

Summary

YUAN-FU LIAO
NATIONAL TAIPEI UNIVERSITY OF TECHNOLOGY

Universal Human-Machine Interface

Voice Assistant

- Who (**Jarvis?**) always listens for your call, anticipates your every need, and takes action when necessary

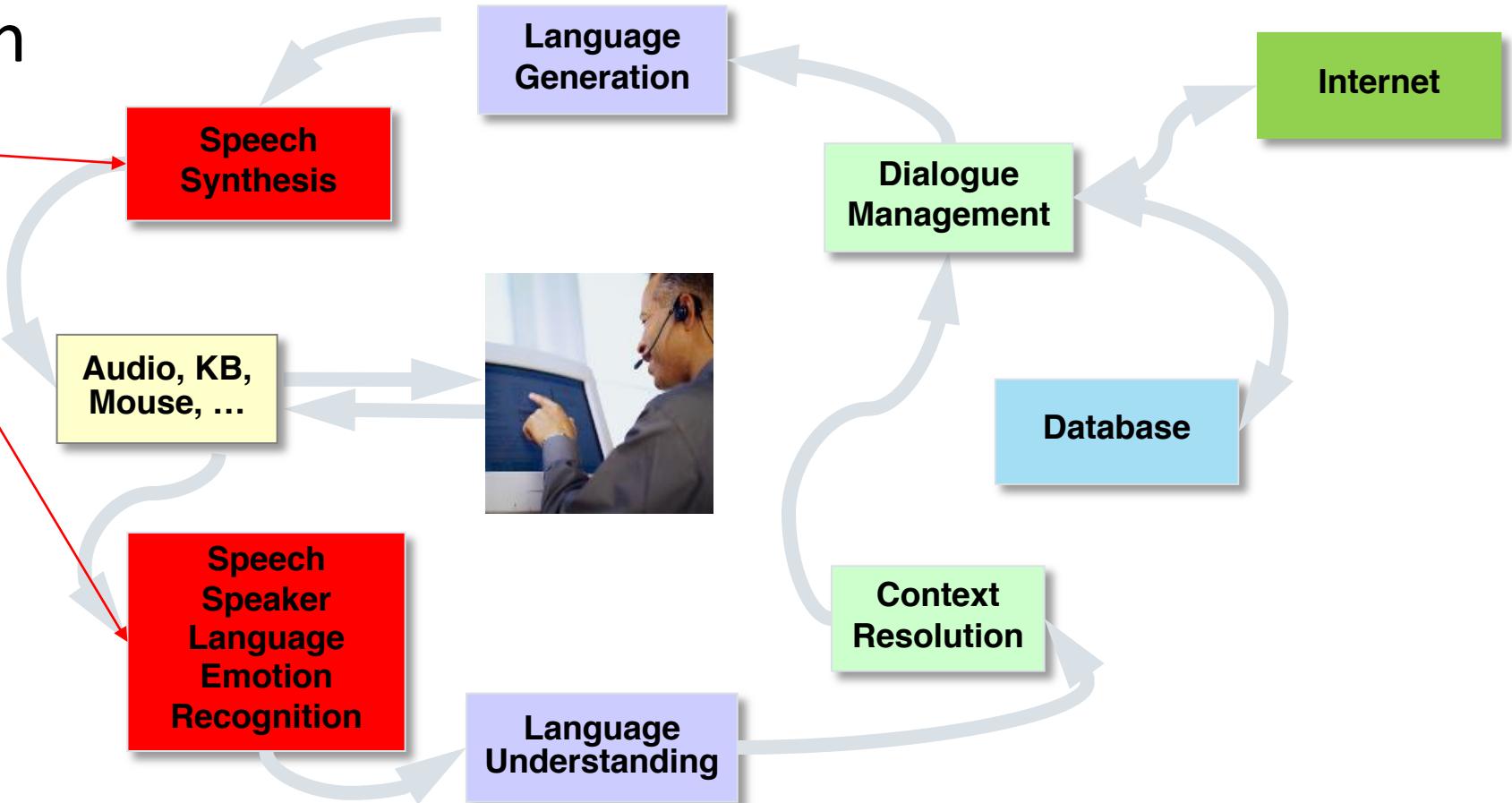


vs.



Toward Universal Human-Machine Interface

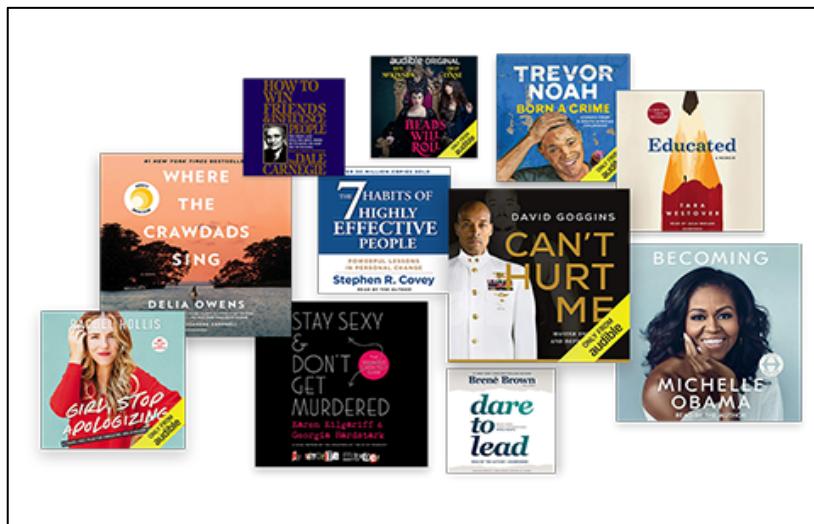
Speech Recognition
Speech Synthesis



Voice Bank 未來推廣

• iVoice.tw

- 與漸凍人協會合作
- 公司支援
- 即時字幕



Indexing

Google

網路搜尋檢索

個人語音電子書下載
即時字幕應用
會議記錄電子化



聲音日記社交網站

自動語音逐字稿轉寫

合成語音音檔與模型下載

網路媒體
網路課程
會議記錄

個人語音日記

即時配音
視障輔具

個人合成語音應用



^ Home

Corpus

✓ Challenge

Workshop

Publications

Formosa Speech in the Wild

The language habits of Taiwanese people are different from other Mandarin speakers (both accents and cultures). In order to boost the development of Taiwan-specific speech recognition techniques, This project will collect and transcribe Real-Life Spontaneous Speech from various sources.

