

# 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

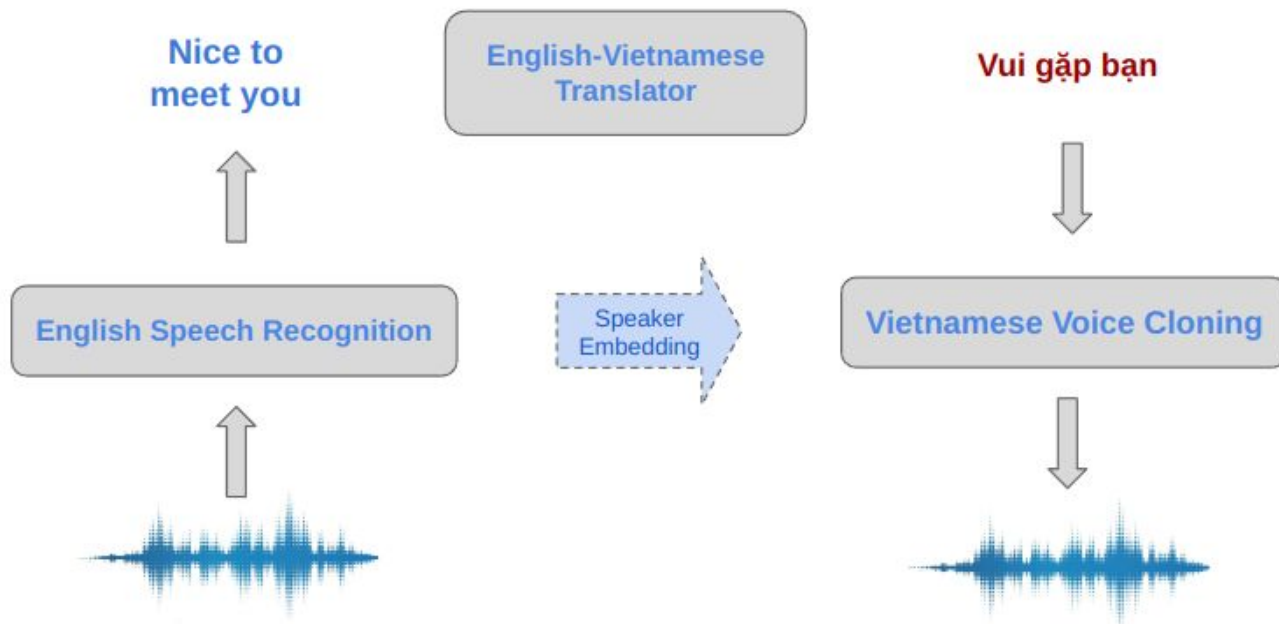


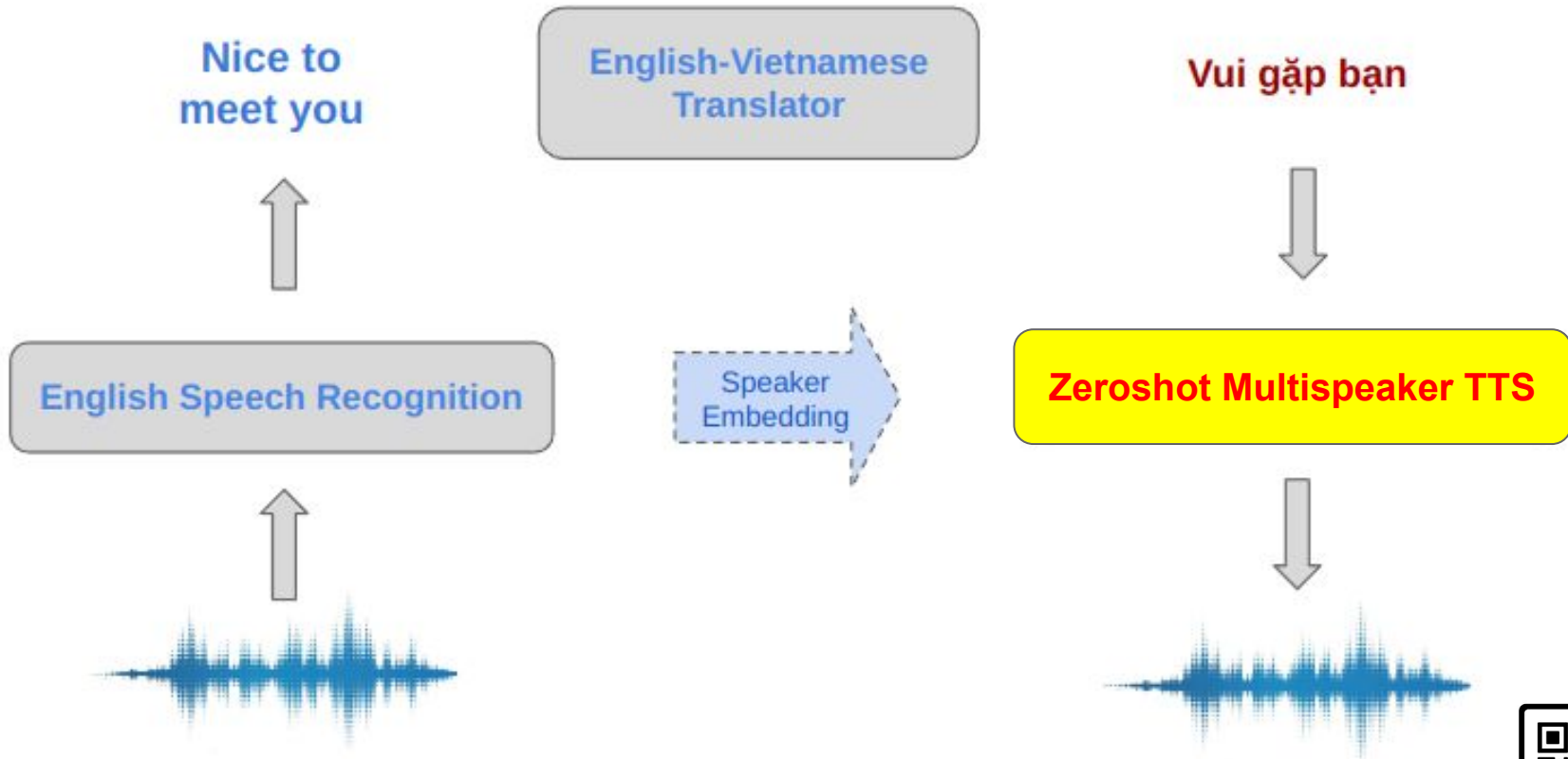
# Main Idea

