Robust Speech Recognition System for Malayalam

KURIAN BENOY

2021MCS120014

7TH APRIL, 2024

Indian Institute of Information Technology Kottayam

CONTENTS

Indian Institute of Information Technology Kottayam

Motivation

1. In Malayalam language, at the moment there are not any automatic speech recognition models which support long-form audio speech transcription, addressing the specific requirements for transcribing extended spoken content with timestamps. This is an essential component in creating subtitles for academic lectures, interviews, movies, serials etc.
2. Even though there has been a lot of works in Malayalam Speech to text. They aren't open-source most of the time. This means leveraging open-source methodologies, the system intends to provide access to datasets, model architectures, and algorithms, promoting transparency, reproducibility, and collaboration in the development of Malayalam ASR technology.
3. Lot of works claim to have achieved 90 percentage accuracy in datasets, even in datasets which are not available in the public domain and kept proprietary. Yet an apple to apple comparison will only ensure that whether model A or model B is better for Malayalam speech.

Indian Institute of Information Technology Kottayam

Problem Objectives

Indian Institute of Information Technology Kottayam

1. Problem Objectives

Develop an Open-Source ASR System:

The project aims to design and implement an open-source ASR system for Malayalam that overcomes the limitations of existing speech-to-text techniques. By leveraging open-source methodologies, the system intends to provide access to datasets, model architectures, and algorithms, promoting transparency, reproducibility, and collaboration in the development of Malayalam ASR technology. It should achieve a key goal of the project is to achieve a Word Error Rate (WER) of less than 0.15 in the developed ASR system for speech to text model accuracy.

Indian Institute of Information Technology Kottayam

2&3. Problem Objectives

Support Long-Form Audio Speech Transcription:

In addressing the dearth of specialized provisions for transcribing long-form audio with timestamps in Malayalam, the project endeavors to develop features and capabilities that cater to the specific requirements of transcribing extended spoken content.

Benchmark Various ASR Models:

The project seeks to compare and benchmark multiple ASR models to evaluate their performance in the context of Malayalam speech-to-text processing. By conducting systematic comparisons, the project aims to identify the strengths and limitations of different ASR methodologies, leading to insights that can inform the selection of appropriate models for specific use cases.

Indian Institute of Information Technology Kottayam

Updated Literature Review

Indian Institute of Information Technology Kottayam

Category	Model/Paper	Key Finding
ASR Models in Malayalam	Cini et al. [7]	Malayalam Numerical ASR is viable with HMM. Word-accuracy of 91%.
	Anuj et al. [8]	Used combinations of HMM and ANN. Word-accuracy of 86.67%.
	Kavya et al. [9]	Hybrid ASR model for open vocabulary speech recognition. Improved WER by 10-7%.
	Vineel et al. [10]	CTC model with a WER of 39.7 in Fleurs dataset.
	Alec et al. [2]	Whisper is an encoder-decoder based model with a WER of 103.2 in Common Voice 9 dataset.
	Barrault et al. [12]	A massively multilingual and multi-modal machine translation as a single model that supports tasks like speech-to-speech translation, speech-to-text translation, text-to-speech translation, text-to-text translation, and automatic speech recognition for up to 100 languages. It works in Malayalam but WER is not reported in paper.
ASR Models in Other Indian Languages	CLSRIL-23 [13]	Multilingual pretraining improves speech representations, Decreases WER by 5% and CER by 9.5% in Hindi.
	End-to-End Tamil ASR [14]	Cost-effective approach to build an ASR using Deep Speech for Tamil.
	Javed et al. [15]	Curated 17,000 hours of data for 40 Indian languages, achieved state-of-the-art results in low-resource languages.
Benchmarking ASR in English	SUPERB [16]	Framework for benchmarking speech processing models.
	ESB [17]	Evaluates performance of ASR systems across a broad spectrum of speech datasets.