Corpora

Activities

[Formosa Speech Recognition Challenge 2018](#)

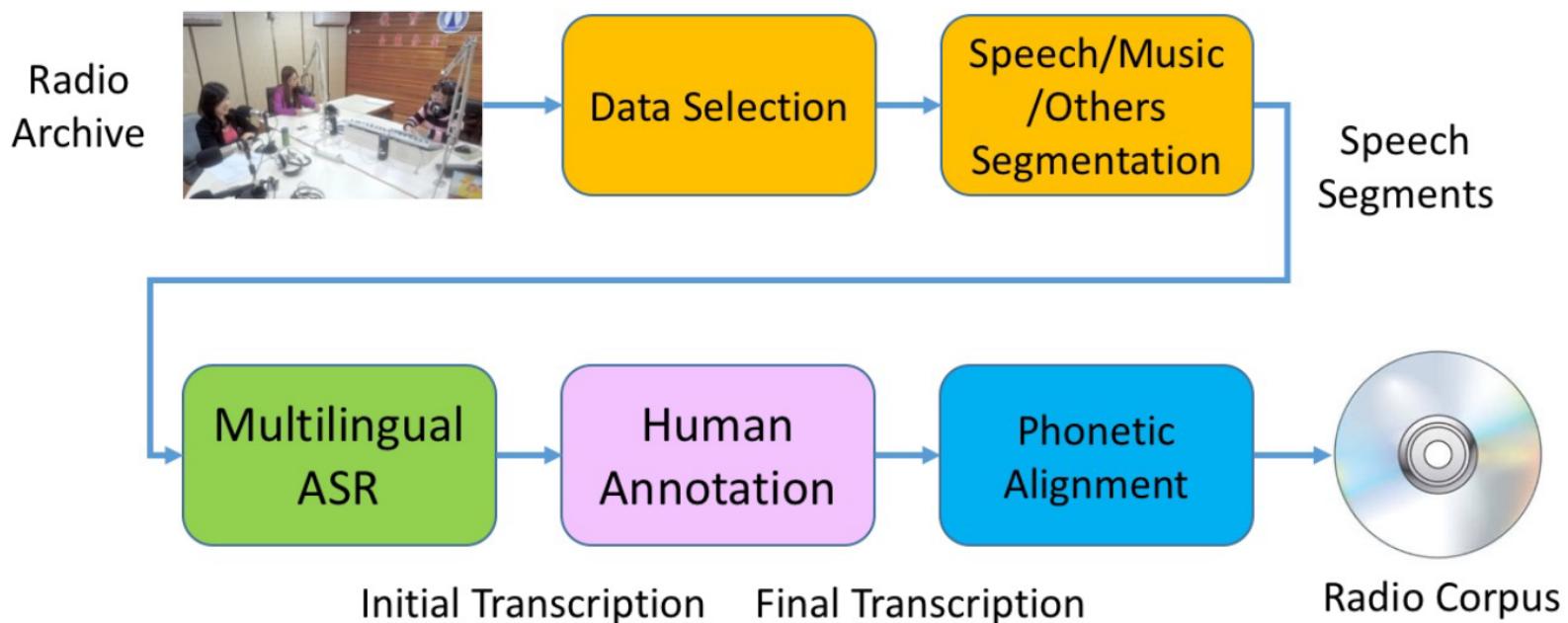
[Formosa Grand Challenge - Talk to AI 『科技大擂台_與AI對話』 Competition](#)

**Our main duty: construct an Large-Scale
Taiwanese Speech Corpus to support AI
development**

- URL: <https://sites.google.com/speech.ntut.edu.tw/fsw>

Semi-Automatic Corpus Processing

- National Education Radio Archive
- Reduce annotation effort as much as possible



