第 三 章　單一語音離散表徵與語音標記的對應模式

3.1　動機

3.2　相關研究

在 HuBERT 出來之後，

3.2.1 語音表徵的語音學分析

3.3 衡量方式

3.3.1 音位（phoneme）長條圖（bar chart）

3.3.2 純度（purity）、熵（entropy）

3.3.3 對齊（alignment）

3.4 語音學分類

3.5 分析結果

3.5.1 基於各自音位的分析

3.5.2 基於語音學分類的分析

3.6 本章總結

第 二 章　背景知識

2.1　真全部深層類神經網路（deep neural network）

深層類神經網路（Deep neural network）是麥氏（McCulloch）在 1943 年提出[1]，取法自生物神經連結的計算模型，旨在模擬生物神經系統的連結，以模仿生物的各項功能，進而透過機器學習的最佳化演算法，使得整個模型能夠藉由資料去貼合理想的函數，以達成應用或工程上所需要的各種任務。。

以發展此模型為主軸的心理學流派，在計算認知神經科學中被稱為「連結派（connectionism）」，其後因為該網路的彈性與平行化的能力，和諸如圖形處理器（graph processing unit，GPU）等硬體裝置能夠最好的利用。并能夠更好的描述資料分佈、達到前所未有的效能，因此近年在電腦科學的機器學習領域中獲得重大進展，並因此現已成為人工智慧發展的主流。

基於深層類神經網路的神經架構有 CNN、RNN、Transformer 等等，由於這些架構在語音與文字處理上都已經被廣泛使用，因此在下面分別介紹：

2.1.1　卷積式（convolutional）類神經網路

卷積式類神經網路一開始是在 cite 中提出，主要是鑑於影像中的局部性（locality），讓 NN 可以在。在語音中，因為語音訊號的資訊是被呈現在時間維度上，因此通常使用一維的卷積式類神經網路，以捕捉時間維度上的局部性特徵，例如本研究特別探討的 phoneme、morpheme 等等。

2.1.2 遞迴式（recurrent）類神經網路

2.1.3 序列至序列（sequence-to-sequence）模型

由于許多實際上的資料都是 2 個序之間互相配對的關系此類的資料包含語音文字。信號等等，都是以時間軸為主要演變方向的資料。因此有一類模型。會被以序列制序列的模式進行訓練。旨在模擬輸入與輸出序列之間的變化與相依關係（dependency）。

此類模型一般的架構是由一個編碼器和解碼器構成其中編碼器是將輸入訊號借由內部表征進行編碼。依據每個時間點輸入訊號的順序來改變其內部表征的狀態接著將最後一個時間點的表徵作為整個序列的特征傳遞給解碼器進行輸出訊號生成。

2.1.4 專注（attention）機制

原本的序列自序列模型本身。需要讓解碼器單純透過最後一個時間點。的表征資訊來完全儲存輸入序列的一切資訊以工解碼器判斷。并生成輸出序列。然而，由于單就最後一個向量進行判斷對於解碼器而言過於不易。因此。盧氏提出。在編碼器中對輸入序列的不同時間點進行注意力機制亦即讓解碼器可以根據當下所需要輸出的內容判斷應該要重新對輸入序列的哪些部份進行更多的加權。

2.1.5 轉換器（Transformer）

其後，由瓦氏（Vaswani）cite 提出的論文中提出了一個完全由注意機制。所構成的序列自序列模型。原先該模型適用於解決機器翻譯。的問題。由於其能夠高度平行化的特性，日後在自然源處理和語音處理，甚至到電腦視覺領域等近乎整個

深層學習的領域都被廣泛的應用。

2.2 表徵（representation）學習

2.2.1 文字的語意表徵

2.2.2 語音特徵與表徵

2.3 語音基石模型與自監督式學習

2.3.1 自監督式學習

2.3.2 語音基石模型

2.3.3 離散單元

2.4 本章總結

參 考 文 獻

[1] W. S. McCulloch and W. Pitts, “A logical calculus of the ideas immanent in nervous activity,” *The bulletin of mathematical biophysics*, vol. 5, pp. 115–133, 1943.