SPEECH RECOGNITION FOR VOICE BASED MACHINE TRANSLATION

Government Engineering College, Wayanad

March 4, 2024

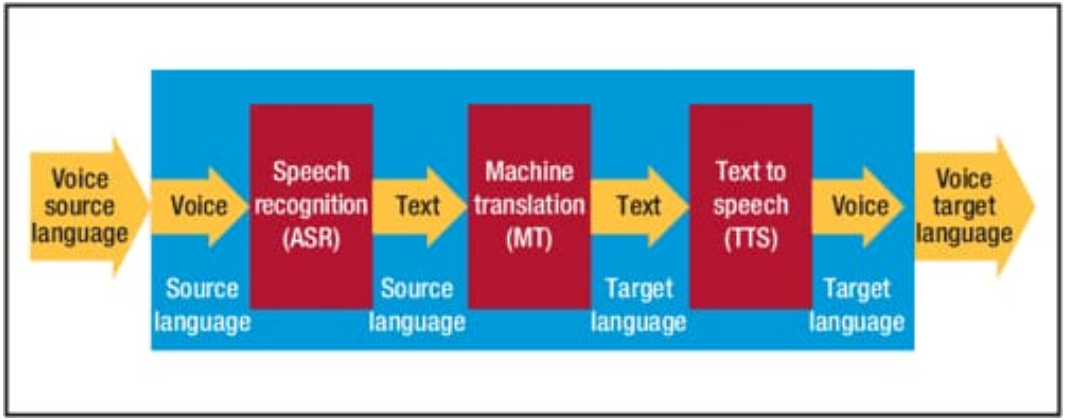
Guide: Prof. Anish Babu

TABLE OF CONTENTS

INTRODUCTION

- Speech recognition for voice-based machine translation involves 3 main steps:
  1. voice recognition
  2. machine translation
  3. speech generation
- Advanced technologies such as Automatic Speech Recognition are essential for accurately transcribing spoken words.

METHODOLOGY

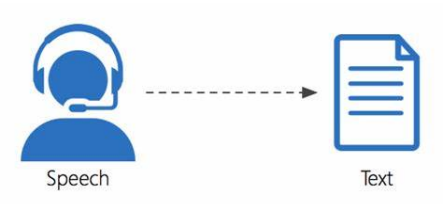


Block diagram of voice translation

METHODOLOGY

1. **Speech To Text**

- SILERO MODEL- high accuracy and efficiency, making them suitable for real-time speech recognition tasks.



METHODOLOGY

- **Installation**:Start by installing silero model from Torch.hub, PyTorch library in python environment
- **Speech Input**:Provide the audio input to the selected silero model
- **Transcription**:The model will process the audio input and generate the corresponding text output
- **Output**:Display the transcribed text
- `https://github.com/snakers4/silero-models?tab=readme-ov-file#speech-to-text`

METHODOLOGY

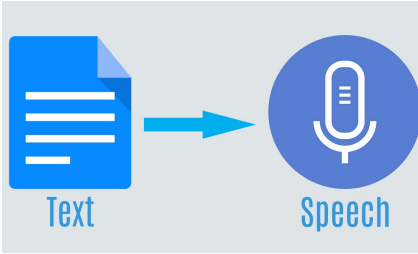
2. **Machine Translation**

- Use an existing package- 'translate '
- The package supports a wide range of languages for translation.
- translate is to make translations in a simple way without the need of bigger effort and can be used as a translation tool like command line
- Translate output in real time
- `https://pypi.org/project/translate/`

METHODOLOGY

3. **Test To Speech**

- **gTTS** - Google Text-to-Speech
- It is a Python library and command-line tool
- Uses the Google Text-to-Speech (TTS) API to convert text into speech
- Easily generate speech files from text strings in various languages.
- `https://pypi.org/project/gTTS/`



METHODOLOGY

Using gTTS :

The gTTS library is a straightforward tool

That converts text into audio, which can be saved as an MP3 file.

Install the gTTS API by running this command in terminal

The output of the program will speak the phrase

DISADVANTAGES

- Silero model supports only a limited amount of languages.
- Duplex communication is impossible with certain combination of languages.
- An appropriate model must be downloaded for the usage of a particular language.

RESULT

FUTURE WORKS

- UI/UX with HTML CSS
- Real time audio recording with Pyaudio
- Create chat-rooms with `https://socket.io/`

CONCLUSION

- The speech recognition system for voice-based machine translation shows promise in overcoming language barriers. By integrating technologies like ASR, MT, and TTS synthesis, real-time translation was achieved.
- Limitations include Silero's language support and duplex communication challenges.
- Future work may involve user-friendly interfaces, PyAudio for audio recording, and socket.io for chat rooms.

REFERENCES

- Multilingual Speech to Speech Translation System in Bluetooth Environment [M.D. Faizullah Ansari, R.S. Shaji]
- Speech Recognition for Voice-Based Machine Translation [Tiago Duarte, Rafael Prikladnicki, Fabio Calefato, and Filippo Lanubile]
- Speech-to text conversion using silero model [Bill Dally and Nita patel]

THANK YOU