NLTM-Speech Technologies in Indian Languages

Annexure XII: Speech Quality Control

KLEF Vaddeswaram -522302, AP, India

(PI: Suryakanth V Gangashetty)

(Co-PIs: Govind D, Akhilesh Kumar Dubey, and Samudravijaya K)

Readme.docx is available in the github at the following link

https://github.com/suryakanthvg/KLEFV_NLTM_STiILs_SQC/tree/main/Help_Directory

GUIDELINES FOR SPEECH CORPUS CREATION (Recording, Transcription, Metadata):

A document containing guidelines for speech corpus creation was created and circulated among all speech data vendors. This document contains detailed guidelines to create metadata, speech recording and transcription. The pdf version of this document namely "SpeechCorpusGuidelines_V11.pdf" is available in the github at the following link

https://github.com/suryakanthvg/KLEFV_NLTM_STiILs_SQC/tree/main/SpeechDataCollection_Transcription_Guidelines

Speech Data Delivery Status

SQC Team KLEFV has consolidated (**from April 2024 to March 2025)** the Speech data and related information received from various vendors. A total of 1547:31:03 (HH:MM:SS) of speech data (along with transcription and Metadata files) from 10 Indic languages have been received from vendors as per the prescribed guidelines.

The xlsx version of the "Speech Data Delivery Status" namely "Mediscribe_Shaip_Speech_Data_Delivery_status_From_June2024_Upto_R013.xlsx" receievd from the vendor "Mediscribe" is availabe in the github at the following link

https://github.com/suryakanthvg/KLEFV_NLTM_STiILs_SQC/tree/main/Speech_Data_Delivery_Status

Speech Corpora

A typical Assmese conversational wideband speech data receved during R013 is available in the github at the following link

https://github.com/suryakanthvg/KLEFV_NLTM_STiILs_SQC/tree/main/SpeechCorpora/as_SHA1P_C_Wideband_r013/

It has the following three folders

AUDIO: This has speech wav files in wav format

TRANSCRIPTION: This has the transcription files in json format

METADATA: Tthis has xlsx file.

Speech Quality Checks

The task of speech quality control of spoken language resources being periodically supplied by multiple vendors and providing appropriate feedback (at multiple levels: metadata, speech wav files and transcription) to vendors to improve the quality data.

Source Codes (Scripts) for Speech Quality chech are available in the github at the following link

https://github.com/suryakanthvg/KLEFV_NLTM_STiILs_SQC/tree/main/SourceCodeforSpeechQualityControl

This has the following two scripts

Speechwavfilestatistics_V2.py

Transcriptionjsonfilestatistics_V3.py

For all received data, quality checks have been carried out at various levels.

(a) At the wav file level (Signal clipping, low amplitude, Channel type, signal bandwidth). Any deviation from the transcription guidelines with respect to these

aspects were detected, and feedback was provided to the vendors. It was ensured that the corrected version was delivered.

The wav file level quality check is donr using the “Speechwavfilestatistics_V2.py” as follws:

Usage:

python3 Speechwavfilestatistics_V2.py fullpath to AUDIO folder

Example

python3 Speechwavfilestatistics_V2.py /home/svg/NLTM_SQC_Git_Upload/SpeechCorpora/as_SHA1P_C_Wideband_r013/AUDIO

After execution of the above script the wav file level quality check report file will be genearted in current folder with the follwoing file name

as_SHA1P_C_r013_SpeechwavfileSignalLevelQualityDetails_file.txt

(b) Transcription file level: Each transcription file was inspected for adherence to the transcription guidelines regarding JSON file format, identification of speech and nonspeech segments, active participation of each of the speakers in a conversation.

The Transcription file level quality check is donr using the “Transcriptionjsonfilestatistics_V3.py” as follws:

Usage:

python3 Transcriptionjsonfilestatistics_V3.py fullpath to TRANSCRIPTION folder/

Example

python3 Transcriptionjsonfilestatistics_V3.py /home/svg/NLTM_SQC_Git_Upload/SpeechCorpora/as_SHA1P_C_Wideband_r013/TRANSCRIPTION/

After execution of the above script for each of the json file the Transcription file level quality check results are displayed on the screen.

The total duration of segments (speech and nonspeech) was cross checked with the duration specified at three locations (namely, transcription file header, audio file header, and Metadata Spreadsheet). The minimum and maximum durations of speech segments by a speaker was monitored and extremities, if any, were identified and feedback was given to vendors. The statistics of code switched words were also collected.

The quality check of all the transcription files of the first few releases of speech by each of the vendors have been manually carried out by native speakers of the languages. Similar quality check is being carried out on a random (10%) basis for the later releases. The quality seems to be satisfactory.

(c) Metadata file level: For each release of speech data by each of the vendors, the received metadata spreadsheet was analysed and the parameters such as diversity in dialect, age group, gender, were monitored and appropriate feedback were given. The consistency in information such as duration of audio in the metadata spreadsheet and the wave file header were checked, and discrepancies, if any, were brought to the notice of the vendors.
