

# Wang TTS (based on VCC 2020 ref. design): results & improvement strategies

---

Sian-Yi Chen

Advisor : Tay-Jyi Lin and Chingwei Yeh

# Outline

## Action item

目標為完成王老師說話風格的 TTS，目前先利用實驗室 IASoC 錄製的語料進行實驗

週報意見：因為 ASR 與 TTS 為獨立的模塊，因此直接討論 TTS，不再討論 ASR 的部分

## Status report

### 上週概要

- 完成 cascade ASR 與 TTS，後續利用實驗室錄製語料進行實驗

### 目前狀況

- 上週 demo 音檔 (語料對象 TMF1)具有一定的品質，使用了同樣的微調方式，微調了兩種不同的語料
  1. 王老師 lab 繆詠青 女生語料 (qing)
  2. 鄭竣尹 男性語料 (jun)
- 上述三者語料品質為：TMF1 > qing >> jun (p.4 有結果音檔)

### 改善策略

- (已執行)
1. 調整腳本中特徵提取的參數
  2. 增加訓練語料
  3. 調整 TTS pretrain model 微調的層數
  4. 更改 vocoder，從 non-AR(autoregressive) 換成 AR，從 Parallel WaveGAN(PWG) 換成 WaveNet

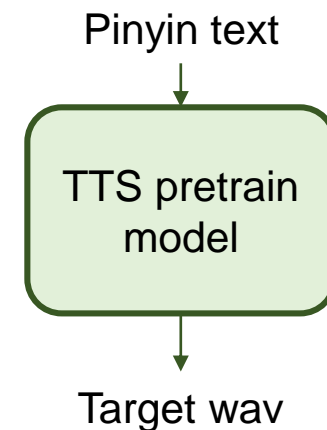


Fig. 1 Conversion process

# Improvement strategies

## 1. 調整腳本中特徵提取的參數

(ESPnet 除了提供預訓練模型，也有提供微調腳本，原本的參數為英文 TTS 的參數，我將其改成中文 TTS 的參數)

- FFT 點數：1024 -> 2048
- 點數位移量：256 -> 300
- windows 長度：無 -> 1200

## 2. 增加訓練語料


- 原本
  - Train: 60 句
  - Dev: 10 句
  - Test: 20
- 增加
  - Train: 290 句
  - Dev: 10 句
  - Test: 20 句

## 3. 調整 TTS pretrain model 微調的層數

## 4. 更改 vocoder，從 non-AR(autoressive) 換成 AR，從 Parallel WaveGAN(PWG) 換成 WaveNet

# Original results & improved results

Demo 語句：他腰痛的老毛病又犯了

	TMF1	qing	jun
原本音色	摩洛哥農業在向歐洲 出口時享受特殊待遇 		
改善前			
調整特徵提取參數			
增加語料			
調整特徵提取參數+增加語料			