Hindi Speech to English Text Translation System

VOXLING
Debanjan Nanda Debayan Datta Ayan Maity

August 27, 2024

Abstract

The VOXLING project develops a Hindi Speech to English Text Translation System that converts Hindi audio to Hindi text and then translates it into English. This system aims to enhance accessibility for non-native Hindi speakers and facilitate communication for Hindi speakers in non-Hindi-speaking regions. By leveraging advanced speech recognition and machine translation technologies, VOXLING addresses real-time translation needs and advances research in low-resource languages.

Introduction

VOXLING is designed to bridge language barriers by converting Hindi audio into Hindi text and subsequently translating it into English. This system supports better communication for non-English speakers and aids in accessing content in Hindi. It also contributes to research and development in multilingual translation technologies.

Methodology

What ?

Model 1: Hindi Speech to Hindi Text

- Dataset Source:OpenSLR

Model 2: Hindi Text to English Text

- Dataset Source: IITB English-Hindi Corpus

Why?

- **Language Barrier Reduction**: Bridges the gap for non-English speakers and enhances real-time translation.
- **Accessibility**: Facilitates access to Hindi content for non-native speakers.
- **Common Ground for Communication**: Assists native Hindi speakers in non-Hindi-speaking regions.
- **Automated Translation System**: Useful in industries like Tourism, Service Centers, and Media.
- **Research & Development**: Promotes R&D in Hindi, a low-resource language.

How ?

- **Model 1:** Takes the Hindi Audio Input from Dataset 1, which contains ID, Hindi Speech, Hindi Text and English Text (collected manually by using Google Translate and considered as ground truth). The model processes the Hindi Speech to predict an intermediate Hindi Text output. [2, 3]
- **Model 2:** Uses Dataset 2, consisting of Hindi Text and English Text. It takes the predicted Hindi Text from Model 1 and translates it into English Text, producing the final Predicted English Text Output. [1]
- **Comparison:** The final output is compared with the actual English Text from Dataset 1 to evaluate the model's performance.

Work Plan & Timeline

Task	Start Date	End Date
Data Collection	20-08-2024	26-08-2024
Literature Survey	27-08-2024	06-09-2024
Data Preprocessing	01-09-2024	11-09-2024
Model 1 Training	12-09-2024	20-10-2024
Model 2 Training	24-09-2024	20-10-2024
Evaluation and Validation	20-10-2024	27-10-2024
Error Analysis and Optimization	27-10-2024	02-11-2024
System Integration	02-11-2024	10-11-2024
Final Testing and Validation	10-11-2024	19-11-2024

Table 1: Timeline of the Project

Work Division

We plan on working collaboratively and will divide tasks based on each other's strengths and weaknesses.

References

[1] Kavit Gangar, Hardik Ruparel, and Shreyas Lele. Hindi to english: Transformer-based neural machine translation. In *International Conference on Communication, Computing and Electronics Systems: Proceedings of ICCCES 2020*, pages 337–347. Springer, 2021.

[2] Ye Jia, Michelle Tadmor Ramanovich, Quan Wang, and Heiga Zen. Cvss corpus and massively multilingual speech-to-speech translation. *arXiv preprint arXiv:2201.03713*, 2022.

[3] Changhan Wang, Anne Wu, and Juan Pino. Covost 2 and massively multilingual speech-to-text translation. *arXiv preprint arXiv:2007.10310*, 2020.