

Evaluating the Efficacy of Large Acoustic Model for Documenting Non-Orthographic Tribal Languages in India

**Tonmoy Rajkhowa, Amartya Roy Chowdhury, Hrishikesh Ravindra
Karande, S.R. Mahadeva Prasanna**

Indian Institute of Technology, Dharwad
Karnataka, India

CONTENTS

- **Introduction**
- **Proposed plan of work**
- **Objectives**
- **Work Directions**
 - *Corpora Creation*
 - *Development of Spoken Language Technologies for documentation*
 - *Performance improvement using Large Acoustic Model*
- **Results**
- **Discussions**
- **Conclusion & Summary**
- **Future works**
- **Ethical Statements**
- **Acknowledgements**

INTRODUCTION

- Majority of the world's languages lack standardized writing system (un-orthographic).
- Under-representation in the digital domain.
 - *Endangered and may face extinction in the near future.*
 - *Loss of cultural heritages, knowledge and traditions passed orally with time.*
- **Viable Solution:** *Preservation by recording the speeches and storing them.*
- **Limitations:**
 - *Exposure only to a limited audience who are well-versed in that language.*
 - *Limited domain coverage.*
- **Primary Challenges:**
 - *To expose these languages to the wider audience.*
 - *Documentation in written form for these unwritten language in digital domain.*

PROPOSED PLAN OF WORK

- **Leveraging Spoken Language Technologies:**
 - *Automatic Speech Recognition (ASR)*
 - *Direct Speech-to-Text Translation (DS2TT)*
- **Technical Challenges:**
 - *Data hungry nature of these Deep Learning based technologies.*
 - *Low-resourced languages (Training data < 10 hours of audio duration).*
- **Addressing the Data Limitation:**
 - *Application of Large Acoustic Model.*
 - *Whisper Large V2 pre-trained model on 680,000 hours of multilingual audios.*

OBJECTIVES

- Introduction of four un-orthographic under-resourced tribal languages of India:
 - ***Soliga & Lambani*** spoken in Karnataka in Southern India.
 - ***Kui & Mundari*** spoken in Odisha in Eastern India.
- Method of documentation in digital domain using ***Spoken Language Technologies***.
- Role of ***Large Acoustic Model*** for improving documentation quality.

WORK DIRECTIONS

- **STAGE 1: Corpora creation**
 - Selection of un-orthographic tribal languages:
 - **Soliga & Lambani** spoken in Karnataka in Southern India.
 - **Kui & Mundari** spoken in Odisha in Eastern India.
 - Adoption of appropriate script for these tribal languages.
 - Manual quality check by linguists.
- **STAGE 2: Development of Deep Learning based Spoken Language Technologies**
 - ASR for generating transcriptions.
 - DS2TT for generating translations.
- **STAGE 3: ASR and DS2TT performance improvement using Whisper Large v2 pre-trained acoustic model.**

CORPORA CREATION

Contents of a Speech Corpus

1. Spoken utterances (audio)
2. Transcription / Transliteration (text)
3. Translation (text)

Adoption of appropriate script for transliteration of tribal languages

- *Popular and widely spoken languages having a standardized written script.*
- *Having a large digital presence.*
- *Geographical neighbourhood.*
 - ***Kannada script** for Soliga & Lambani*
 - ***Odia script** for Kui & Mundari*

DATA COMPILATION & PRE-PROCESSING

1) Compilation of text sentences in English

- Based on day-to-day conversations.
- Topics related to the tribal communities.
 - *Examples: Agricultural practices, herbal medicines, tribal cuisines.*
- Content from educational textbooks with reference to tribal communities.

2) Data cleaning and Pre-processing

- *Removal of incomplete & meaningless sentences.*
- *Removal of sensitive and political contents.*
- *Manual verification by linguists to ensure quality content.*

TRANSLATION TO CONTACT LANGUAGES

3) Translations to Kannada & Odia texts from compiled English sentences

- Compiled English sentences subsequently translated manually to Kannada & Odia.
 - **Kannada** as contact language for *Soliga & Lambani*.
 - **Odia** as contact language for *Kui & Mundari*.
- Translation by experts well-versed in English and contact languages.
- Translation quality check by linguists.
 - To ensure if sentences in contact languages are parallel
 - To ensure if translated sentences convey the same meaning reference to English.
 - Secondary level checking for sensitive statements, irrelevant topics.
 - Manual checking and correction of grammatical and spelling mistakes.

