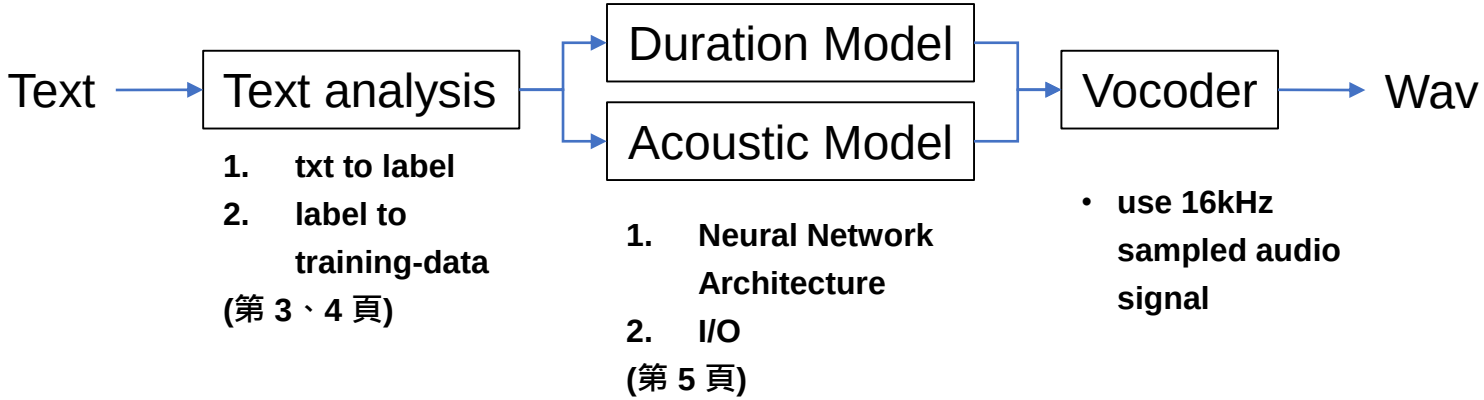
Traditional TTS design (block diagram with clear I/O & processing of each block)

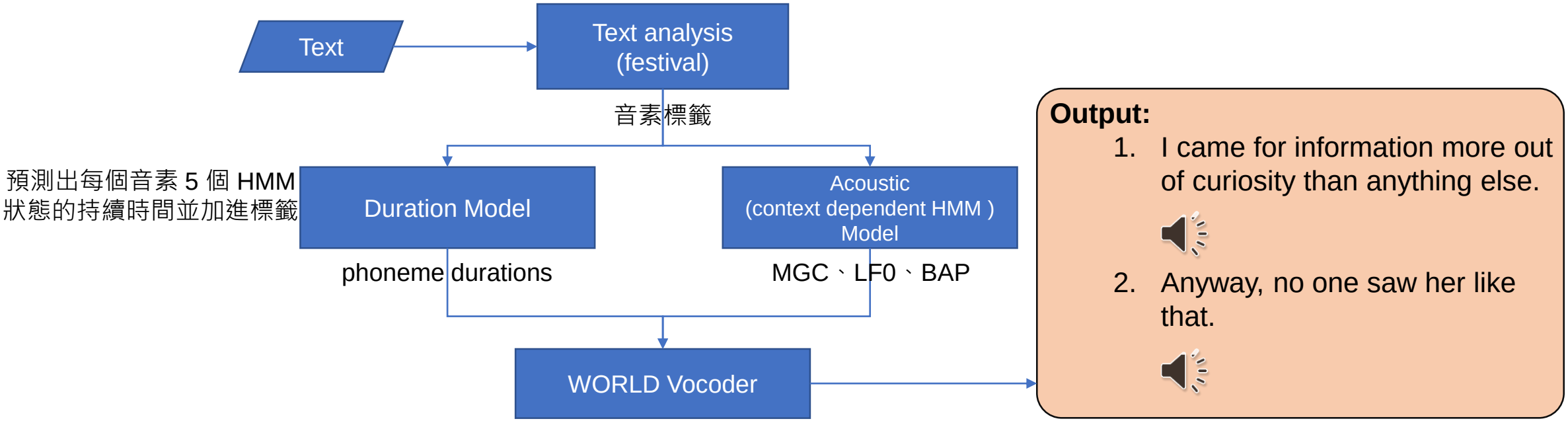
Sian-Yi Chen

Advisor : Tay-Jyi Lin and Chingwei Yeh

Block-diagram of speech synthesis system



Synthesis part



Text analysis (txt to label)