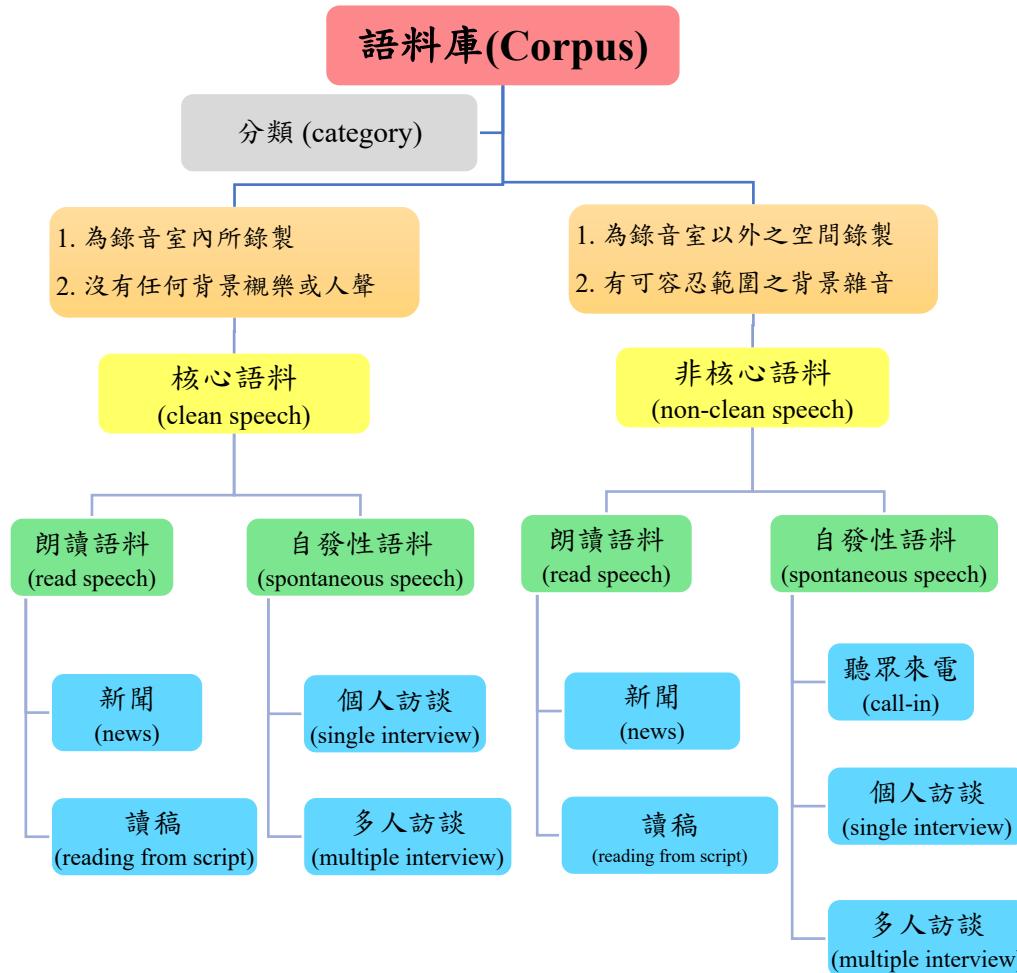
Corpus Design

- **Core (clean)**

→ Human Correction and
Annotation

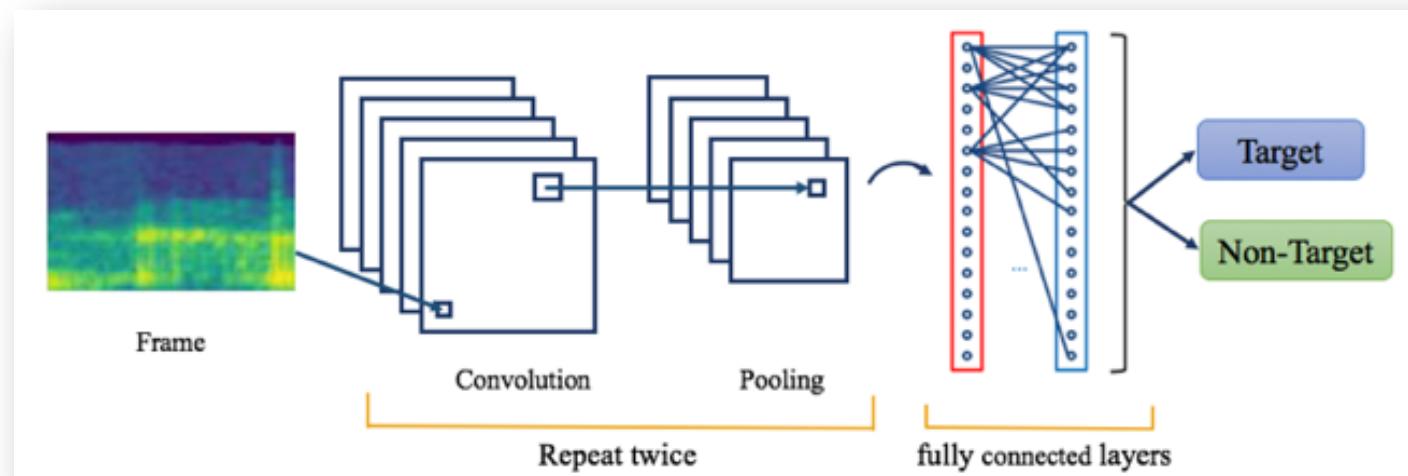
- **Non-Core (non-clean)**

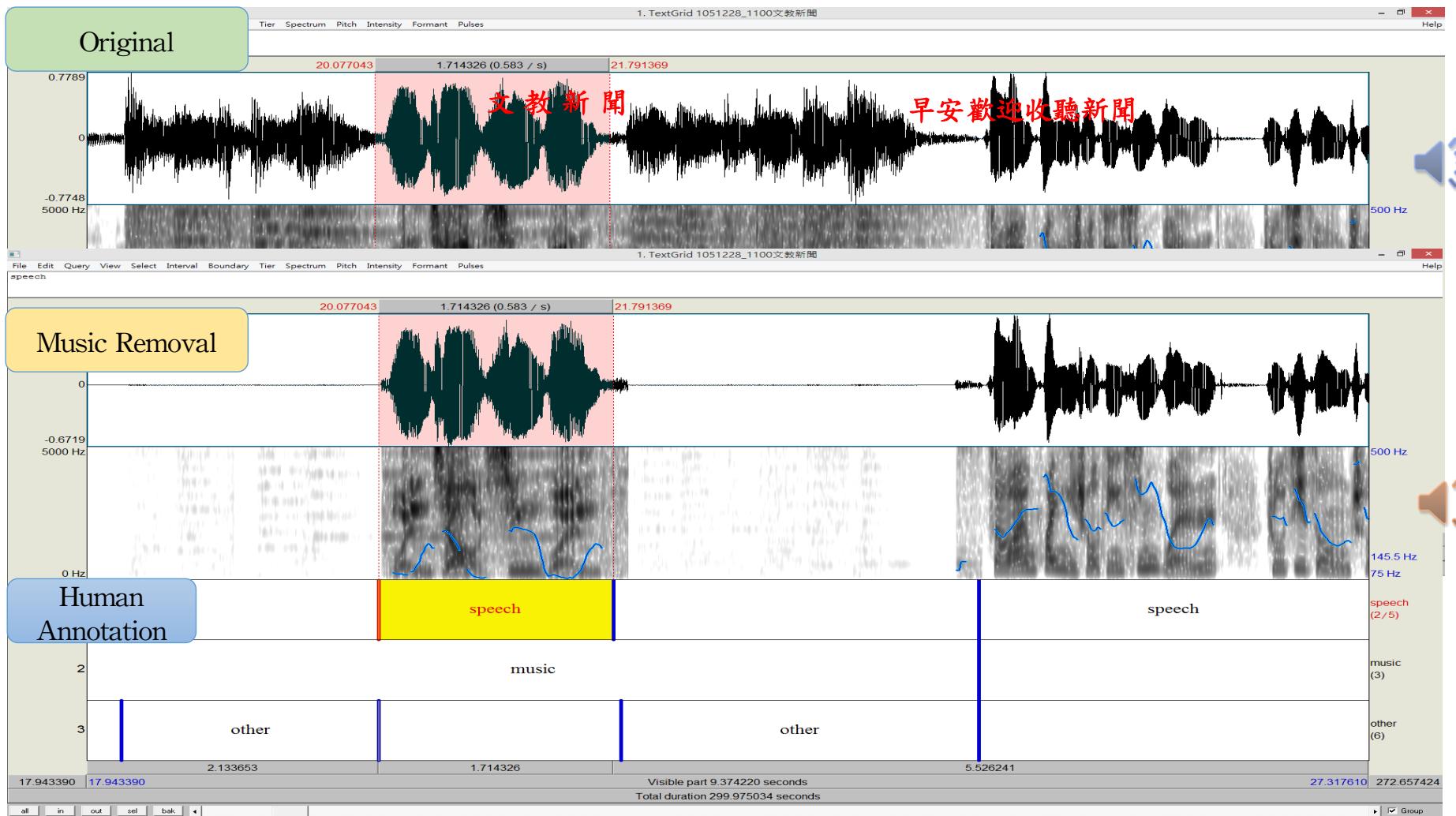
→ Semi-supervised training



Audio Event Detection

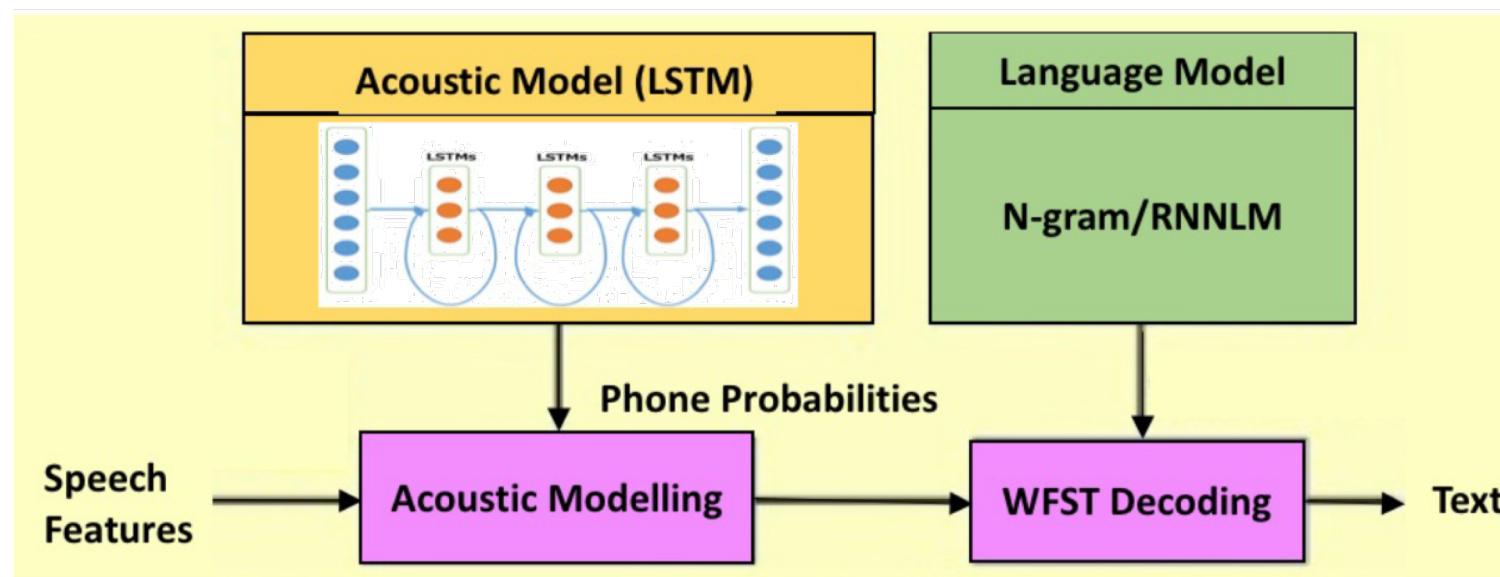
