

IEEE Spoken Language Technology Workshop



SLT 2022

Happening on
9-12 JANUARY 2023

Doha, Qatar

Voice-Preserving Speech Machine Translation

<https://www.youtube.com/watch?v=g7tuzM4YHLA>

Voice-Preserving Speech Machine Translation

- **SLT Hackathon 2022**
- **Mentor:** Amir Hussein
- **Members:**
 - Do Tri Nhan
 - Nguyen Ngoc Minh Khanh
 - Nguyen Ngoc Bang Tam
 - Igli Mlloja



Motivation

- Language **barrier** with **locals** when traveling to Vietnam
- Vietnamese is a tonal language with 6 **different tones** -> difficult for non-native speakers to **absorb**

=> We propose a cross-lingual speech translator

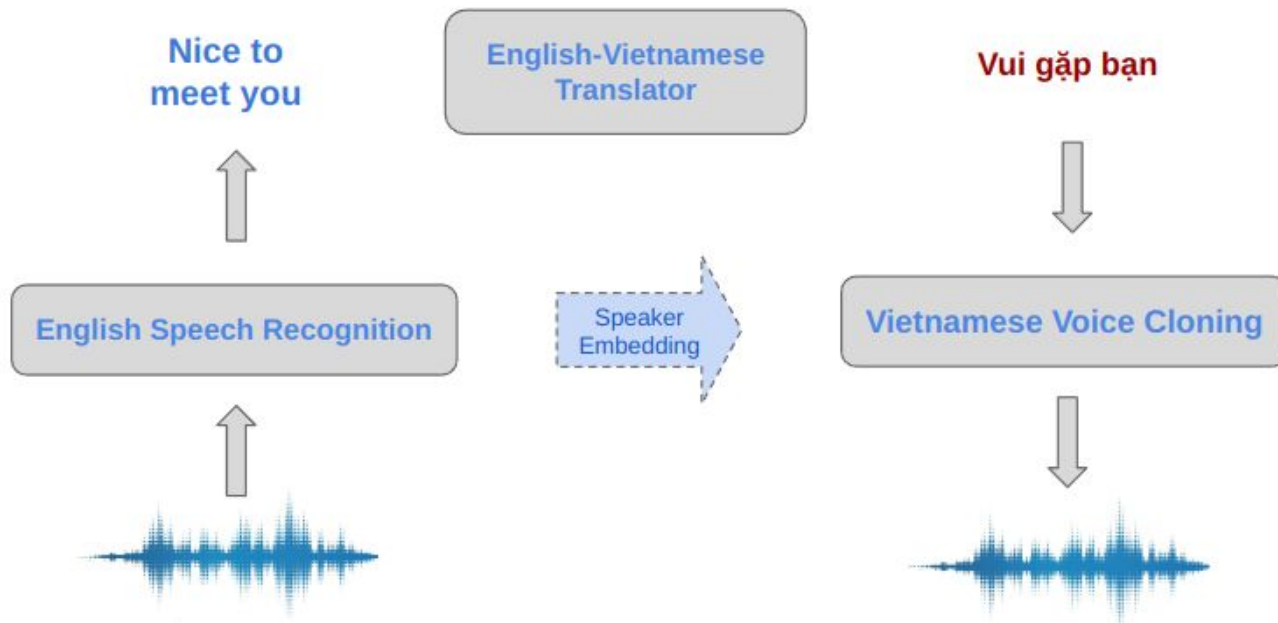


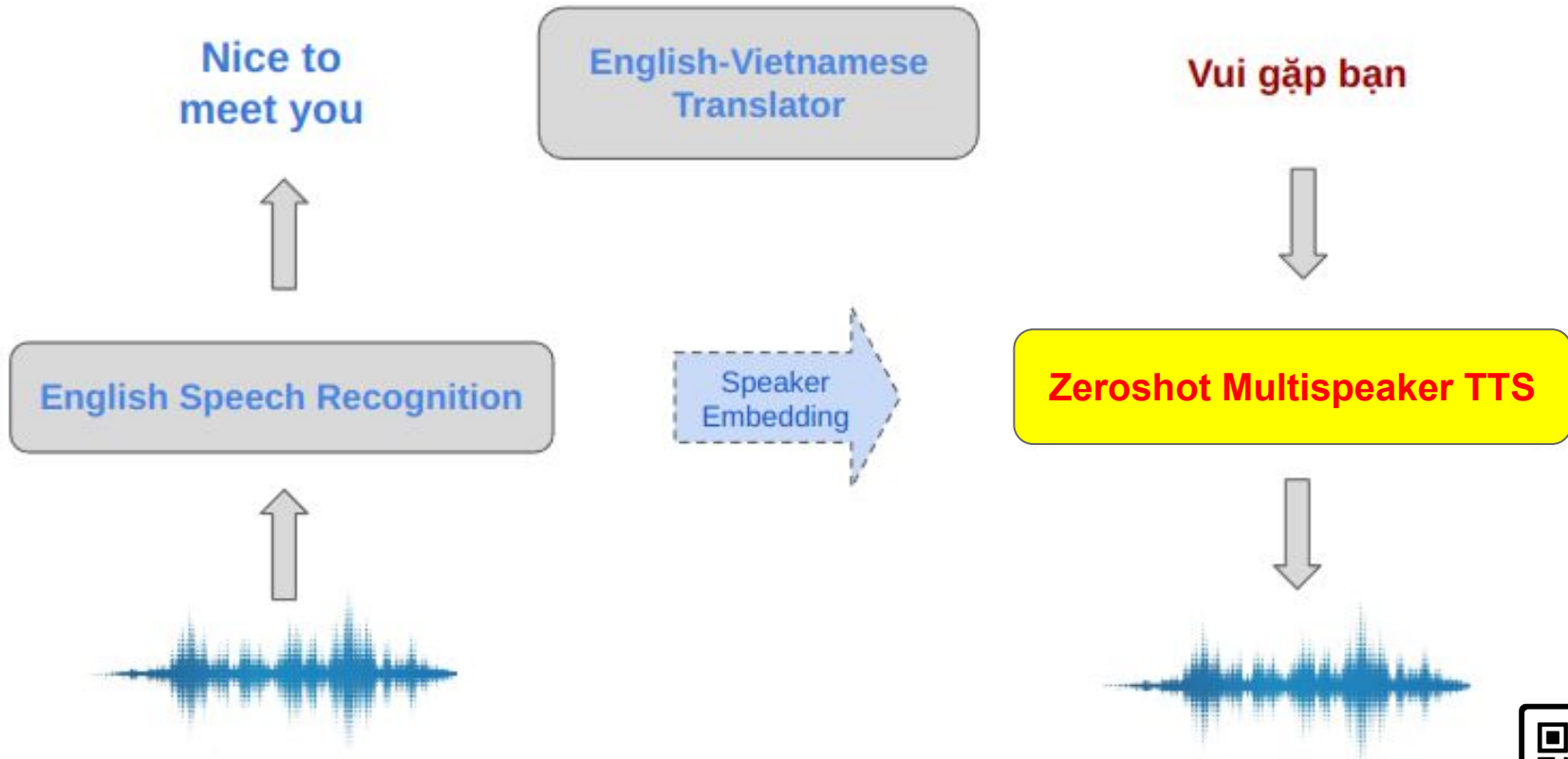
Main Idea

- Speech Translation for **Low Resource Language** with Voice I/O and **Preserve the Characteristics** of the Voice Input



