

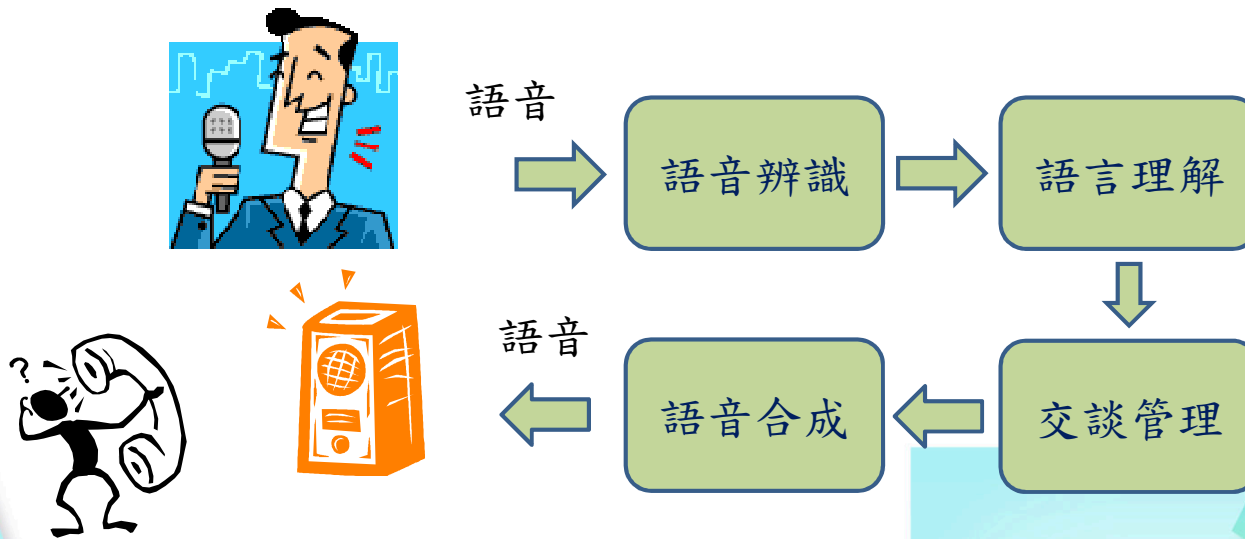


第六章

機器聽覺

語音交談系統

- ◆ 機器人具有智慧的重要表徵
- ◆ 主要模組
 - ◆ 語音辨識: 辨識說話聲音包含了哪些詞彙
 - ◆ 語言理解: 剖析並解釋其語意資訊
 - ◆ 交談管理: 整合資訊以產生回應的訊息
 - ◆ 語音合成: 將回應訊息合成為聲音