- Convolutional Neural Network (**CNN**) + raw **spectral** features
- Audio Event Database (manual **annotation**, ~60 hours)





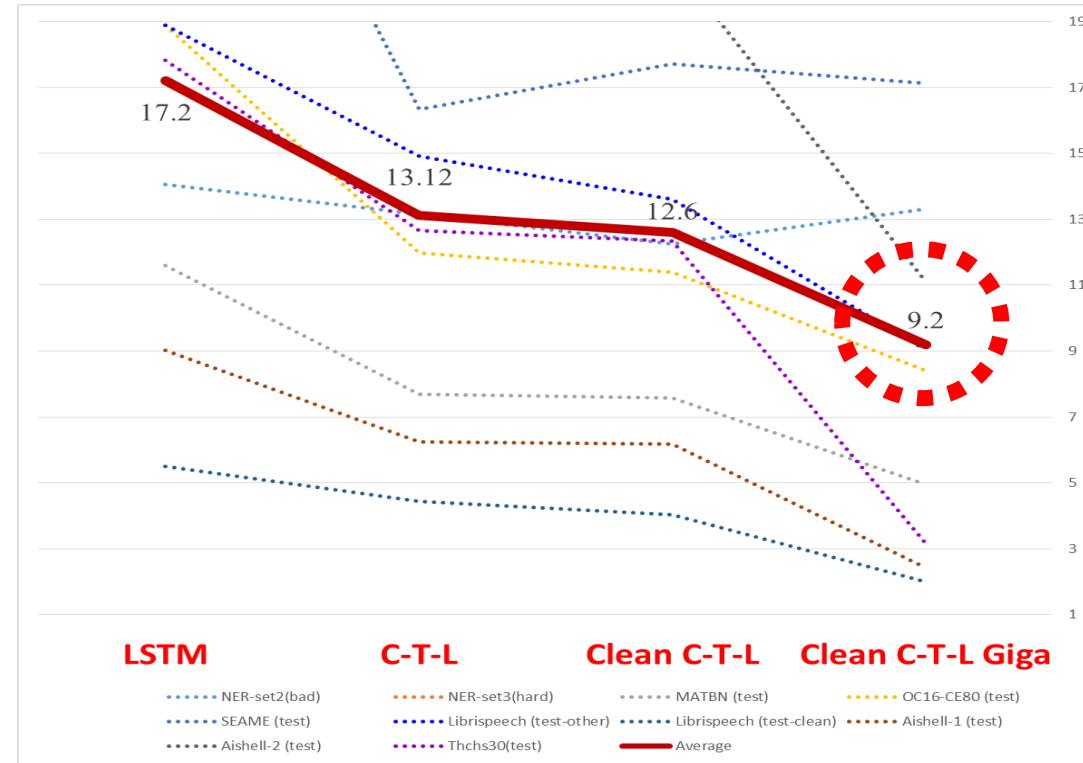
Automatic Radio Speech Transcription

- Mixed Chinese/English ASR
 - Trained using about **22 million words, 400 hours speech**