在Merlin提供的標籤中有兩種類別，分別是state align用狀態對齊與phoneme align用音素對齊，預設使用state align方式對齊。

State align使用HTK(Speech Recognition Toolkit)生成，以發音狀態為單位的標籤文件，而每個音素都由多個狀態組成，這邊則是指定生成5個HMM狀態的標籤。

txt to label

生成txt檔(一)，並使用festvox中一個名為EHMM的工具生成全文標籤，

EHMM(ergodic HMM)是一種對齊方法，它解釋了音素標籤之間可能存在停頓的可能性。

將txt檔轉換成全文標籤後，並依照HMM-base標籤格式(二)生成具有5個狀態得HMM標籤(三)

(一)：Text 檔

- (arctic_a0001 "Author of the danger trail, Philip Steels, etc.")
- (arctic_a0002 "Not at this particular case, Tom, apologized Whittemore.")

(二)：Context-dependent label format for HMM-based speech synthesis in English

p1^p2-p3+p4=p5@p6 p7
/A:a1 a2 a3
/B:b1-b2-b3@b4-b5&b6-b7#b8-b9$b10-b11!b12-b13;b14-b15|b16
/C:c1+c2+c3
/D:d1 d2 /E:e1+e2@e3+e4&e5+e6#e7+e8 /F: f1 f2
/G:g1 g2 /H:h1=h2^h3=h4|h5 /I:i1=i2
/J: j1+ j2- j3

lab_format.pdf

(三)：具 5 個狀態的 HMM 標籤

- 0 50000 x^x-sil+sil=ao@x_x/A:0_0_0/B:x-x-x@x-x&x-x#x-x$x-x!x-x;x-x|x/C:0+0+0/D:0_0/E:x+x@x+x&x+x#x+x/F:0_0/G:0_0/H:x=x@1=2|0/I:0=0/J:14+8-2[2]
- 50000 100000 x^x-sil+sil=ao@x_x/A:0_0_0/B:x-x-x@x-x&x-x#x-x$x-x!x-x;x-x|x/C:0+0+0/D:0_0/E:x+x@x+x&x+x#x+x/F:0_0/G:0_0/H:x=x@1=2|0/I:0=0/J:14+8-2[3]

Text analysis (label to training-data)

進神經網路訓練之前，需要將標籤檔轉換成二進位檔或是向量化，也就是現在神經網路做的Embedding，在Merlin中有兩種轉換的文件，差別為生成檔案的維度不同，分別為416與600維，此文件稱為**問題集(Question file)**。
問題集針對不同的語言需要自行設計，這邊使用的是416維的問題集，也就是由416道題目所組成，內容包含判斷前後文的聲韻母為何？聲母、韻母、韻律、位置特徵劃分等等。

Neural Network Architecture

Input: 416 dimensions label binary file
Duration model (DNN): 4*512 (tanh)

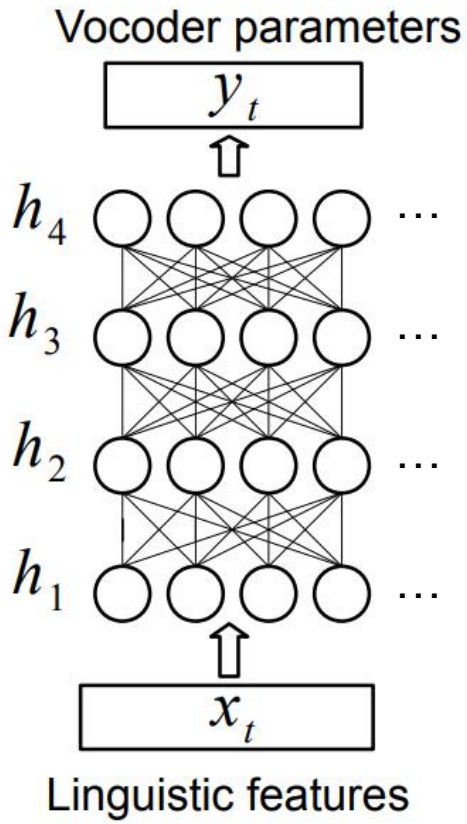
- **Output: 預測出每個音素 5 個狀態的持續時間**

- Batch size: 256
- Learning rate: 0.002
- Train file number: 50
- Valid file number: 5
- Test file number: 5

Acoustic model (DNN): 4*512 (tanh)

- **Output: mgc: 60維; bap: 1維; lf0: 1維;**

- Batch size: 64
- Learning rate : 0.002
- Train file number: 50
- Valid file number: 5
- Test file number: 5



圖一：前饋神經網路(DNN)