語音的產生

◆發聲器官協同合作

- ◆唇、齒、舌、口腔、鼻腔、咽喉腔、聲帶、支氣管、肺等。

- ◆氣流造成聲帶的週期性振動

- ◆咽喉腔、口鼻腔及唇、齒、舌、顎等器官所形成的發聲通道

◆聲波

- ◆在空間傳遞的縱向疏密波

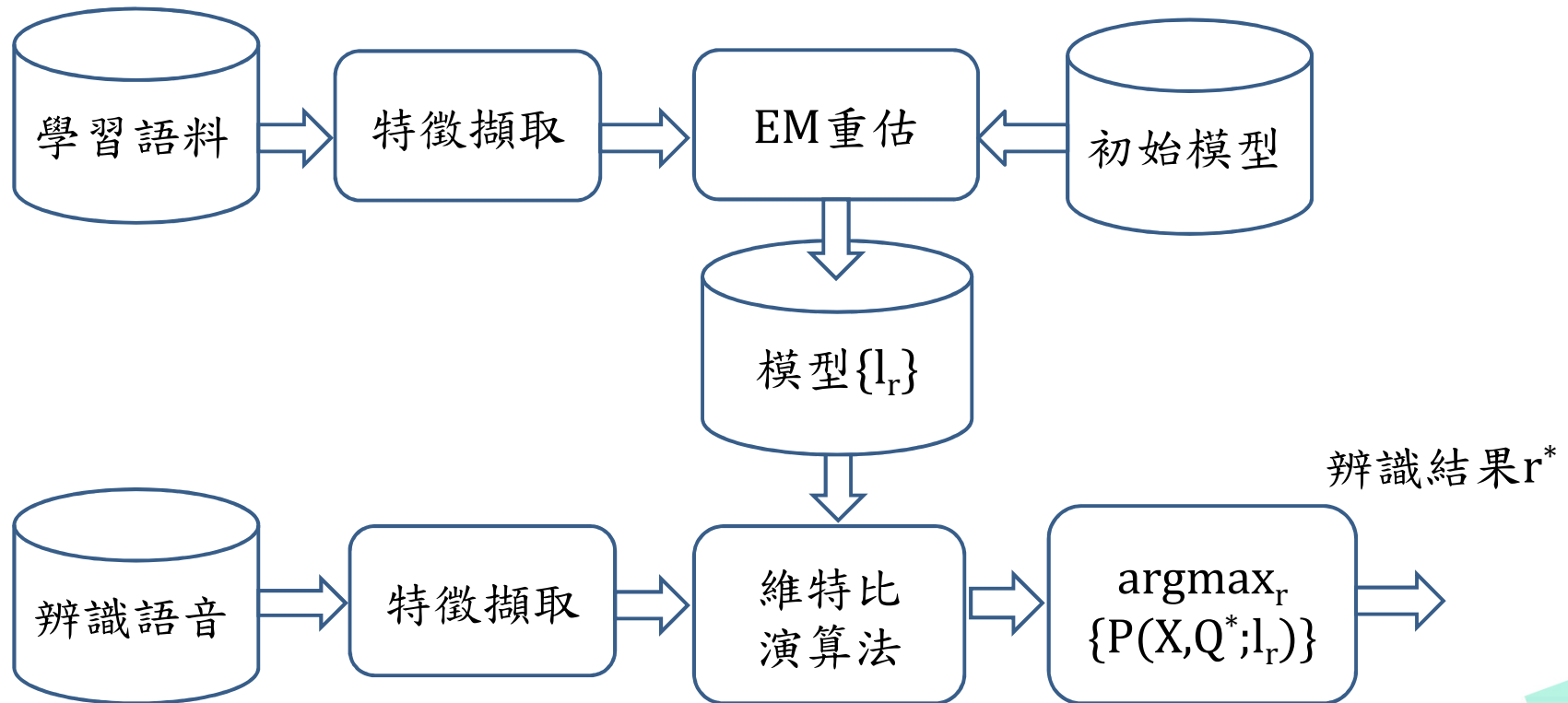
語音的儲存

- ◆ 麥克風
- ◆ 類比數位轉換裝置(ADC)
- ◆ 取樣頻率
 - ◆ 通常設為16k Hz (音樂CD 44,100 Hz)
 - ◆ 人類語音主要分布在8kHz以下
- ◆ 量化方式
 - ◆ 通常設為16 bits PCM (-32,768 ~ 32767)
 - ◆ 以short int儲存

學習問題

- ◆ 據學習資料，找到最佳化的參數集
- ◆ 期望值最大化演算法 (Expectation Maximization, EM)
 - ◆ 先給定初始模型參數 l ，根據公式重估參數 l' 。將重估後的參數 l' 再代入模型參數，就可求取下次的重估參數，依此方式迭代直到期望值收斂為止。