TRANSLATION TO TRIBAL LANGUAGES

4) Translations to Tribal Languages from Contact Languages

- Kannada & Odia sentences subsequently translated to Tribal Languages.
 - *Kannada sentences* translated to *Soliga & Lambani text* using *Kannada script*.
 - *Odia sentences* translated *Kui & Mundari text* using *Odia script*.
- Translation by experts and linguists well-versed in tribal languages and script of Contact Languages.
- Translation quality check by linguists and experts from respective tribal communities.
 - Ensure conveyance of similar meaning with reference to Contact Languages.
 - Evaluation of translated sentences by multiple experts for each tribal language.
 - Ensure colloquial rather than formal bookish language.
 - Sentence reframing (if required).

AUDIO RECORDING OF TRIBAL SENTENCES

5) Audio Recordings of Tribal Languages by native speakers

- Audio Recording Setup
 - Developed a GUI where each sentence is displayed for recording.
 - Utilized 16-bit 16 kHz sampling frequency and mono-channel recordings.
 - Recorded using microphones, ZOOM Recorder.
 - Recorded in studio as well as outdoor environments.
 - To ensure a robust dataset capable of handling noises.
- Audio Quality Check
 - Manual quality check by linguistic experts and tribal representatives.
 - Checking of correct pronunciations, expressiveness etc.
 - Speech intelligibility when recorded in outdoor environments.

STATISTICS OF THE CORPORA

Table 1: *Statistics of the corpora depicting the number of parallel text sentences along with the recorded total audio duration information (in hours) for each of the tribal languages.*

SI No.	Language Type	Languages	# of Sentences	Audio Duration
1	Reference	English	10,000	-
2	Contact	Kannada	10,000	-
3		Odia	10,000	-
4	Tribal	Soliga	10,000	9.7 hours
5		Lambani	10,000	10.1 hours
6		Kui	10,000	9.5 hours
7		Mundari	10,000	10.3 hours

GLIMPSES OF THE CORPORA

- Parallel corpora for all 4 language pairs.

Table 2: *Samples of Lambani and Soliga examples transliterated using Kannada alphabets along with their translated meaning in English.*

SI No.	English	Lambani	Soliga
1	Buffalo playing in the mud	ಅಂಗೋಳೋ ಆರೋಚ್	ಎಮ್ಮೆ ಬದೀಲಿ ಆಟ ಆಡಿದ್ದ
2	The bull hit the cart	ಬಳದ ಆನ ಗಾಡಿನ ಮಾರ್ದಿಮೋಚೆ	ಎತ್ತು ಬಂದು ಗಾಡಿಕೆ ಗುದ್ದಿಕಿತು

Table 3: *Samples of Kui and Mundari examples transliterated using Odia alphabets along with their translated meaning in English.*

SI No.	English	Kui	Mundari
1	Buffalo playing in the mud	ରଣ୍ଡେ କୋରୁ ଗେଦେତାନି କାହାଲମାନେ	ಎମೆଆଁ କେଡ଼ା ଲସକ୍ ରେ ଏନେକତାନେ
2	The bull hit the cart	ରଣ୍ଡେ ଷଣ୍ଡକୋଡ଼ି ଶଗଡ଼ି ଗାଡ଼ିତିନି ତୁଝୁଦାଅତେ	ମିଆଁ ସିଂଶ ଉରିକ୍‌ଗାଡ଼ିକେ ଧକ୍କାଲୋଃ

