# Automatic Speech Recognition System Of Low Resource Malvi Language

**Capstone project report**
Submitted in partial fulfillment for the award of the degree of
**Bachelor of Technology**

**In**

**DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING (AI & CYBERNETICS)**

Submitted to

## VIT BHOPAL UNIVERSITY (M.P.)

**Submitted by**


## LAXMI PARMER (21BAC10033)
## RAKESH LODHI (21BAC10034)
## MANISH MEENA (21BAC10038)


Under the Supervision of
**Dr.ANIRBAN BHOWMICK**


**SCHOOL OF ELECTRICAL AND ELECTRONICS ENGINEERING**

**VIT BHOPAL UNICERSITY**

**BHOPAL (M.P.)-466114**


**DEC – 2024**

# VIT BHOPAL UNIVERSITY BHOPAL (M.P.) 466114

SCHOOL OF ELECTRICAL & ELECTRONICS ENGINEERING

**DECLARATION**

I hereby declare that the Dissertation entitled "*Automatic speech recognition system of low resource Malvi language*" is my own work conducted under the supervision of DR.ANIRBAN BHOWMICK, Assistant Professor Senior, SEEE at VIT University, Bhopal.

I further declare that to the best of my knowledge this report does not contain any part of work that has been submitted for the award of any degree either in this university or in other university / Deemed University without proper citation.

Laxmi Parmar (21BAC10033)

Rakesh Lodhi (21BAC10034)

Manish Meena (21BAC10038)

**VIT UNIVERSITY BHOPAL (M.P.) – 466114**

SCHOOL OF ELECTRICAL & ELECTRONICS ENGINEERING

# CERTIFICATE

This is to certify that the work embodied in this Capstone report entitled **"Automatic Speech Recognition System Of Low Resource Malvi Language"** has been satisfactorily completed by **Ms. LAXMI PRAMAR** (21BAC10033), **Mr. RAKESH LODHI** (21BAC10034) and **Mr. MANISH MEENA** (21BAC10038) in the School of Electrical & Electronics Engineering at VIT University, Bhopal. This work is a bonafide piece of work, carried out under my/our guidance in the School of Electrical & Electronics Engineering for the partial fulfilment of the degree of Bachelor of Technology.

**PROJECT GUIDE**

Dr.Anirban Bhowmick,
Assistant Professor Senior ,
School Of electrical and electronic engineering ,
VIT BHOPAL UNIVERSITY

Forwareded by

Dr.Soumitra K. Nayak,
Program Chair,
School Of electrical and electronic engineering ,
VIT BHOPAL UNIVERSITY

Approved by

Dr M Suresh,
Dean,
School Of electrical and electronic engineering ,
VIT BHOPAL UNIVERSITY

# Acknowledgement

In the first place I would like to record my gratitude to "***Dr. Anirban Bhowmick*****"** for his supervision, advice, and guidance from the very early stage of this thesis work as well as giving me extraordinary experiences throughout the work. Above all and the most needed, he provided me unflinching encouragement and support in various ways. His words have always inspired me to work in an efficient and comprehensive way. I would like to thank him for his constant encouragement that enabled me to grow as a person. His presence has definitely improved me as a human being.

I express my gratitude towards **"Dr M Suresh"**, Associate Professor & Dean, School of Electrical and Electronics Engineering Department, VIT Bhopal University, for providing me all the help and permitted me to work in the laboratory with no time limits.

I shall ever remain indebted to **"Dr. Senthil Kumar Arumugam",** Vice Chancellor, VIT Bhopal University, for providing me institutional and administrative facilities during my project work at VIT.