斷開單詞語音辨識系統



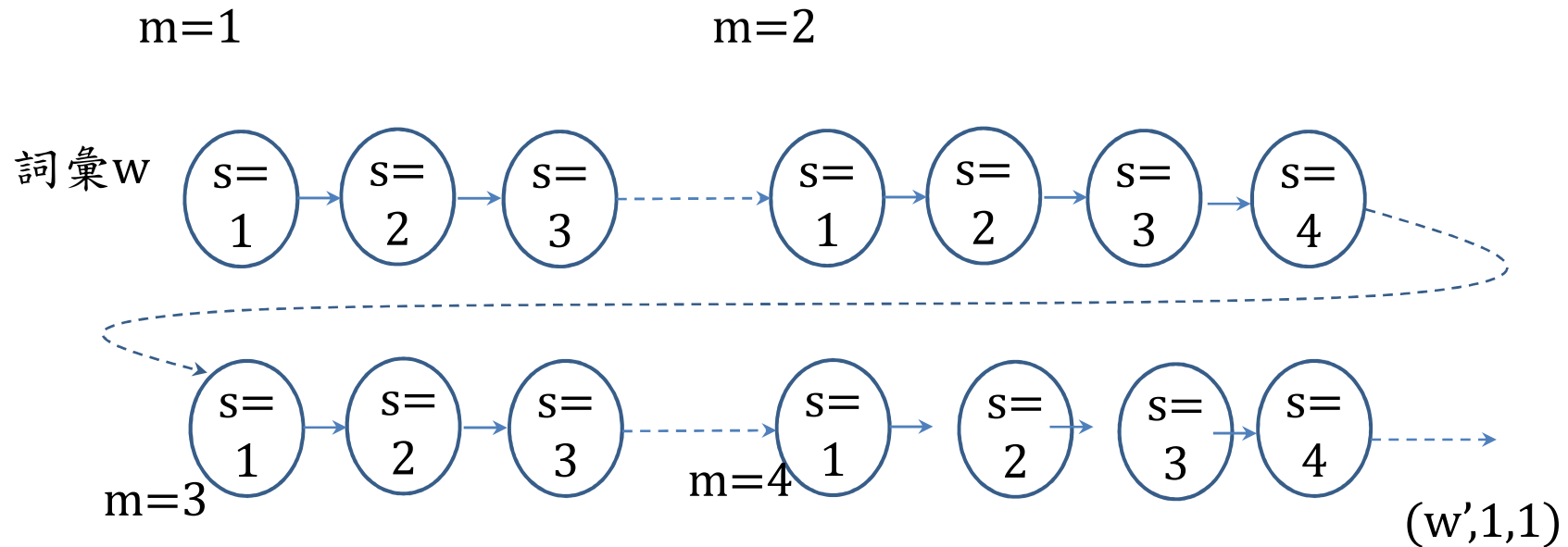
語音辨識技術的困難

- ◆ 連續語音辨識
 - ◆ 大量詞彙導致搜尋空間和計算量大
- ◆ 語者口音的變異
 - ◆ 語者特定/語者不特定
 - ◆ 語者調適
- ◆ 地域發音的變異
- ◆ 錄音通道的變異
- ◆ 環境的雜訊

聲學模型的基本單位

- ◆ 和語言結構有關
- ◆ 中文(國語)
 - ◆ 詞 (例: 打開)
 - ◆ 音節 (例: da kai)
 - ◆ 聲、韻母 (例: d a k ai)
 - ◆ 音素 (例: d a k a i)
- ◆ 以「詞」為辨識基本單位缺乏彈性
 - ◆ 詞彙量大時不易錄製足夠訓練語料

