Real-Time Transcription and Translation for Code-Switched Speech in Lecture Video

Presenter : Dr. Hay Mar Soe Naing

University of Computer Studies, Yangon, Myanmar

Background :

- Multilingual phenomenon has become increasingly common in real life, working environment and education.
- Increasingly globalized learning environments, lecture videos have become a vital medium for knowledge dissemination.
- Teachers and students often employ the **use of intra-sentential code-switching while teaching** Information Technology (IT) subject.

Problem Statement :

- **Specific Challenge**: This is a significant problem for low-resource languages.
- **Technical Failure**: Traditional transcription system struggle with such mixed-language **input.**
- **Real-Time Barrier**: Real-time translation systems cannot effectively handle code-switched conversations.
- **Real-World Impact**: This gap creates barriers in multilingual communities, education, and business.

Research Objectives

- Build a **Real-Time Transcription and Translation Engine** for code-switched speech in video content.
- Leverage Multilingual AI Models to significantly enhance speech recognition accuracy.
- Integrate a Context-Aware Translation Mechanism that handles mixed-language input.

Proposed Method:

A technical solution would be to integrate –

- Multilingual ASR (pre-trained model e.g., OpenAI Whisper)
- Language identification
- Context-aware segmentation
- Cross-lingual Neural Machine Translation into a low-latency streaming pipeline (e.g., streaming Whisper)
- Post-processing to improve accuracy and readability

Impact:

- Speech technology is transforming digital learning and accessibility worldwide.
- Often employ the use of intra-sentential code-switching while teaching IT subject.
- Frequently alternate between languages to explain the meanings of individual IT technical terms or to differentiate between similar terms for students.

Impact (Cont'd.):

- Code-switching speech data is scarce, especially in academic and technical domains in Myanmar
- Aims to tackle the challenges of code-switching in automatic speech recognition, especially in IT education.
- Bilingual/ Multilingual subtitling for input lecture contents.

Analysis of Mix-language Lecture Speech :

- Existing Myanmar ASR system only focus on mono language.
- Analysis of mixed-language lecture speech based on 10 hours durations with 10 lecturers

Total Utterances	11, 945 utterances
Total Duration	10 hours
Mixed-language utterances	8,142 utterances
Percentage (%)	68.16%

Scope of the Project:

🌐 **Initial Language Pairs**: Initial focus was on Myanmar speeches with English terminology lecture videos.

📥 **Input Types**: Compatible with both recorded files and live streams.

📝 **Core Function**: Automatic transcription of mixed-language audio.

⚡ **Key Feature**: Real-time translation capabilities.

🎞️ **Deliverable**: Generation of bilingual/multilingual subtitles especially for ASEAN languages.

Output/Outcome:

- A Toy Model for automatic speech recognition (ASR) in Myanmar-English code-switched lecture speech.
- Based on around 10 hours mixed-language lecture speech including 10 lecturers.

Type	Speaker	Utterance	Duration
TRAIN	10	8,452	7.5 hours
TEST	6	3,493	2.5 hours
TOTAL	**10**	**11,945**	**10 hours**

Preliminary ASR Results:

- Build the pipeline ASR with GMM-HMM speaker adaptive training and DNN model using Kaldi Toolkit.
- Use Statistical N-gram language model in decoding.
- Evaluated on ~2.5 hours lecture speech (3,493 utterances) containing 6 lecturers.

ASR model	Word Error Rate(%)
GMM-HMM (LDA-MLLT) model	31.95
GMM-HMM (SAT) model	31.31
DNN model	**26.69**

Conclusion:

- To captures natural intra-sentential code-switching from actual teaching settings, especially in IT education
- To support the development of codeswitching ASR systems by offering realistic and varied speech data
- To enable the **End-to-End ASR models** for bilingual academic contexts
- Plan to grow the generation of bilingual or **multilingual subtitles using cross lingual NMT** capabilities

Summary

We provide real-time transcription and Translation for Code-Switched Speech in lecture video.

SUPPORT US TODAY!

Contact Us:

haymarsoenaing@ucsy.edu.mm

+95 9 794309926

Website: www.ucsy.edu.mm