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



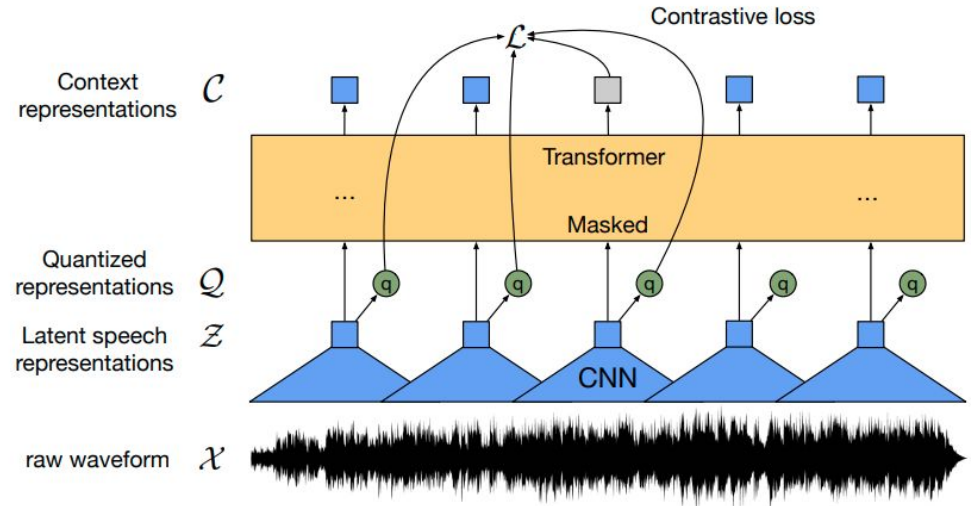
# Main Idea





# English/Arabic ASR module

We intend to use large pretrained English wav2vec2.