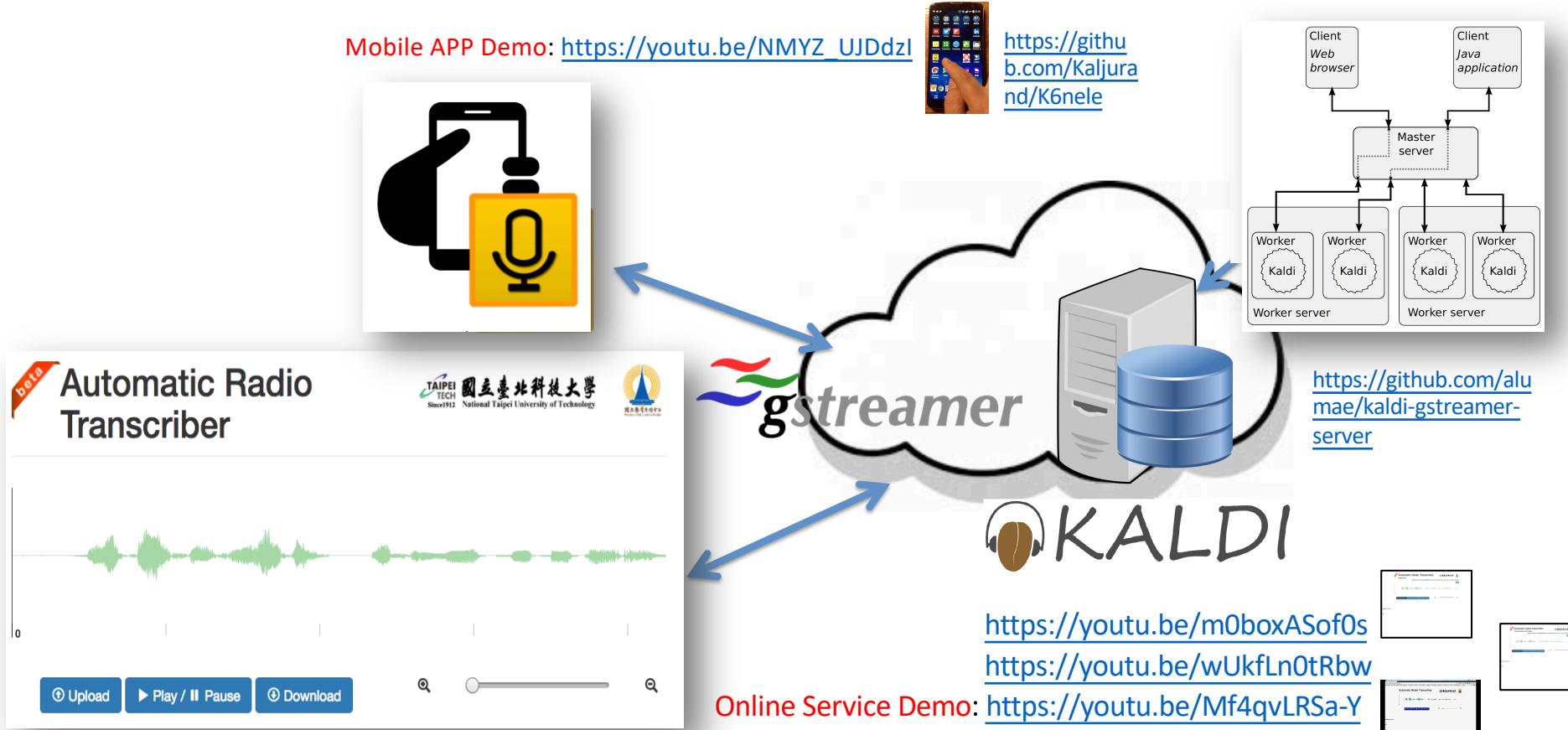


Performance of NTUT's Speech Recognizer

- CNN+(DNN+LSTM)*3
 - Conv1 → Conv2 → DNN → DNN → DNN → LSTM → DNN
 - DNN → DNN → LSTM → DNN → DNN → DNN → LSTM
- 800 hours Training data
 - MATBN
 - NER-Trs-Vol1
 - SEAME
 - LibriSpeech
 - AiShell
 - OC16-CE80
 - Thchs30