SUMMARY OF CORPORA CREATION PIPELINE

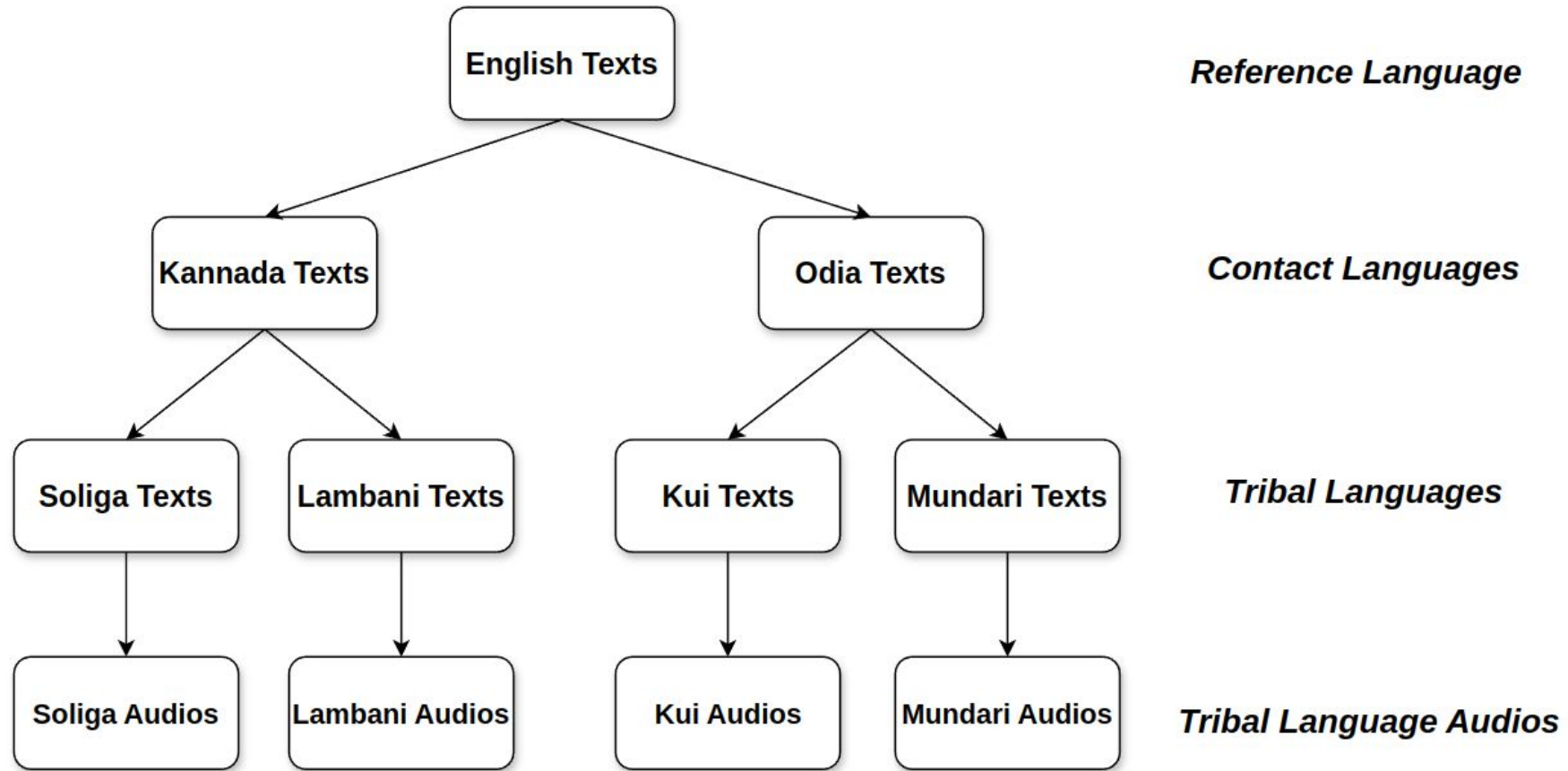


Figure 1: Block diagram depicting the pipeline of parallel corpora creation from text of English language to audios of tribal languages.

UTILITY OF THE CORPORA

- MACHINE TRANSLATION
- AUTOMATIC SPEECH RECOGNITION
- TEXT-TO-SPEECH SYNTHESIS
- SPEECH-TO-TEXT TRANSLATION
- SPEECH-TO-SPEECH TRANSLATION

