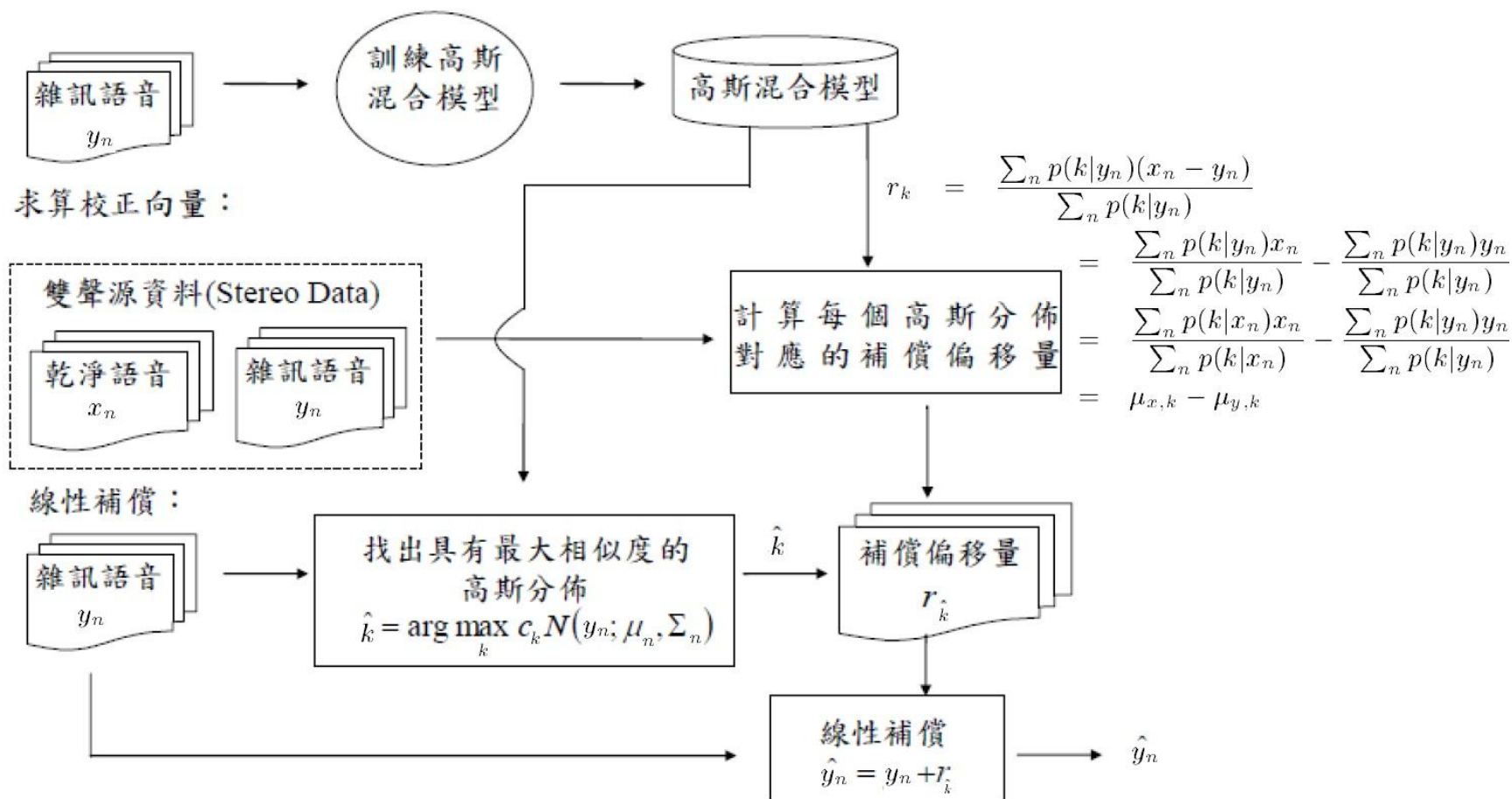


20101003進度報告

使用HTK訓練GMM 與 SPLICE實驗

Reporter: 吳柏鋒
Professor: 陳嘉平

SPLICE架構圖



雙聲源為基礎之分段線性補償流程圖

補償公式

$$\hat{x} = \sum_k p(k|y)(y + r_k)$$

\hat{x} : 補償後參數

k : 為高斯元件個數

y : noisy 參數

補償公式

$$\begin{aligned}r_k &= \frac{\sum_n p(k|y_n)(x_n - y_n)}{\sum_n p(k|y_n)} \\&= \frac{\sum_n p(k|y_n)x_n - \sum_n p(k|y_n)y_n}{\sum_n p(k|y_n)} \\&= \frac{\sum_n p(k|y_n)x_n}{\sum_n p(k|y_n)} - \frac{\sum_n p(k|y_n)y_n}{\sum_n p(k|y_n)} \\&\stackrel{.}{=} \frac{\sum_n p(k|x_n)x_n}{\sum_n p(k|x_n)} - \frac{\sum_n p(k|y_n)y_n}{\sum_n p(k|y_n)} \\&= \mu_{x,k} - \mu_{y,k}\end{aligned}$$

事後機率

$$\begin{aligned} p(k|y) &= \frac{p(k, y)}{p(y)} \\ &= \frac{p(y|k)p(k)}{\sum_k p(y|k)p(k)} \\ &= \frac{\alpha_k \mathcal{N}(y; \mu, \Sigma)}{\sum_k \alpha_k \mathcal{N}(y; \mu, \Sigma)} \end{aligned}$$

HTK擷取feature

- 重新擷取的feature有14維，第13維為 C_0 ，第14維為log energy
- 透過修改FrontEnd，在此使用 C_0 取代log energy

HTK擷取feature

- 並改用power spectral density取代原先預設的magnitude spectrum