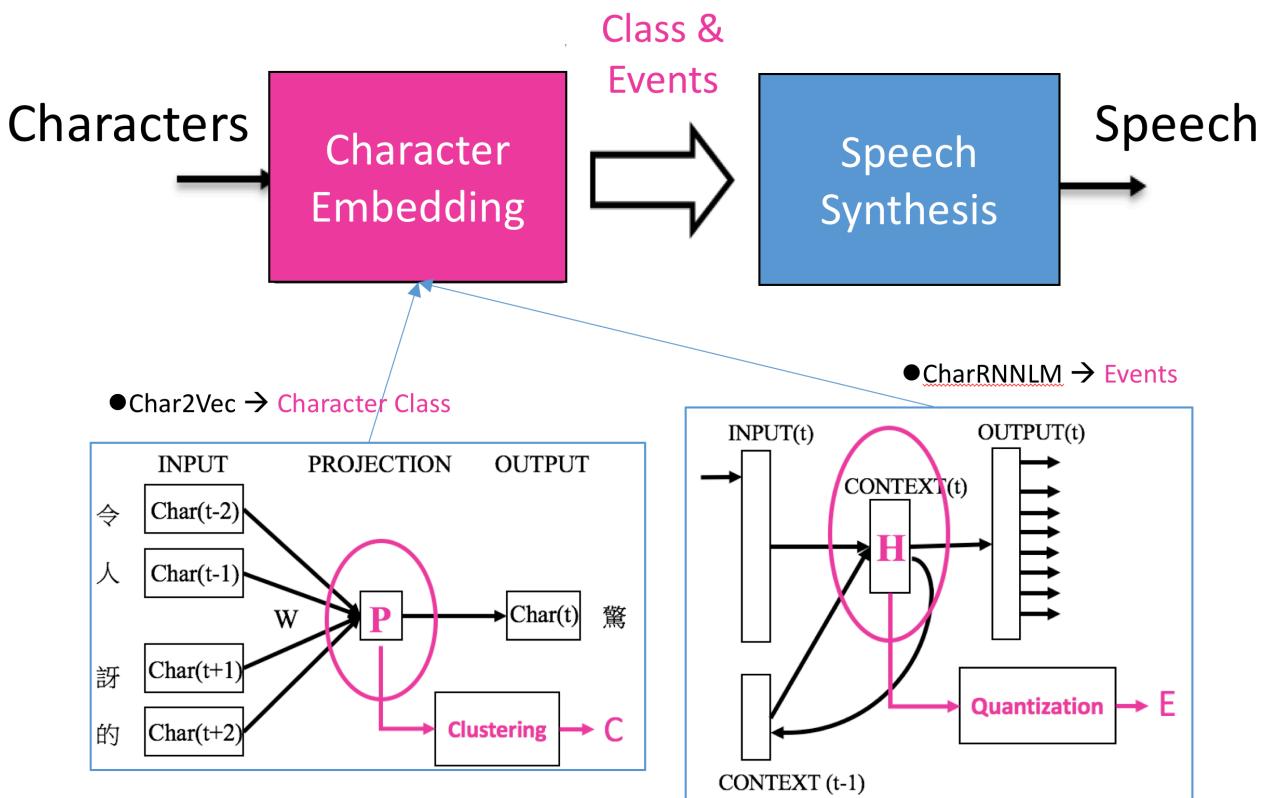


NTUT's Online Speech Recognition Server



NTUT's Speech Synthesizer

- Mixed Chinese/English TTS



- Chinese



- Mixed Chinese-English

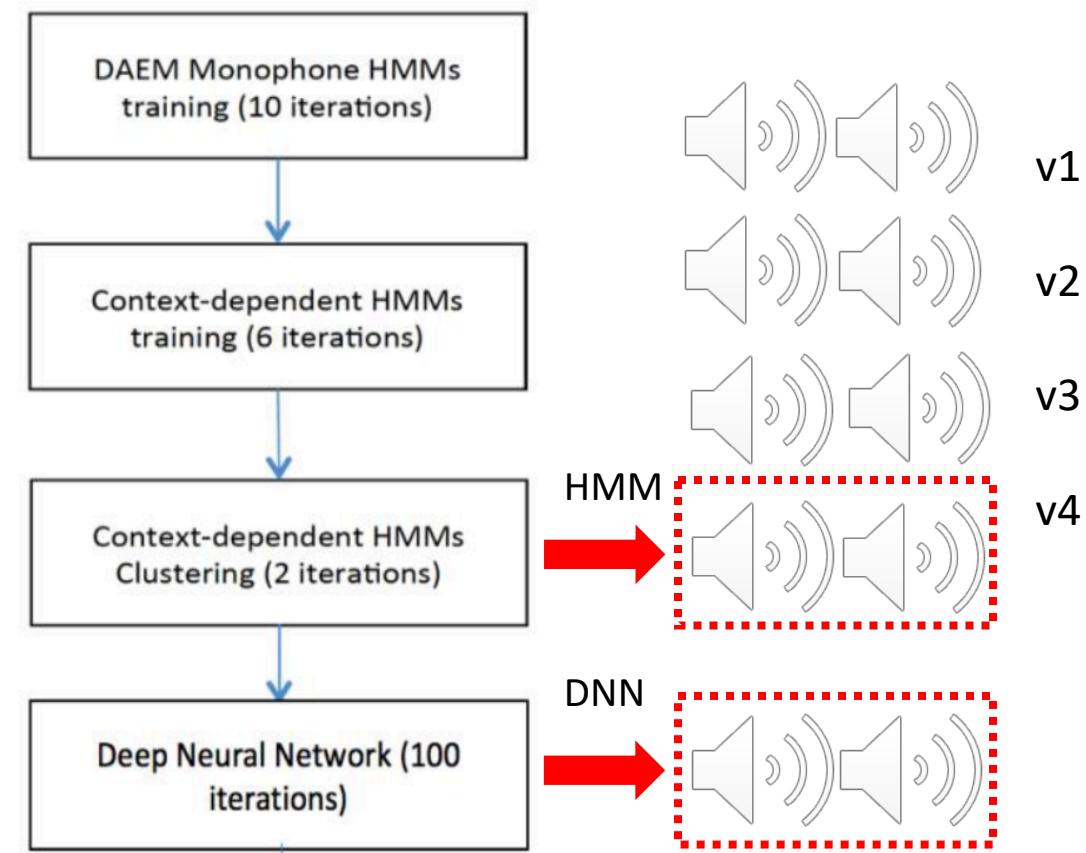


Speech Synthesis Backend (Improvement, 1/2)

- Chinese TTS
 - HMM, 34-dim. MGC+F0, MLSA, Speaker-Dependent ASR
 - DNN, 60-dim. World Vocoder (on-going)



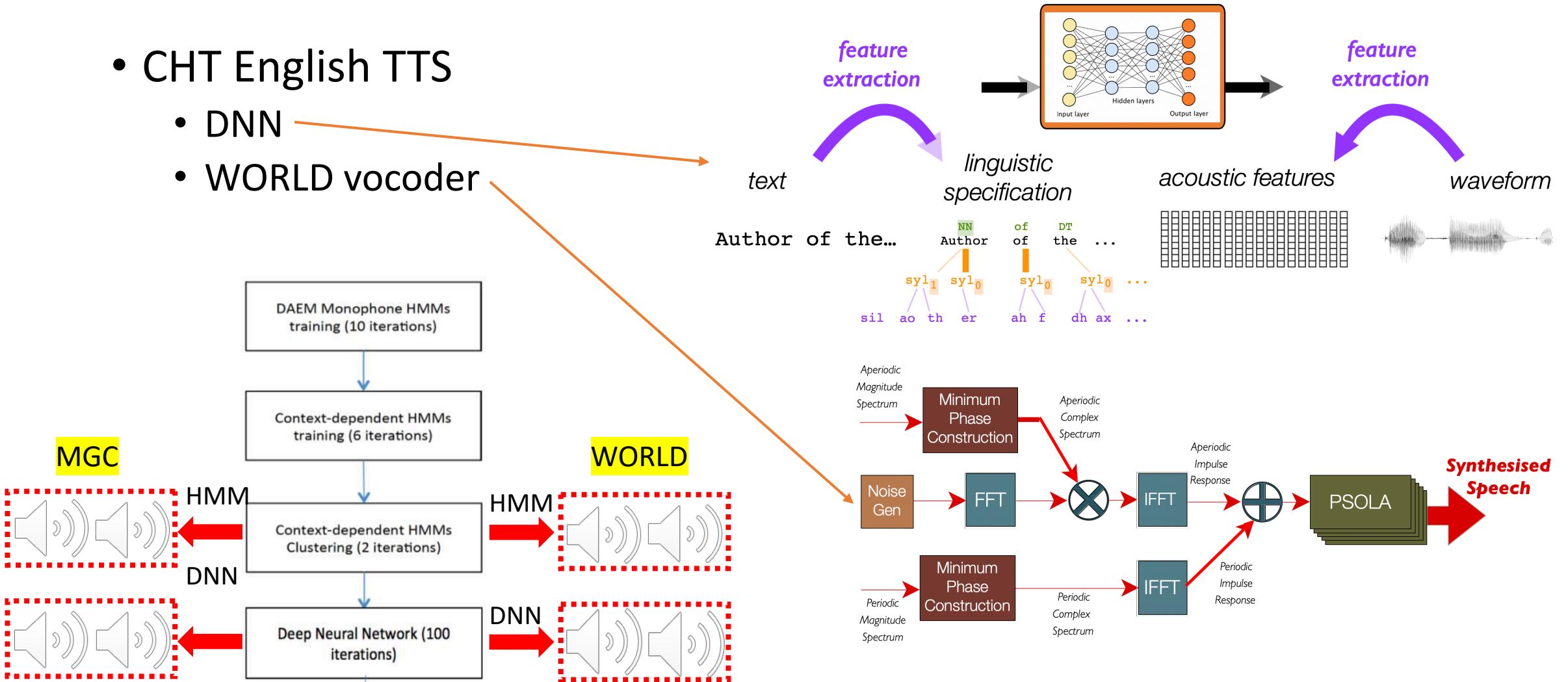
前一段时间我看到一篇文章，说上世纪六零年代的“大逃港”时，香港人真是好，全民动员救助逃过去的吃不饱饭的大陆人，可见香港人的精神境界有多高。有一次，我问一位老香港人，说这事是真的吗？他犹豫了一下说：真的，但是没有他们说的那么崇高。怎么回事呢？你想，香港原来就是个小渔



Speech Synthesis Backend (Improvement, 2/2)

