

# VCC 2020 baseline design

---

Sian-Yi Chen

Advisor : Tay-Jyi Lin and Chingwei Yeh

# Outline

- Action item

- 將 [1] 所提供的 open online resource 建構起來

- Status report

- 論文中使用到的 ASR 與 TTS 皆提供自 ESPnet (end-to-end speech processing toolkit) , 且 ESPnet 遵循 Kaldi (speech recognition toolkit) 風格的數據處理、特徵提取/格式和配方 (Fig. 1)

- 實作步驟：

- 1. Setting up environment (需要 GPU 且為 Linux 環境)

- Step1 : Install requirements package

- Step2 : Install Kaldi

- Step3 : Install ESPnet

- 2. Using ASR、TTS model provided by ESPnet

- 3. Pretraining and fine tuning TTS model

- 4. Training neural vocoder (Parallel WaveGAN, PWG)

- 5. Training by NTU student dataset

- GPU issues：

- **失敗原因**：原先嘗試的方法皆做到一半就走到死路 (e.g.,

- 1. 與東昇共用 1080Ti，但執行到上述 1-Step3 時電腦空間不足

- 2. Google Colab 有提供免費 GPU，但使用時間限制為12小時，且安裝到 1-Step2 時提供的空間不足

- 3. 再次嘗試 PowerAI，但因為 ppc64le 較為特別的架構，在 1-Step1 許多相依的套件無法順利安裝)

- **解決方案**：目前已與 Charlie 學長討論好，學長會協助調動資源，目前預計禮拜二 (10/19) 下班前會完成

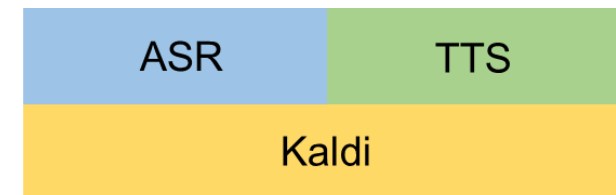


Fig. 1: The composition of ESPnet