

VCC 2020 reference design for Mandarin VC

Sian-Yi Chen

Advisor : Tay-Jyi Lin and Chingwei Yeh

Outline

Action item

完成 Mandarin VC (cascade ASR + TTS)

Status report

本周進度

- 完成將 VC 從 English 換成 Mandarin

- Demo

▣ 轉換對象

Source: 張明祥

Target: baseline female 1

▣ 轉換語句

使用 IASoC 錄製 320 句的其中 20 句作為輸入

▣ 轉換過程 (fig. 1)

1. 將明祥語料輸入 ASR model 並輸出 Mandarin 文字檔，目前會有錯字 or 漏字 (p. 4)
2. 將 Mandarin 轉換成拼音輸入 TTS model 並透過 PWG (parallel waveGAN) 合成語音

後續可優化方向

1. ASR 可能因為繁簡體差異，因此辨識率略為低
2. TTS 使用女生語料表現較男生語料好

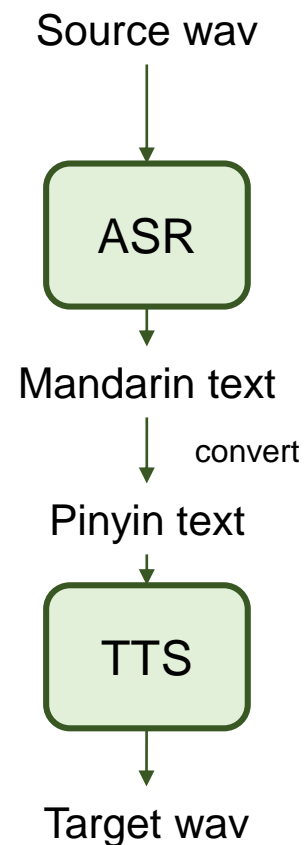


Fig. 1 Conversion process

Conversion result

■ Target 音色示例



摩洛哥農業在向歐洲出口時享受特殊待遇

■ Mandarin input 、 output



小朋友們在看螞蟻搬家
小朋友们在看嘛一般家



Source 01

Target 01



這個池塘裡養了很多魚
这个词汤里养了很多余



Source 02

Target 02



這裡將要建一座發電廠
这里将要建议做发电厂



Source 03

Target 03



他下山時被蛇咬了一口
她下算时被子咬了一口



Source 04

Target 04



他計畫今年買一台電腦
他计划今年买一台电脑



Source 05

Target 05



這間公司要招聘工程師
这间公司要招聘工程师



Source 06

Target 06

Conversion result

紅色為錯字 or 缺字

測試語句	ASR input	ASR output	
1	小朋友們在看螞蟻搬家	小朋友們在看嘛一般家	
2	這個池塘裡養了很多魚	这个词汤里养了很多余	
3	這裡將要建一座發電廠	这里将要建议做发电厂	
4	他下山時被蛇咬了一口	她下算时被子咬了一口	
5	他計畫今年買一台電腦	他计划今年买一台电脑	
6	這間公司要招聘工程師	这间公司要招聘工程师	
7	便衣警察抓到一名小偷	便宜检查抓到一名小偷	
8	他是那座大樓的設計師	是那做大楼的设计时	
9	他們擺好姿勢準備拍照	他们白好只是准备拍照	
10	今天老師宣布提前放學	今天老师宣布提前放学	

測試語句	ASR input	ASR output	
11	桌子上擺了一大盤瓜子	说__上百了一大盘瓜子	
12	中國萬里長城中外名聞	中国万里长城中外名文	
13	那片原始森林發生大火	那片延死森林发生大火	
14	棒球飛過來打破了窗戶	榜车飞过来打破的仓库	
15	他的哥哥養了一群白鴿	他的哥哥养了一群百哥	
16	他做完功課才上床睡覺	他做晚公课才上床睡觉	
17	一群孩子在那裏玩跳繩	一群孩子在那里玩跳省	
18	這裏的風俗習慣很特別	这里的风俗习惯很特别	
19	他腰痛的老毛病又犯了	他腰痛的老毛病又犯了	
20	他不小心把茶杯碰翻了	它不小心把茶被碰烦的	

附錄

Source 、 target 、 fine-tune sound


ASR input & TTS output




用於微調 TTS model 其中 5 句語料




摩洛哥農業在向歐洲出口時享受特殊待遇



不幸的是，其他人因為積怨分開



音像產業非常重要



可再生能源也不例外



當局已經裁掉了他們的工作

Execution details

- Demo cascade ASR + TTS 結果

- Demo 影片：<https://youtu.be/Szy09AIR-Ak>
 - ▣ 展示執行指令，進行 fig. 1 流程並生成音檔
- 輸入指令：`./run.sh --stage 11 --stop_stage 13 --srcspk ming_24 --trgspe TMF1 --tts_model_dir exp/TMF1_train_pytorch_train_pytorch_transformer+spkemb.tts1_en_zh`

1. 開始、停止 階段

stage 11: ASR model 辨識並輸出文字檔

stage 12: 將中文轉換成拼音，並將拼音文字檔輸入 TTS model、標記化 (tokenization)、解碼 mel 濾波器

stage 13: 使用 PWG(Parallel WaveGAN) 合成語音

2. 指定 **source**、**target speaker**

- source: 張明祥
- target: baseline female 1

3. 指定 **TTS** 微調預訓練模型

- 下一頁為 ASR input、ASR output、target wav 結果展示

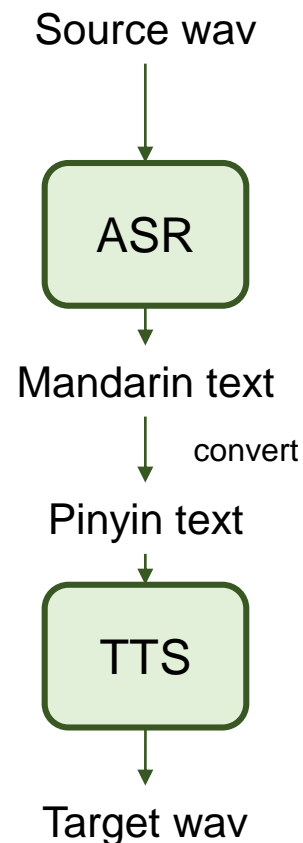


Fig. 1 Conversion process