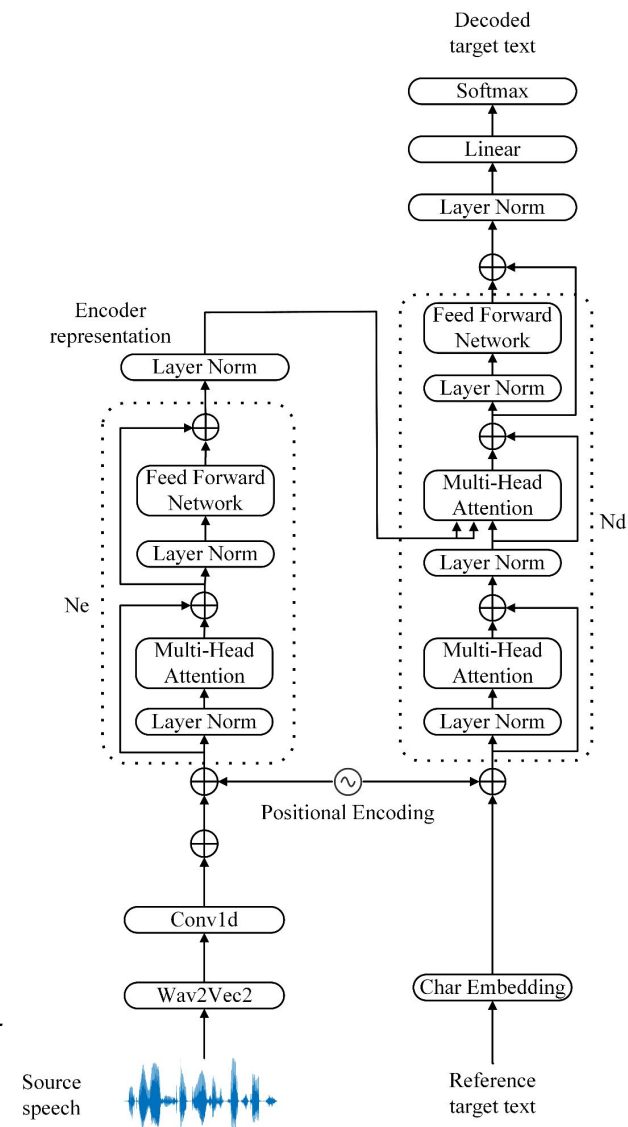
DEVELOPMENT OF SPOKEN LANGUAGE TECHNOLOGIES

- **AUTOMATIC SPEECH RECOGNITION:**
 - To generate transcriptions/transliterations.
- **DIRECT SPEECH-TO-TEXT TRANSLATION:**
 - To generate translations.

Implementation strategies

- *Training from scratch.*
- *Fine-tuning using Whisper large v2 pre-trained model.*

Figure 2: Architecture of an Encoder-Decoder based ASR / DS2TT system^[1] using Transformer units



[1] Wang et.al. , “Fairseq S2T: Fast Speech-to-Text Modeling with Fairseq”, ACL 2020

RESULTS: AUTOMATIC SPEECH RECOGNITION

Table 4: *Baseline and Whisper fine-tuned evaluations for ASR task on the four tribal languages using Word Error Rate (WER) metric (in percentages).*

SI No.	Language	WER (baseline)	WER (fine-tuned using Whisper)
1	Kui	106.53	64.57
2	Mundari	101.30	70.00
3	Lambani	100.37	49.61
4	Soliga	103.24	22.09

RESULTS: DIRECT SPEECH-TO-TEXT TRANSLATION

Table 5: *Baseline and Whisper fine-tuned evaluations using BLEU metric (in percentages) for DS2TT task with the tribal languages as source speech translated to English text.*

SI No.	Source-Target Language pairs	BLEU (baseline)	BLEU (fine-tuned)
1	Kui-English	2.44	5.58
2	Mundari-English	1.67	5.09
3	Lambani-English	1.47	7.62
4	Soliga-English	1.44	10.07

DISCUSSIONS

- *Performance improvement when fine-tuned with Whisper Large Acoustic pre-trained model.*
- *Better performance for languages that are linguistically closer to Kannada than Odia for both ASR and DS2TT task.*
 - *Additionally Whisper model supports Kannada and not Odia.*
- *Better performance for Soliga as it contains similar words with Kannada.*

CONCLUSIONS & SUMMARY

- *Created a parallel corpora containing four tribal un-orthographic languages.*
- *Proposed a novel method of documentation using Spoken Language Technologies.*
- *Evaluated the efficacy of Whisper acoustic model for enhancing documentation quality.*

FUTURE WORKS

- *Extend the size and audio duration of corpora.*
- *Invoke Generative AI based methods to augment data.*
- *Develop Large Language Models for these tribal languages.*
- *Perform a comparative study on the behaviour of various Large Acoustic Models such as Wav2Vec2, Conformer etc.*

ETHICAL STATEMENTS

- This work is funded by ***Ministry of Electronics & Information Technology, Govt. of India*** for the project “***Tribal Speech-to-Speech Translation***”.
- Linguists and other experts employed were paid appropriate honorariums.
- Consents were taken in written form from the tribal representatives for recording their voices.
 - Honorariums were paid as per the duration of recordings by each speaker.
- Written consents were also taken from head of those tribal communities to begin the project in their area.

ACKNOWLEDGEMENT

- Ministry of Electronics & Information Technology (MeiTY), Government of India
- Consultants and Data Associates
- Linguistic experts
- Fellow mentors and researchers
- Representatives from each of the tribal communities
- Indian Institute of Technology for Computing Facilities

THANK YOU