I would like to express my sincere gratitude to **Ph.D. Scholar Mr. Palash Jain** for his invaluable guidance and continuous support throughout this project. His expertise, encouragement, and willingness to assist whenever needed have been crucial to the successful completion of this work

I am very thankful to all

Laxmi Parmar 21BAC10033
Rakesh Lodhi 21BAC10034
Manish Meena 21BAC10038

# Abstract

The Malvi language, an Indo-Aryan dialect predominantly spoken in the western regions of Madhya Pradesh and southeastern Rajasthan, faces significant challenges due to its limited representation in the digital domain and lack of linguistic resources. As a low-resource language, Malvi has been largely underrepresented in speech technologies, making it difficult to develop tools that could preserve and promote its use in modern communication. This project aims to bridge this gap by developing an Automatic Speech Recognition (ASR) system for Malvi, focusing on the creation of a custom dataset through the collection and annotation of audio recordings from native Malvi speakers. These recordings were carefully selected to represent a variety of dialects, accents, and speech contexts, ensuring that the dataset encapsulates the linguistic diversity within the Malvi-speaking community.To address the challenges inherent in working with low-resource languages, advanced speech recognition techniques were employed to process and analyze the speech data. The ASR model was trained using this dataset, and its performance was evaluated through key metrics such as Word Error Rate (WER) and Character Error Rate (CER). Initial results showed promising accuracy levels, indicating that it is indeed possible to build a functional ASR system for Malvi, despite the scarcity of resources. These early findings demonstrate the potential for effective transcription and voice-enabled applications in Malvi, which could help preserve the language and provide a platform for its continued use in the digital age.This work makes a significant contribution not only to the technological advancement of Malvi but also to the broader field of low-resource language processing. By laying the foundation for Malvi's inclusion in modern AI-driven communication technologies, this project opens doors for future research aimed at enhancing the accessibility and inclusivity of such technologies. The creation of an ASR system for Malvi is a step toward fostering linguistic diversity in

the digital world, ensuring that this underrepresented language is not left behind in the rapid development of speech technologies. Moreover, the methods and techniques employed in this project can be adapted for similar low-resource languages, contributing to the preservation and revitalization of linguistic heritage globally.

Keywords: ASR, audacity, WER, CER, Wav2Vac2.0

# List of figures


| Figure no. | Caption /tittle | Page no. |
|---|---|---|
| 1.1.1.1 | Malwa Region | 2 |
| 1.1.1.2 | Malwa Region | 3 |
| 3.2.1.1 | Ground Work | 9 |
| 3.2.1.2 | Ground Work | 9 |
| 3.2.1.3 | Ground Work | 10 |
| 3.2.2.1 | Audacity | 11 |
| 3.3.1 | Flow Chart | 12 |
| 4.1 | Text Data | 15 |
| 4.2 | Audio Data | 16 |
| 5.1 | Wav2vec2.0 model | 17 |

# List of Tables


| Table no. | Caption /tittle | Page no. |
|---|---|---|
| 1.1.2.1 | Malwa Region Districts | 4 |
| 1.1.3.1 | Total Malwa Speakers | 5 |
| 1.1.3.2 | Rural Speakers | 5 |
| 1.1.3.3 | Urban Speakers | 5 |
| 4.1 | Age Wise Speakers | 14 |

# Table of contents

**Page No.**


Front page …….. ..I

Declaration……II

Certificate…… III

Acknowledgment…… IV

Abstract…….V

List of figures ……..…. VI

List of Tables……. VII

**Contents**

Chapter – 1 Introduction…….. 01-06

1.1 Malvi and Malwa region

1.1.1 Malwa Region…….

1.1.2 Malvi Language…...

1.1.3 Malvi Speakers……

1.2 Motivation of work…..…….....

1.3 Objective of the work……. 06

Chapter – 2 Literature review……..... 07

Chapter – 3 Problem formulation and proposed methodology….