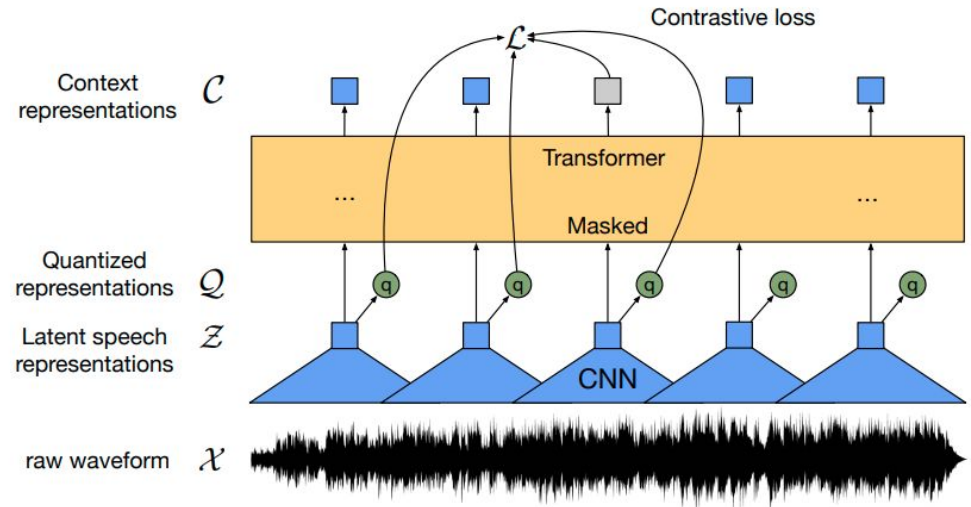
Main Idea





English/Arabic ASR module

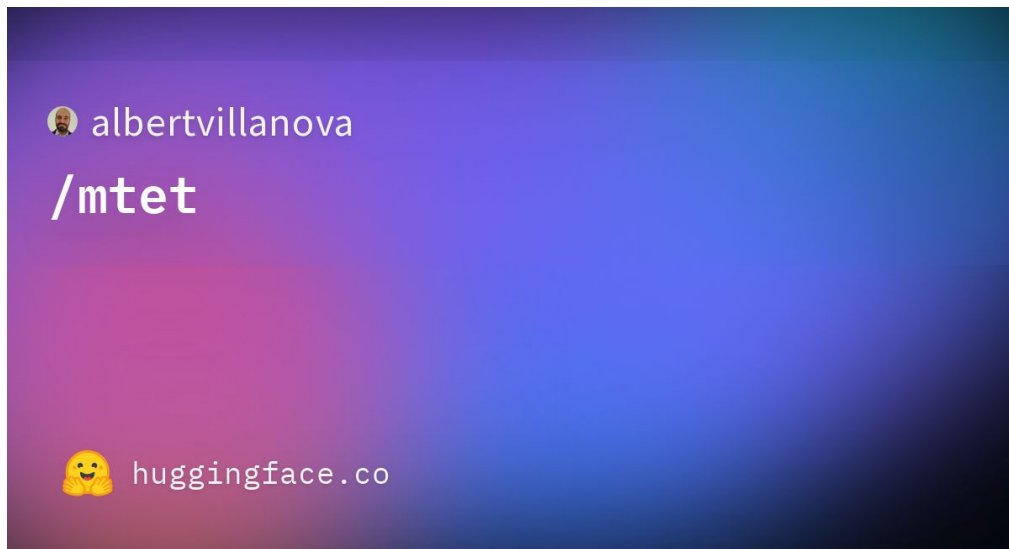
We intend to use large pretrained English wav2vec2.



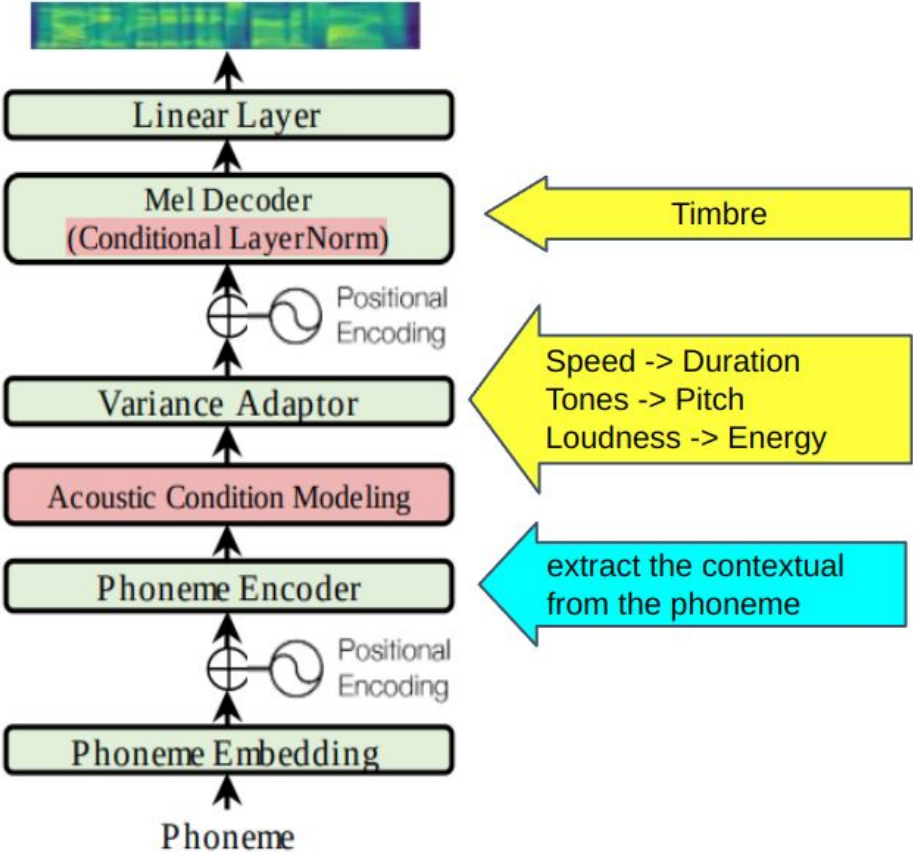
English-Vietnamese machine translation module

MT module from HuggingFace

MTet: Multi-domain Translation for **English-Vietnamese**: uses a larger and better quality Machine Translation dataset