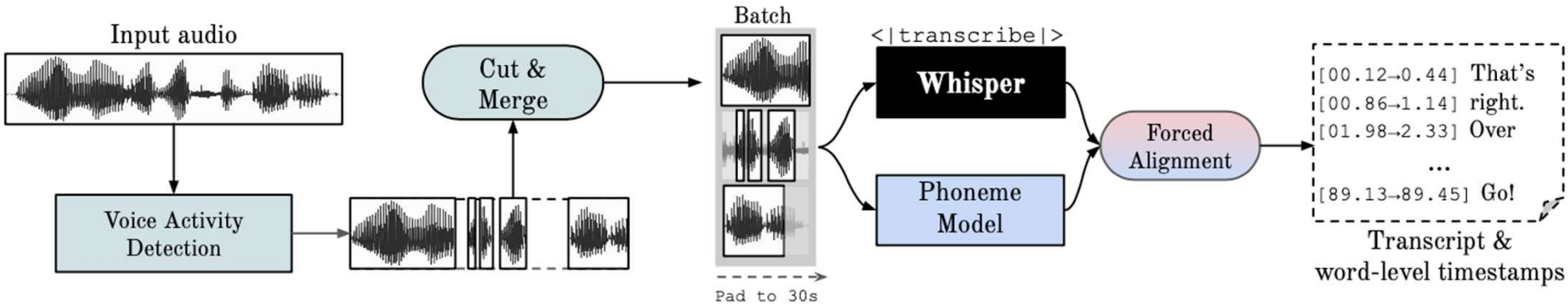
Table 2.1: Summary of Literature Survey

Indian Institute of Information Technology Kottayam

Support Long-Form Audio Speech Transcription

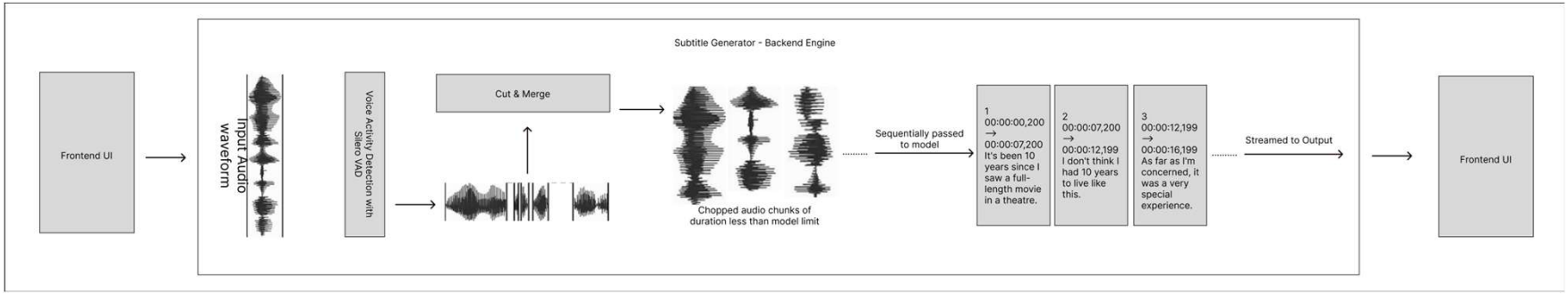
Indian Institute of Information Technology Kottayam

WhisperX architecture



Indian Institute of Information Technology Kottayam

Long Form Transcription



Indian Institute of Information Technology Kottayam

Indic-Subtitler

- It's an open source subtitling platform 💻 for transcribing and translating videos/audios in Indic languages.
- We are building this for an Opensource AI hackathon sponsored by Meta, which we were shortlisted for.
- Support for transcribing and translating in 10+ Indic languages including Malayalam with SeamlessM4T[2], WhisperX[6] and faster-whisper[5].

- Let me demo it: https://indicsubtitler.vercel.app/

Indian Institute of Information Technology Kottayam

Features

- Supports 4 powerful SOTA models for Speech Transcription
- Ability to download subtitles in SRT, JSON, TXT and VTT formats.
- Built-in Editor for manual tweaking
- Streaming Response Generation
- Stores past transcriptions in a library locally
- Easy to add custom models and self host
- Supports both transcription and translation of audio & Video files
- Extremely simple and easy to use UI/UX
- Live Transcribe feature

Indian Institute of Information Technology Kottayam

Humblebrags

- **100% Accessbility Score** as per Google Lighthouse
- Huge reach of over **750+ visitors to date.**
- Got **50+ Github stars**
- Testimonials from **10+ industry leaders** endorse our platform's effectiveness.

