一般使用者使用之辨識系統， 包含Google Voice Search以及微軟Bing搜尋等等， 行動裝置上的辨識系統不需要將聲音傳送至雲端伺服器， 因此可以在沒有網路的情景下運作。同時由於行動裝置往往是個人化的設備， 如何收集使用者資訊， 包含

慣用字詞或是使用者聲音， 也是行動裝置所獨有之處， 同時由於行動裝置系統可以離線運作， 也避免了將使用者資訊上傳至雲端的安全風險。本專題第一階段先讓同學熟悉實驗室既有的iOS端行動辨識系統， 第二階段視狀況在iOS或Android端進一步開發介面， 同時進行針對使用者個人化的資訊收集與實驗。
基礎需求: 語音辨識系統， 至少進行兩學期以上專題之意願
額外需求: Java程式語言， C++/Java整合開發
助教: 葉青峰 d00942013@ntu.edu.tw

10. 語言學習技術(Computer Aided Language Learning)中基於不變結構(Invariance structure)之發音評量
學習華語對外國學生來說是十分困難的，由於在課堂的學習資源有限，老師很難給予外國學生充分的時間練習，所以Computer Aided Language Learning(CALL)系統變得很流行，因為學生可以隨時學習、且避免因為尷尬導致學生不願意學習。本研究是給定我們有華語老師對外國學生的中文字發音標注分數的訓練語料，透過機器從訓練語料中學習老師是怎樣評分的，讓日後我們可以用電腦取代老師的評分。不變結構(invariance structure)是指每人的聲音都不一樣，但假設如果把每一個基本音都能唸得彼此相當不同不易分辨清楚，老師就會給較高分數；這個各個基本音之間的距離是所謂的不變結構結構。
助教：蘇嘉雄 r01942034@ntu.edu.tw