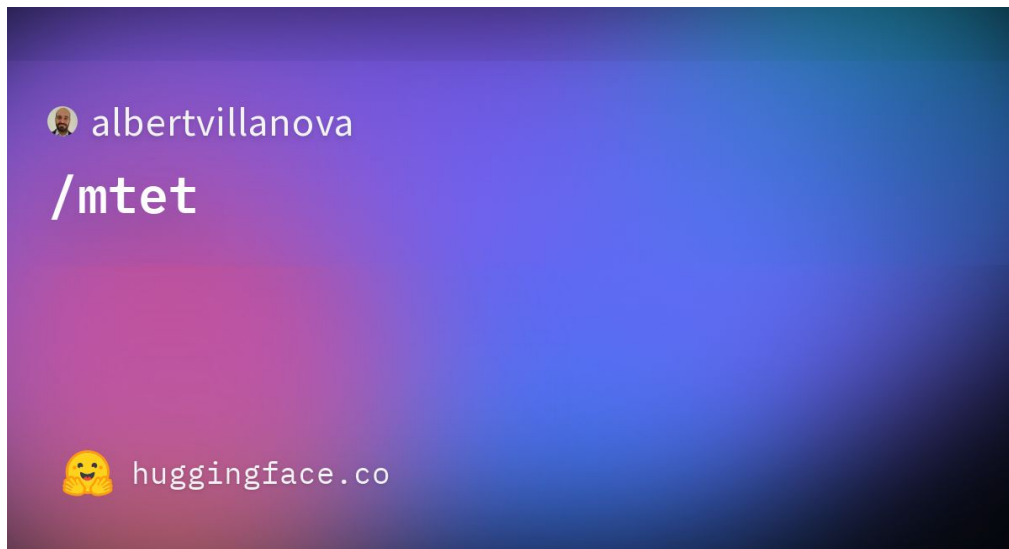




# English-Vietnamese machine translation module

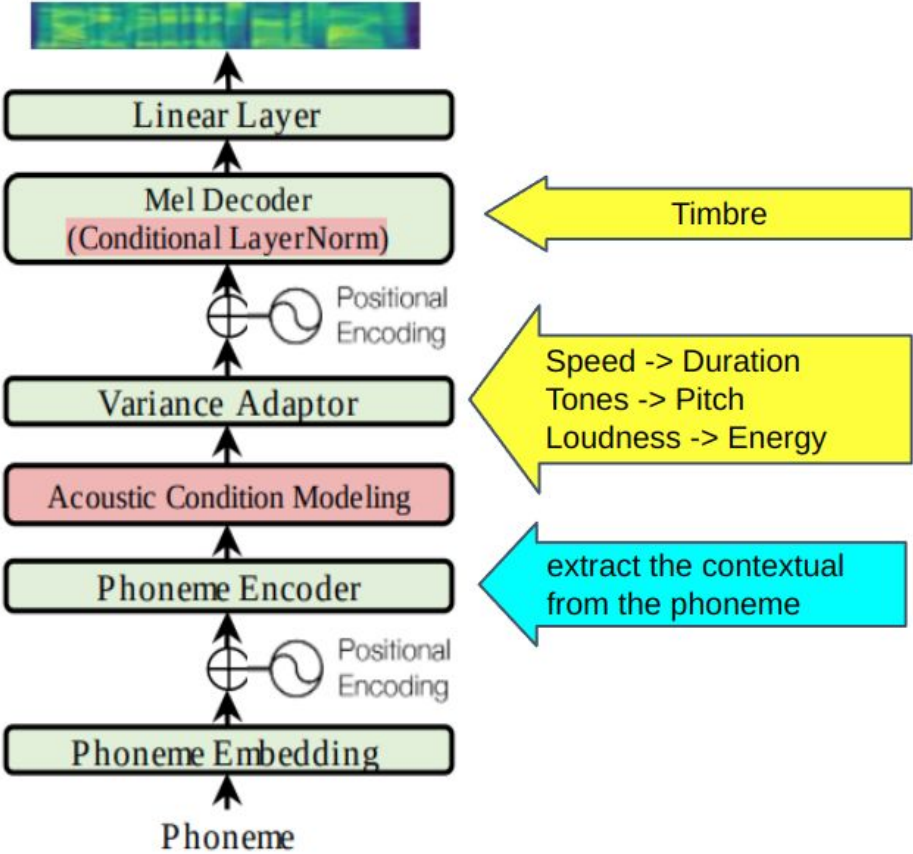
## MT module from HuggingFace

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





# Vietnamese TTS module





ASR

MT

TTS

## Voice Preserving Speech Machine Translation

Machine Translation Input

Machine Translation Output

3000 Characters Remaining

TRANSLATE

ASK

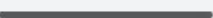


PLAY

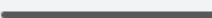




## Sample Demo




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