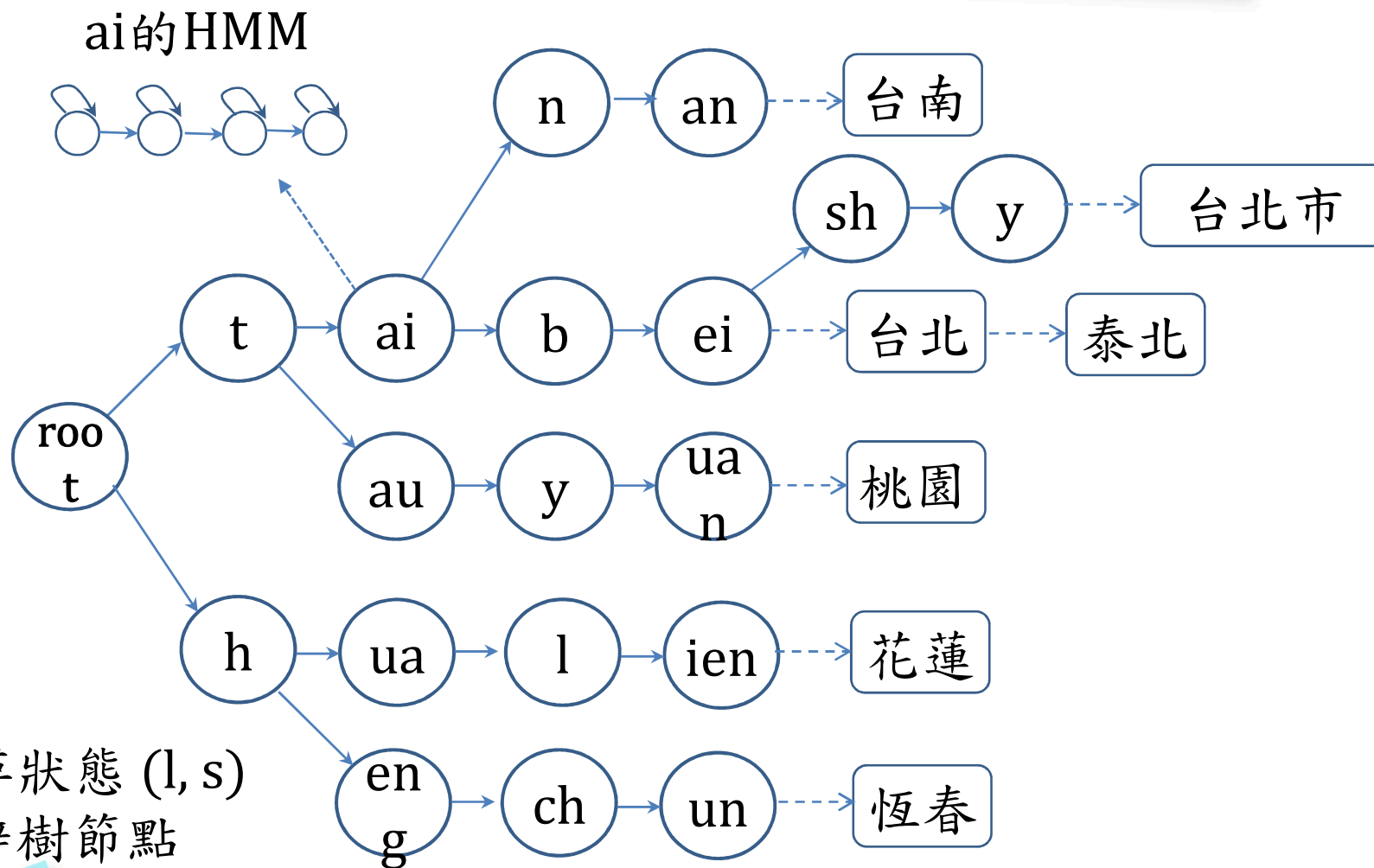
以序列結構為基礎的搜尋



搜尋狀態 (w, m, s)

w : 詞彙, m : 模型索引, s : 狀態

以樹狀結構為基礎的搜尋



搜尋狀態 (l, s)