3.1 Problem formulation…….. 08

3.2 Proposed methodology…… 09

3.2.1 Data Collection…… 09

3.2.2 Corpus Format and Storage…….. 10

3.3 Implementation……. 12

Chapter – 4 Results and Discussions…….. 14-16

Chapter – 5 Conclusion Future scope……. 17

References ……. 18

# Chapter 1

# 1. Introduction

Language plays a pivotal role in shaping human communication, culture, and identity. However, many regional and indigenous languages, particularly those in linguistically diverse nations like India, face challenges related to limited recognition and documentation. One such language is Malvi, a mother tongue grouped under the Hindi language family and classified as an Indo-Aryan language. Malvi is predominantly spoken in the Malwa region of India, encompassing districts such as Ujjain, Indore, Dhar, and Jhalawar. Despite its significant cultural relevance, Malvi remains a low-resource language, with minimal linguistic and technological resources to support its preservation and growth.Malvi is written in the Devanagari script, read from left to right, and is closely related to Nimadi and other dialects within the Rajasthani language group. Influenced by Rajasthani, with traces of Gujarati and Marathi, Malvi showcases linguistic diversity and cultural richness. While nearly 75% of the Malwa population can converse in Hindi, Malvi retains a distinct identity, with major dialects including Ujjaini, Rajawadi, Umadwadi, and Sondhwadi.This project seeks to develop an Automatic Speech Recognition (ASR) System for the Malvi Language, addressing the lack of digital tools and resources for this language. A key component of the project is the creation of a comprehensive corpus for Malvi. To ensure the dataset's authenticity and representativeness, the team will conduct extensive groundwork, engaging directly with native speakers in the Malwa region. This involves collecting audio samples and corresponding text transcriptions from various dialects, thereby creating a diverse and accurate dataset for the ASR model.The end-to-end Wav2Vec2.0 framework has been chosen for building the ASR system due to its ability to process raw speech and deliver high accuracy, even for low-resource languages. The ASR model aims to transcribe Malvi speech into text, enabling better communication and advancing the digital representation of the language.By focusing on both the technical development and the cultural preservation of Malvi, this project aspires to contribute to the broader goals of linguistic documentation, accessibility, and revitalization. The outcomes will pave the way for future research and technological advancements for Malvi and other low-resource languages, ensuring their continued relevance in the digital era.

## 1.1 Malvi and Malwa Region

### 1.1.1 Malwa Region

The area of the Malvi-speaking people encompasses a large tract of land in central India, specifically the western part of Madhya Pradesh (MP) and the lower southeastern side of Rajasthan. The Malwa region comprises territory from 21 to 25 degrees north latitude and from 73 to 80 degrees east longitude, or from Chittore in Mewar in the north to the Tapti River in the south, and from Bundelkhand in the east to Gujarat in the west. It consists largely of plateau, but it also has low ranges of hills and rivers running roughly west to southeast. Malwa is a traditional cultural area consisting today of the districts of Ujjain, Rathlam, Indore, Dewas, Shajapur, Mandsaur, Nimuch, Rajgarh, Sehore, Dhar, and Bhopal of MP and Jhalawar district of Rajasthan. The traditional boundaries of Malwa are attested to in a well-known poetic verse as 'between Chambel, Betwa and Narmada in the south.' In fact, the Narmada River in the south, the Betwa in the east and the Chambel in the northwest roughly marked its boundaries. The provinces of Kanthat and Bagat separated Malwa from Gujarat and Rajaputana, while the tract known as Hadothi formed the most extreme limit on the northwest. Bundelkhand and Gondwana surrounded Malwa on the eastern and southeastern side. The region within the province was mainly plateau, with alluvial tracts scattered all over the land. The jungle was dense in many places.