- 主要針對前13維的向量來做SPLICE補償之運算

訓練模型

- 之前是在windows環境下HTK相關套件訓練GMM，在產生的mean跟cov的數據有所誤差
- 透過重新改寫AURORA2中訓練語音模型的script，將CleanTR與MultiTR部分各別作改寫，改寫成訓練GMM的scripts

SPLICE公式相關參數

- 使用clean和noisy資料共同訓練出的GMM，並根據其各別mean值求得 r_k
- 再使用noisy資料來訓練GMM，並將獲得的mean、cov和weight等相關變數，代入SPLICE補償公式中作mfcc參數補償

Mfcc參數補償方式

- 主要對multiTR與test_set作SPLICE補償
- 實驗一：
兩者只針對noisy部分作補償
- 實驗二：
兩者針對clean部分跟noisy部分都作補償

實驗流程

- 利用clean和noisy資料，訓練GMM模型獲得SPLICE公式中的相關參數
- 並對原先擷取到的mfcc參數作SPLICE補償
- 最後在將補償過後的mfcc參數重新作語音模型訓練，並進行語音辨識實驗

實驗數據一

Multi_Testa	subway	babble	car	exhibition
Clean	84.23	83.17	83.45	84.56
20dB	80.63	79.23	82.28	75.13
15dB	76.63	79.47	81.57	70.90
10dB	69.94	73.91	75.22	60.88
5dB	58.37	57.98	60.66	39.59

實驗數據二

Multi_Testa	subway	babble	car	exhibition
Clean	86.03	85.73	85.65	85.78
20dB	82.16	80.44	83.92	72.14
15dB	77.37	78.57	81.90	67.82
10dB	67.73	72.70	73.55	58.13
5dB	54.68	55.20	54.37	36.93

如何改進實驗辨識率

- 在訓練GMM時候的參數重估步驟由3次修改增加為4次
- 檢查補償公式程式碼是否有誤
- 訓練GMM時候的分裂方式在大於8個 component時，下一次分裂增加次數為原先的一半