

# 畢業論文

---

Sian-Yi Chen

Advisors : Tay-Jyi Lin and Chingwei Yeh

# Outline

- 論文題目：使用Transformer達到具個人風格並接近人類音質的TTS
  - 預期論述方式
  - 預期實驗結果
- 創新、突破點
  - 對傳統TTS做詳細的介紹
  - 實作Wang老師TTS
  - 提升音質
  - 做實驗證明串接ASR與TTS這種簡單架構是有競爭力的
- 目前進展：
  - statistical parametric synthesis
    - ❑ 老師週報的意見：加速收斂到一個簡單的介紹
    - ❑ 目前關於參數合成的技術已大致看完，但對於許多公式僅知道公式的目的，無法理解算式過程或為什麼要這樣計算
    - ❑ 內容分為synthesis與training part，其中training part的三音素模型的實作過程還尚未清楚
  - 後續目標
    1. 先對statistical parametric synthesis整理一個易於理解的版本（每一個步驟做了什麼、如果該步驟有公式盡量簡單呈現）
    2. 補缺失不完整的部分

# ■ 預期論述方式

## Abstract

- 簡述TTS，並說明TTS的目標，點明自己的三項貢獻

## Introduction

- 描述題目背景：說明TTS從古至今各方法的概述，
- 研究動機：針對傳統中最具代表性的statistical parametric synthesis作介紹，傳統方法的缺點
- 研究方法：介紹 VCC2020 baseline，並將其從英文版本更改成中文版本
- 結論與貢獻

## Body

- 完整介紹statistical parametric synthesis與VCC2020 baseline
- 說明問題定義、系統架構、實作過程、實驗結果

## Conclusion

- 將問題、方法、結果用別人看完的角度重新描述一次
- 關於這個研究，未來可以發展的方向

# ■ 預期實驗結果

- 使用王老師的聲音做為語料微調 Transformer，並輸出具王老師音色、音調的音檔。
- VCC2020 baseline所使用的預訓練模型，是使用女聲語料作為訓練資料，因此模型對於高頻學習較好，相對對於男生的聲音表現較差。
- 將實作完成的Wang TTS（男生、中文）與VCC2020 baseline（女生、英文）的MOS分數比較，Wang TTS的MOS分數應略低於VCC2020 baseline，但音質仍已足夠好。
- 將fine-tune完成的TTS串接各ASR，說明ASR的辨識錯誤並不會影響TTS的效果，並再次證明VCC2020 baseline中表示ASR+TTS這種簡單架構對於語音合成來說是具有競爭力的。

# ■ 突破點

1. 對傳統TTS ( statistical parametric synthesis ) 做詳細的介紹
2. 使用自行準備 ( 王老師 ) 的語料，對現有的**TTS**模型做微調，並合成出帶有王老師說話風格的語句
3. 請實驗室所有人幫忙測試**MOS**分數，證明音質與**baseline**相比有所提升
4. 將完成的**TTS**串接**ASR**，並再次證明**ASR+TTS**這種以往認為過於簡單的架構合成出來的語音是具有競爭力的

# ■ 目前進展

## statistical parametric synthesis

### ● Synthesis part

- Context-dependent Label
- Conversion Process of Text-to-Label
- Letter to Sound Rules
- 持續時間模型
- 參數生成spectral與excitation序列
- Excitation生成和語音合成MLSA濾波器

### ● Training part

- Excitation & spectral 提取
- Hidden Markov Model (HMM)介紹
- 訓練HMM模型
- X 單音素HMM擴展至三音素HMM
- 利用決策樹做上下文聚類取得最佳參數

## multi-speaker, x-vector Transformer-TTS model

- Feature representation (MFCC)
- Embedding (x-vector)
- TTS model (Transformer)
- Attention mechanism
- Vocoder (Parallel WaveGAN)

## 實作 Wang TTS

- 前處理王老師音檔
- fine-tune
- MOS評分