Indian Institute of Information Technology Kottayam

User Flow

Upload Video/Audio

Generate Subtitles

Make optional edits

Download .srt/.vtt file

Indian Institute of Information Technology Kottayam

Tech-Stack

Models Used:

- SeamlessM4T
- WhisperX
- faster-whisper
- Vegam-whisper

Frontend

- NextJS
- TailwindCSS
- DaisyUI

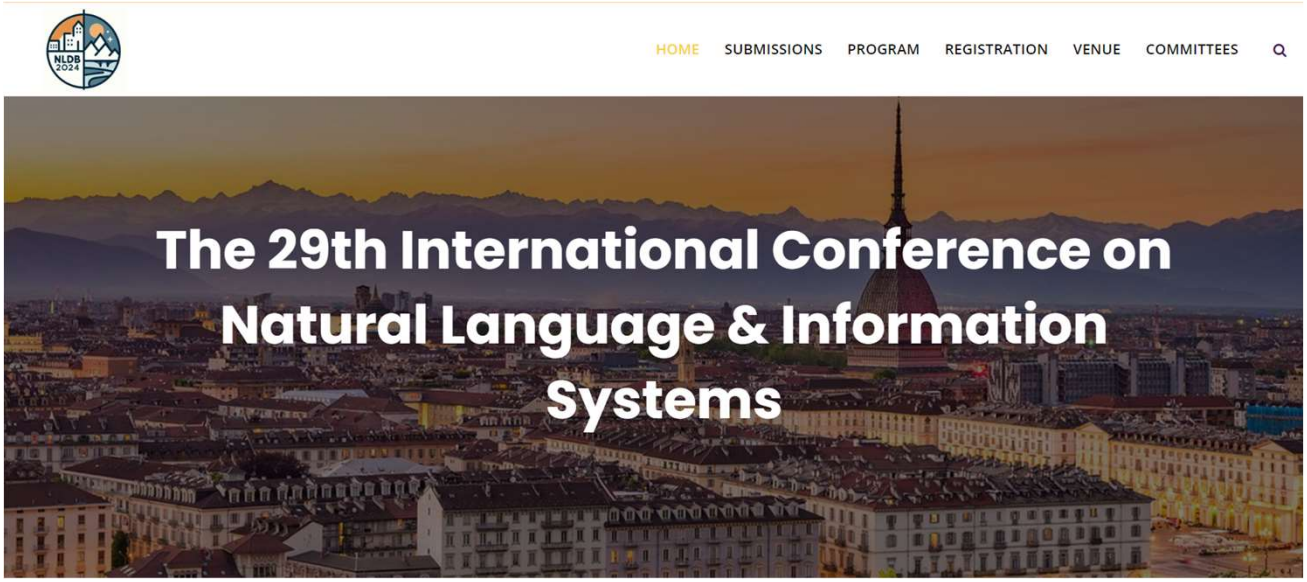
Backend

- FastAPI
- Modal.com

Indian Institute of Information Technology Kottayam

Paper submitted for NLDB 2024

- We submitted a paper for NLDB 2024 titled : An Open source platform for generating subtitles for Indian Languages



Indian Institute of Information Technology Kottayam

REFERENCES

1. Radford, Alec, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. "Robust speech recognition via large-scale weak supervision." In *International Conference on Machine Learning*, pp. 28492-28518. PMLR, 2023.
2. Barrault, L., Chung, Y. A., Meglioli, M. C., et al. "SeamlessM4T-Massively Multilingual & Multimodal Machine Translation." In: AI Meta Publications, 2023 [2] https://ai.meta.com/research/publications/seamlessm4t-massively-multilingual-multimodal-machine-translation/
3. Pratap, Vineel, Andros Tjandra, Bowen Shi, Paden Tomasello, Arun Babu, Sayani Kundu, Ali Elkahky et al. "Scaling speech technology to 1,000+ languages." *In* AI Meta *publication* (2023).
4. Manohar Kavya et al., ASR for Malayalam, In: https://gitlab.com/kavyamanohar/asr-malayalam
5. Klein Gullimane et al., faster-whisper, In: https://github.com/SYSTRAN/faster-whisper
6. Koluguri, Nithin Rao, et al. "Investigating End-to-End ASR Architectures for Long Form Audio Transcription." In. *Nvidia nemo website*(2023). https://nvidia.github.io/NeMo/blogs/2024/2024-01-parakeet/

Indian Institute of Information Technology Kottayam

REFERENCES

7. Bain, Max, Jaesung Huh, Tengda Han, and Andrew Zisserman. "WhisperX: Time-accurate speech transcription of long-form audio." *In: Interspeech conference* (2023).

8. Gopinath, Deepa P., and Vrinda V. Nair. "IMaSC--ICFOSS Malayalam Speech Corpus." *arXiv preprint arXiv:2211.12796* (2022).

9. Benoy Kurian et al., In: https://github.com/kurianbenoy/whisper_normalizer

10. Dinesh S Akshay, Thottingal Santhosh et al., In: https://github.com/libindic/normalizer

11. Kunchukuttan Anoop et al., In: https://github.com/anoopkunchukuttan/indic_nlp_library

Indian Institute of Information Technology Kottayam