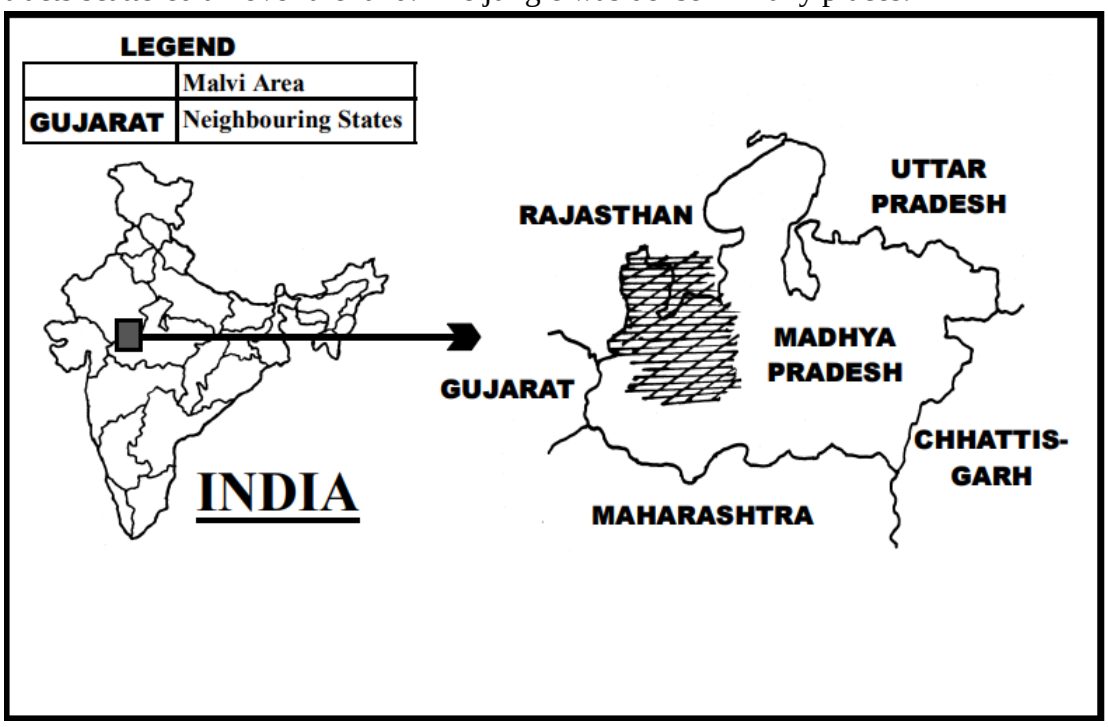



Figure 1.1.1.1 Malwa region

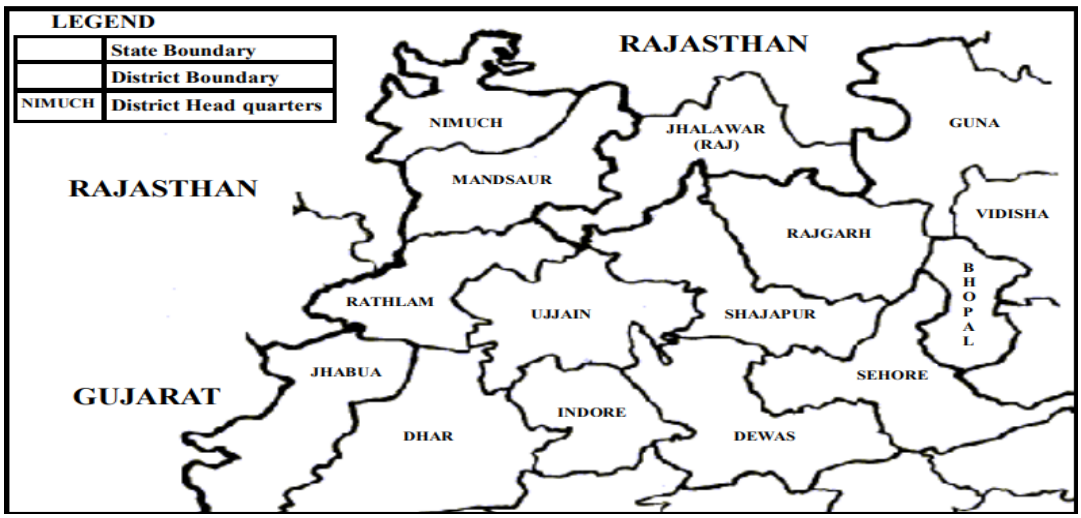
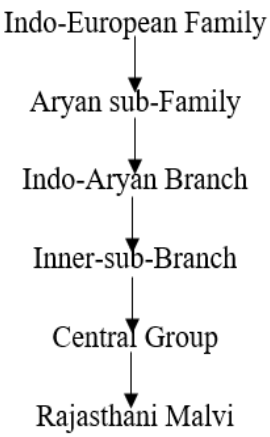