l: 辭樹節點

S: HMM狀態

樹狀結構搜尋法的優點

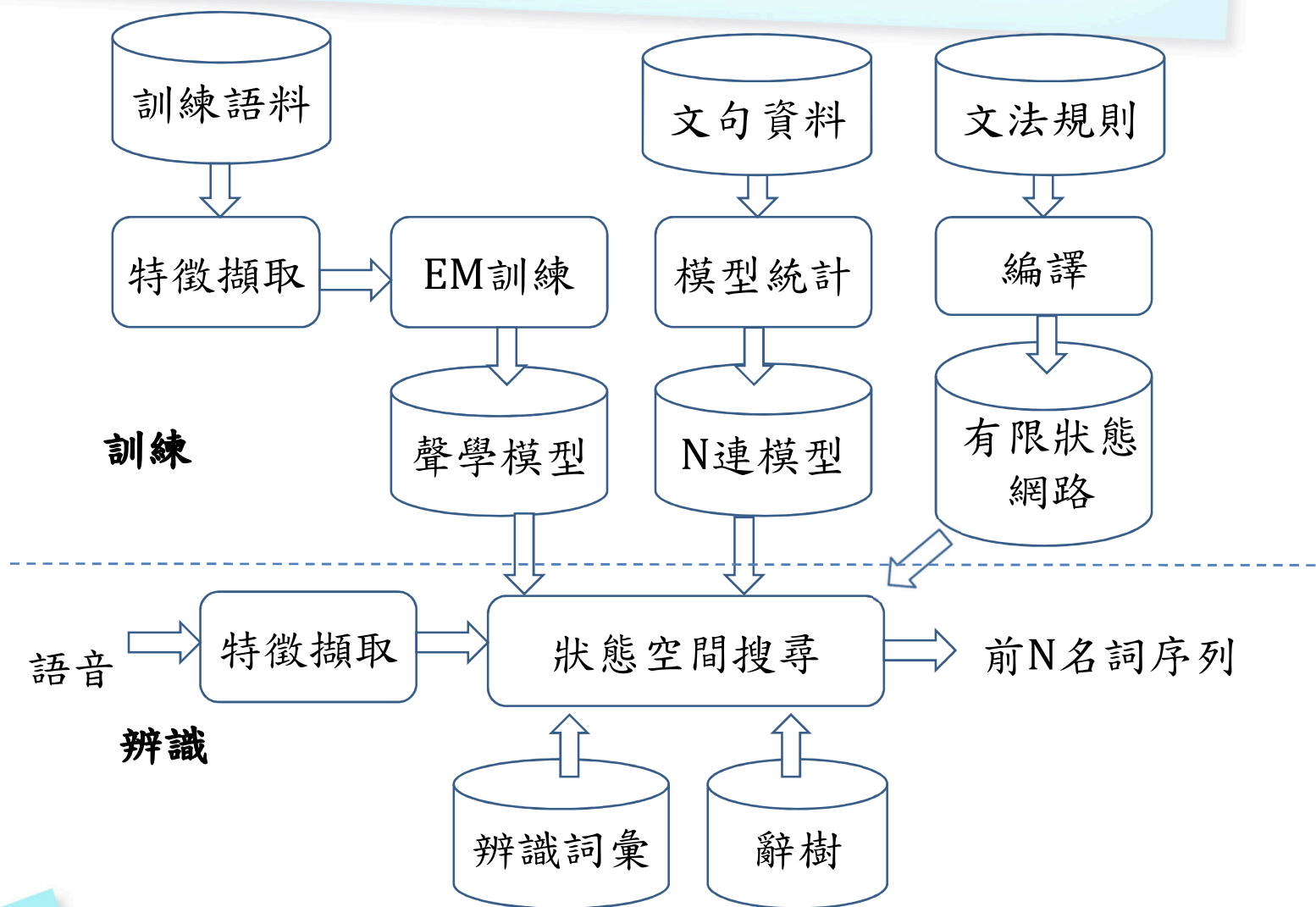
◆ 節省記憶空間

- ◆ 樹狀結構較序列結構節省記憶空間
- ◆ 詞彙量大時更明顯

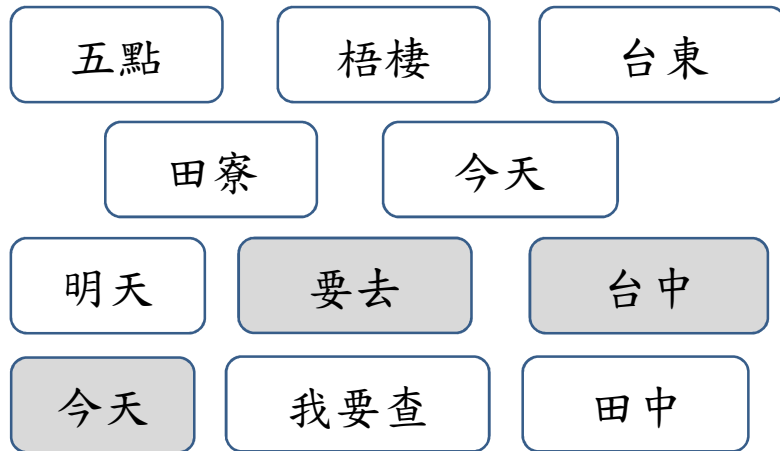
◆ 減少搜尋時間

- ◆ 搜尋時重疊部分計算不重複
- ◆ 台北、台南：t和ai的部分共享
- ◆ 台北、台南、桃園：t的部分共享

整合聲學和語言模型辨識架構



詞圖格式的輸出



「今天要去台中」輸出詞圖

詞圖/格狀詞組

- 起訖時間/路徑機率
- 較佳「詞彙含入率」
- 較具有彈性，可利用語意交談資訊找到更好的結果
- 較複雜之自然語言交談系統