

## A Novel Approach to Malayalam Speech-to-Text and Text-to-English Translation

Ms. ANCY K SUNNY, DENI THOMAS, JASHLIN S SIMON, MOHAMMED ZAIN RAFEEQUE, THAHA MUHAMMED YASEEN

### Abstract:

This paper presents a novel approach to facilitate Malayalam speech-to-text transcription and subsequent translation into English text. The proposed system leverages advancements in speech recognition, natural language processing, and machine translation techniques. We demonstrate the effectiveness of our approach through a practical implementation and evaluation.

### Introduction:

The ability to accurately transcribe spoken language and translate it into other languages has numerous applications in today's digital world. However, the development of such systems for languages with complex structures, such as Malayalam, presents unique challenges. In this paper, we propose a solution to address these challenges by combining state-of-the-art technologies in speech recognition and machine translation.

### Literature Review:

Previous studies have explored various approaches to speech-to-text transcription and machine translation. However, few have focused specifically on the Malayalam language. Existing systems often struggle with accurately transcribing and translating Malayalam due to its complex morphology and syntax.

### Methodology:

Our approach consists of several key steps:

*Speech Recognition* : We employ the SpeechRecognition library to transcribe spoken Malayalam into text.

*Text Preprocessing*: The transcribed text undergoes preprocessing, including tokenization and normalization, using the IndicNLP library.

*Translation*: The preprocessed text is translated into English using a custom-built translation model implemented with CTranslate2 and SentencePiece.

### Results:

We evaluated our system using a dataset of spoken Malayalam sentences. The system achieved a high accuracy in speech recognition and produced fluent translations into English.

### Discussion:

Our results demonstrate the feasibility and effectiveness of our approach in accurately transcribing and translating spoken Malayalam. However, certain challenges remain, such as handling dialectal variations and improving translation quality for complex sentences.

**Conclusion:**

In conclusion, we have presented a novel approach to Malayalam speech-to-text transcription and text-to-English translation. Our system shows promising results and opens up possibilities for further research and development in this area.

**References:**

- [1] S. K. Sheshadri, B. S. Bharath, A. H. N. S. C. Sarvani, P. R. V. B. Reddy, and D. Gupta, "Unsupervised neural machine translation for english to kannada using pre-trained language model," pp. 1–5, 2022.
- [2] A. H. Patil, S. S. Patil, S. M. Patil, and T. P. Nagarhalli, "Real time machine translation system between indian languages," pp. 1778–1783, 2022.