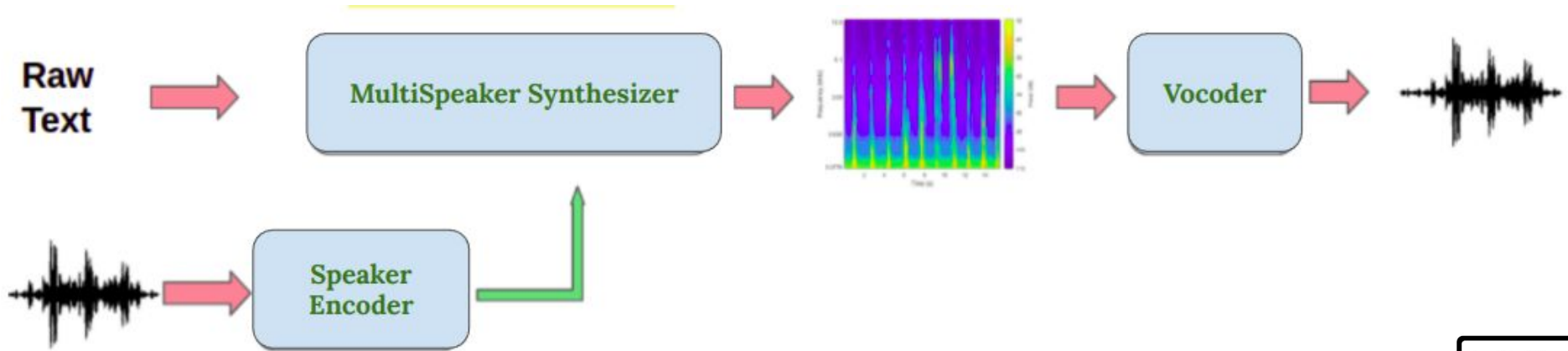


Vietnamese TTS module



Voice Cloning

- we build the multi speaker speech synthesizer based on finetuning **Fastspeech2** then inject the **speaker embedding of English speaker** to convert the voice output.



ASR

MT

TTS

Voice Preserving Speech Machine Translation

Machine Translation Input

Machine Translation Output

3000 Characters Remaining

TRANSLATE

ASK

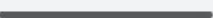


PLAY

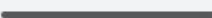




Sample Demo

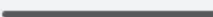


Source - EnScript - ViScript - CloneVoice

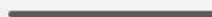


908-157963-0013.wav

▶ 0:00 / 0:04   

▶ 0:00 / 0:03   

260-123440-0012.wav

▶ 0:00 / 0:05   

▶ 0:00 / 0:02   

Realtime Demo

Source Music

Speaker audio with 16000Hz - 5s long



Drag and drop file here

Limit 200MB per file

Browse files

Translate My Voice



Thanks for yours listening

- ★ Thanks for the enthusiasm of our mentor
- ★ Thanks for sponsoring the organizers



- **Motivation:**

- What is the motivation for the project? What problem do you solve? Why is it important and for whom?
- Why is voice preservation important? (male/female voice preservation etc)
- What language are you focusing on and why?

- **Approach:**

- Describe the system pipeline on high level?

The proposed system consists of 3 core modules as illustrated below: (1) ASR to transcribe English audios to corresponding texts, (2) English-Vietnamese machine translation, and (3) voice cloning-based TTS which can generate Vietnamese-spoken audios using the original speaker's voice.

- Describe the speech recognition model part in details? What model you used? How you trained/finetuned the model? What data you used and how did you get it?
- What challenges have you faced to develop a speech recognition model? Language specific challenges: Vietnamese language is tonal language etc
- Describe the TTS model part in details? What model you used? How you trained/finetuned the model? What data you used and how did you get it?
- What challenges have you faced to develop a TTS model?
- Describe the voice conversion model part in details? What model you used? How you trained/finetuned the model? What data you used and how did you get it?
- What challenges have you faced to develop a voice conversion model?

TODO TASK

- Write a script for Igli Mlloja
- Make a video for backup presentation and show demo
- Edit slide: detail the model and technique

Instructions

- Please **edit directly on this google slide deck**. During the presentation, you will use a provided laptop for the presentation.
- The final presentation should consist of **3 min presentation + 1-2min QA from judges**. Please stick to the time as we will stop presentations that exceed 5 min.
- In your presentation please consider the following:
 - Goal of the project and what social or economic impact could it create
 - What it makes interesting and/or innovative ?
 - Challenges you have overcome
 - What have you learned from it ?
 - What makes the project special or gives your proposal an edge over similar solutions in the market ?

TIPS and guidelines

- Please do not copy the contents from other materials (if it is very difficult to redraw, it is acceptable with the appropriate citation information).
- It depends on the audience, but it is a good idea to spend some time clearly presenting the introduction/motivation/problem setups
- Use a simple picture to emphasize your method/concept
- Long sentences in slides are not a good idea
- If you are showing numbers, please extract important numbers or highlight important numbers
- Add a take-home message in your final part