Zero-shot Text To Speech With Singaporean Accent

Student: Li Haoyang   Supervisor: Assoc Prof Chng Eng Siong

Objectives:
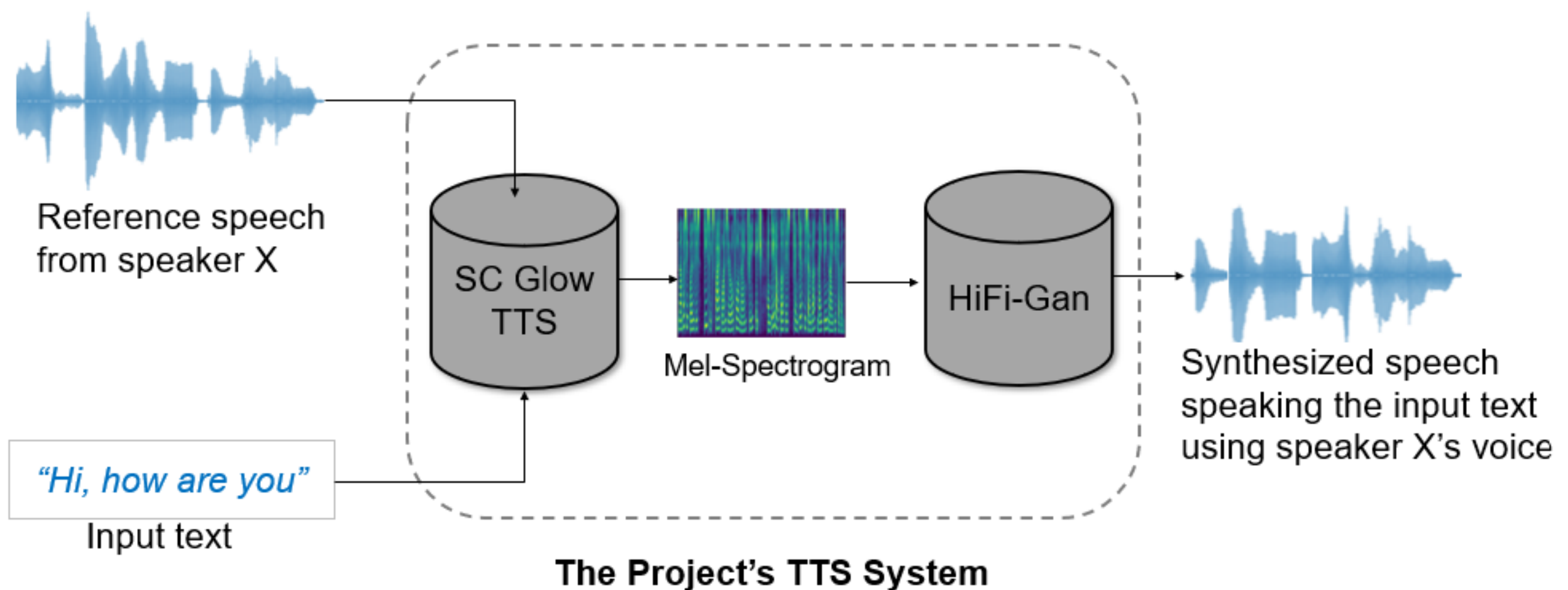
Develop a text-to-speech (TTS) system that can

➤ Generate speech in Singaporean accent

➤ Adapt quickly to voices unseen during training

The developed TTS can

➤ Enhance Singaporean users' experience on using TTS services, since there is currently lack of development on zero-shot TTS for Singaporean accented speech

Methodologies:

➤ Transfer learning using pretrained TTS model

➤ Domain adaptation via fine tuning pretrained TTS model on Singaporean accented corpus

The Project's TTS System

Results:

Evaluation on speech quality

Source of audios	MOS
Ground-truths	4.81
Fine-tuned TTS	4.14
Baseline TTS	2.52

Evaluation on speaker similarity for speaker adaptation

Source of audios	Cosin similarity
Ground-truths	1.000
Fine-tuned TTS	0.933
Baseline TTS	0.861

Observations:

➤ Domain adaptation significantly improves both the speech quality and speaker similarity