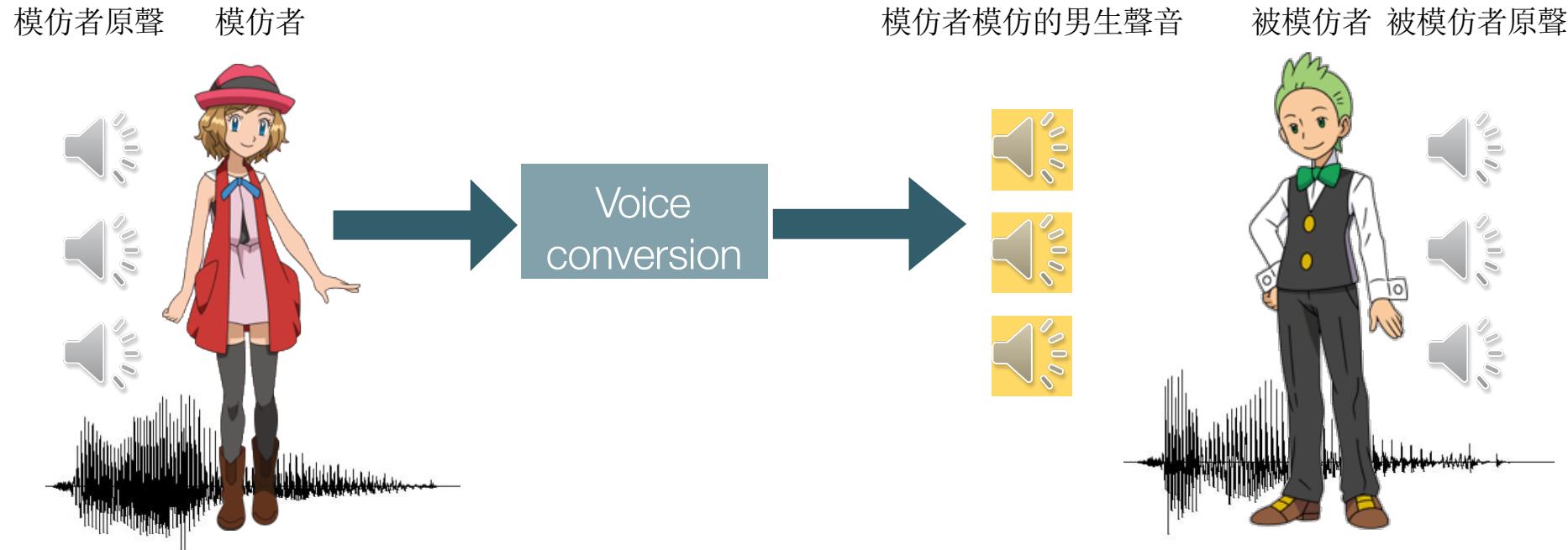
- CHT English TTS

- DNN
- WORLD vocoder



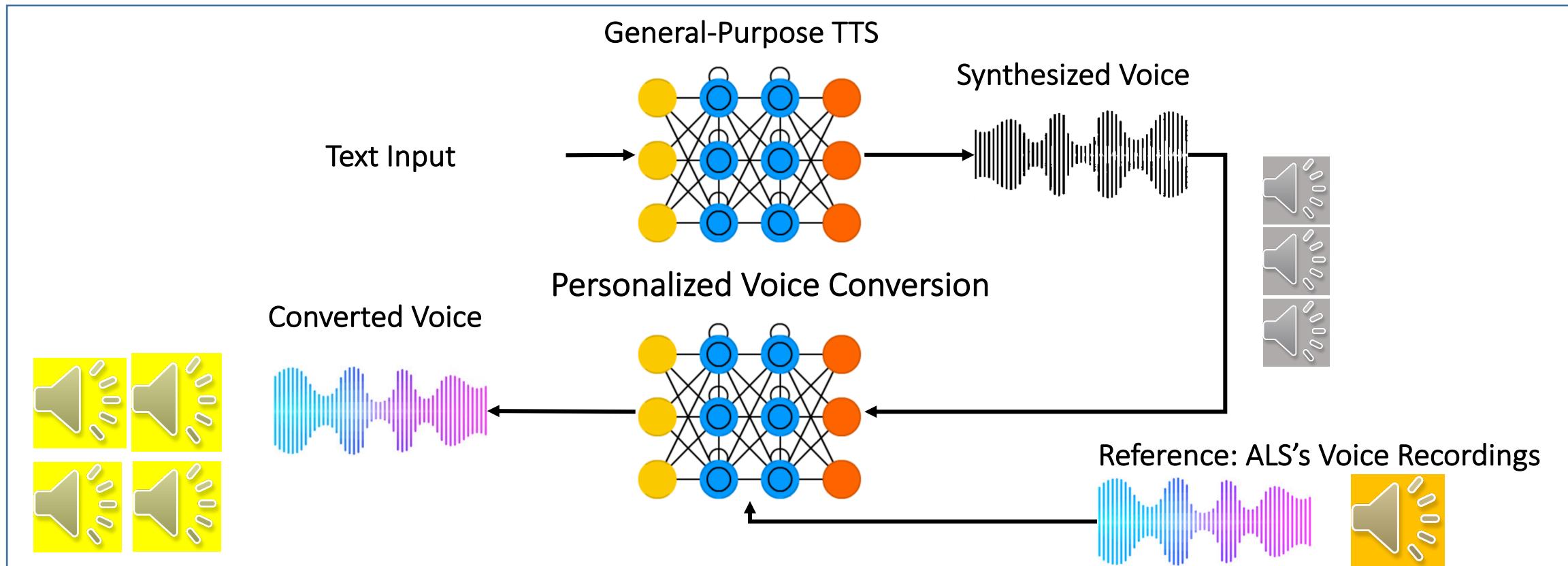
NTUT's Voice Conversion

- Manipulate source speaker's voice to sound like target without changing language content



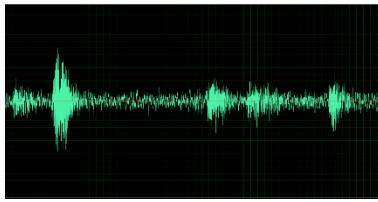
Personalized Voice Output Communication Aid (VOCA) for ALS (2/2)

- Real Case
 - Only Few Minutes Speech, Low-Quality Recordings (8 kHz, Noisy, MP3)

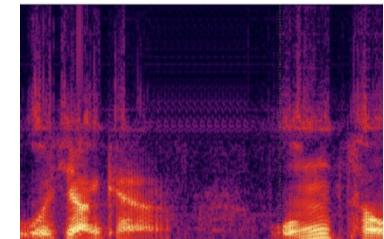
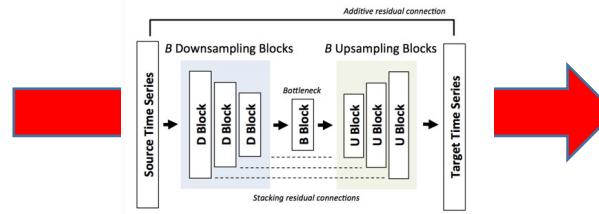
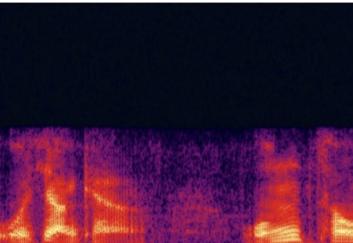


Approaches for Low-Quality, Noisy Samples

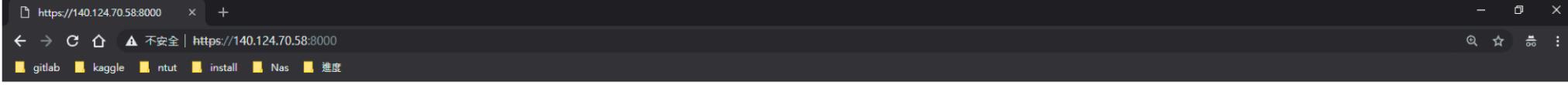
- Auxiliary Corpus
 - High-Quality Speech Corpus (20 hours, ~60 speakers) + Background Noises
- Denoising Frontend
 - WaveNet (**noisy vs. clean** speech pairs)



- Bandwidth Expansion Backend
 - U-Net (**8 vs. 16 kHz** speech pairs)



NTUT's Question/Answering System



(臺北)臺北市五星級國賓飯店,昨晚發生男子.疑似不耐久候電梯,踹門摔落電梯井的意外!男子送醫後,仍因傷重.宣告不治;事發原因.正由檢警調查!記者 錦傑 報導

臺北市消防局週三晚間9點多獲報:中山北路國賓飯店,一名參加公司尾牙男子,自二樓意外摔落地下一樓.電梯井;於是立刻出動包括特種車在內的7部車輛,19人.前往救援;不過.男子被救出時,因為頭步嚴重受創,已經失去生命徵象;送往馬偕醫院急救,仍因傷重不治!國賓飯店公關.黃國瑋說~VOICE

轄區臺北市警局中山分局初步調查:這名42歲的鄒姓男子,週三晚間參加所屬塑膠公司年終尾牙,到.晚間9點多,略帶酒意,在2樓搭電梯離去時;疑似因為久等不耐煩,踹電梯門,導致滑條鬆脫開啟,摔落地下一樓電梯井!事發後,轄區中山警分局,已調閱現場監視錄影畫面,釐清意外原因!而.國賓飯店上午也發布聲明表示:對此意外遺憾難過,向死者家屬表達慰問哀悼;並將全力配合檢警調查.釐清事發原因!

Ans: 中山北路 國賓 飯店

請問男子墜落意外的發生地點於哪家飯店？

Get Answer

NTUT's Toy #1

- English ChatBot
 - <http://140.124.72.159:8888>



The screenshot shows a web-based chat interface. At the top center is a circular profile picture of a white robot with a yellow antenna, set against a green background. Below the profile picture, the text "Machine Learning @ NTU" is visible. The main area contains a conversation between a user and the bot:

I'm a Chatbot trained on movie dialogue (beta)

Let's gossip about movies :)

Bot: I asked you.

You: Do you love me?

Bot: Eighteen.

You: How old are you?

Bot: California. oakland.

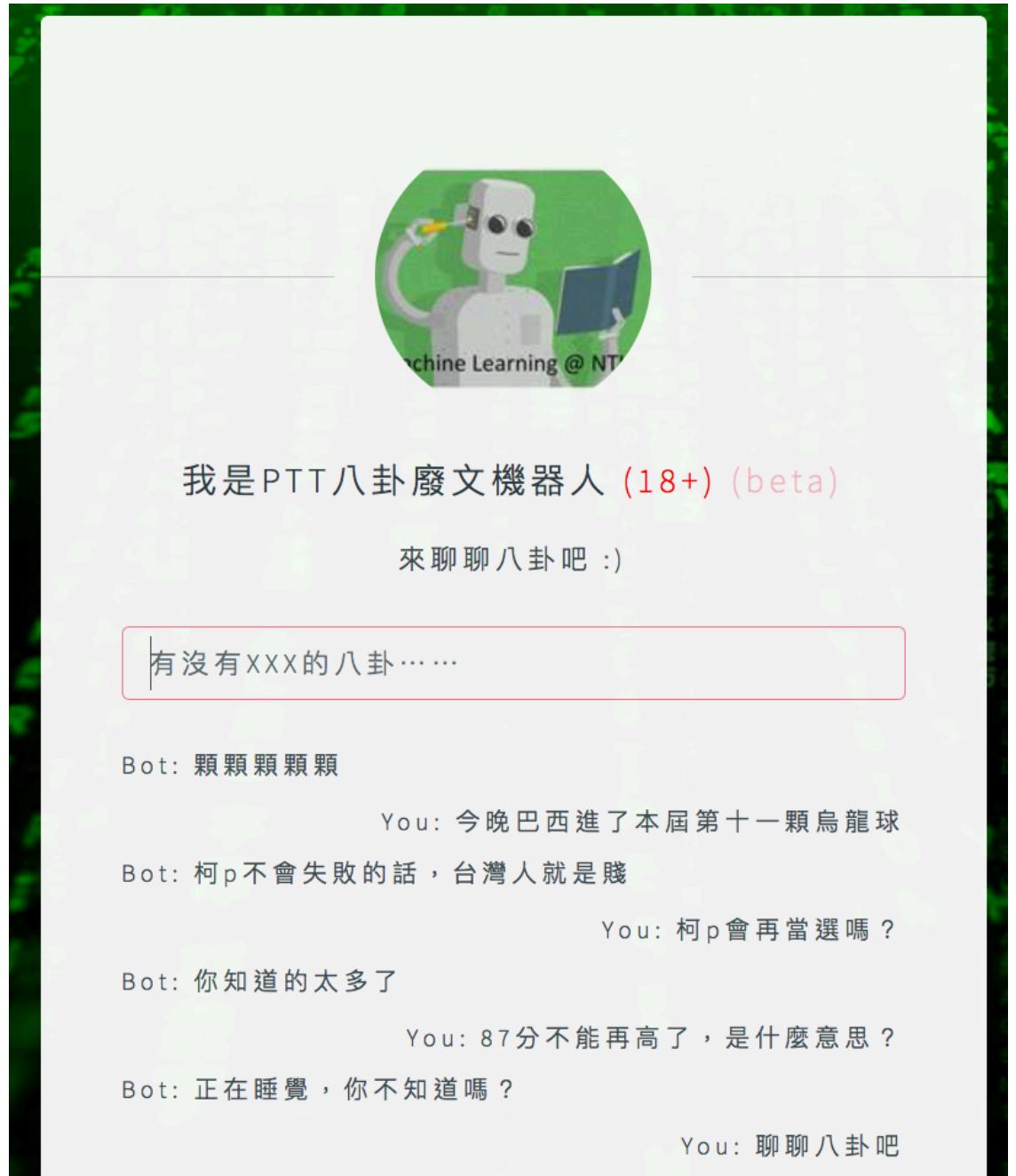
You: Where do you come from?

Bot: The disco boys.

You: Who are you

NTUT's Toy #2

- Chinese ChatBot
 - <http://140.124.72.159:8889>



NTUT's Toy #3

- Chinese Text to Taiwanese Romanization Translator
- <https://140.124.72.159:8990>