Figure 1.1.1.2 Malwa Region

### 1.1.2 Malvi Language

Malvi is a regional language spoken in the northwestern part of MP and in Jhalawar district of Rajasthan. It is functioning as a language for intra-group and inter-group communication. The language name is Malvi because it is spoken in the region of Malwa. It has alternative names such as Malwada, Mallow, and Ujjaini, and alternate spellings such as Malwi and Malavi. Malvi is a language of the Indo-European, Indo-Iranian, IndoAryan subfamily, Central group. It was previously considered that Malvi is a dialect of South Rajasthani. Grierson further suggested, 'Nimadi is really a form of Malvi, but it has such marked peculiarities of its own that it must be considered separately'. Malvi is not a dialect of South Rajasthani nor Eastern Hindi, but instead comes under Western Hindi and as a language has its own identity. The Malvi language is said to be a direct descendant of '**Avanthi Prakrith'3** of the mid-Indo-Aryan Family. Malvi is surrounded by different languages and dialects of Gujarati, Rajasthani, Hindi, and Marathi. Along its boundaries are: to the northwest, Mewari to the southwest, Gujarati to the north, the Jaipuri dialect of Rajasthani to the east, the Bundeli dialect of western Hindi to the south, Nimadi and the Khandeshi dialect of Marathi.



Within the Malvi language area, there are various 'islands' of Bhili and Gondi dialects of MP. Over the ages Malvi has fallen under the influence of the main neighbouring languages such as Hindi, Marathi, and Gujarati. For example, originally Malvi had a counting system of numbers up to twenty, and then for twenty they would say 'kozi'. After twenty, they would say 'one kozi', 'two kozi', etc. Perhaps because of the influence of Hindi, this Malvi numbering system is no longer much in use. Dialects Malvi has its own identity, literature, and dialects. The Ethnologue lists the following as dialects of Malvi: Bachadi, Bhoyari, Dholewari, Hoshangabad, Jamral, Katiyai, Malvi Proper, Patvi, Rangari, Rangri, and Sondhwari. However, regional scholars consider the four major dialects of Malvi to be Ujjaini (also called Avanthika or Malvi Proper), Rajwadi, Umadwadi, and Sondhwadi.

| District | Dialect |
|---|---|
| RAJGARH | Umadwadi |
| MANDSAUR | Sondhwadi |
| NEEMUCH | Rajwadi Malvi |
| RATLAM | Bhili Malvi ( influenced by local dialects)+Rajwadi |
| UJJAIN | Ujjaini (Malvi Proper) |
| AGAR | Sondhwadi Malvi |
| SHAJAPUR | Sondhwadi |
| DEWAS | Adarsh Malvi |
| SEHORE | Adarsh Malvi |
| DHAR | Adarsh Malvi |
| JHALAWAR | Sondhwadi |
| KOTA | Kota Malvi (variation influenced by local culture) |
| PRATAPGARH | Rajwadi Malvi (local variations) |
| BANSWARA | Bhili Malvi (with tribal influences) |

Table 1.1.2,1 Malwa Region Districts

**1) Ujjaini** is spoken in the area of Ujjain, Dewas, Indore, Sehore, northern Dhar, and southern Shajapur districts of MP. Ujjain was the capital of the Avanthi kingdom.

**2) Rajwadi** is found in Nimuch, Mandsaur, Rathlam, and some parts of Indore districts of MP. The people of this area prefer to call their language Malvi rather than Rajwadi. Since the Rajaputana kings were previously ruling the Mewad kingdom (including the region of Chittaurgarh and Banswara in Rajasthan and Mandsaur, Rathlam, and Nimuch in MP), the Malvi in this area was influenced by Rajasthani dialects and so it came to be called Rajwadi or Rangari.

**3) Umadwadi** is spoken in Rajgarh district of MP. The kings who formerly ruled in this area were Umads, a clan of Rajputs, and therefore this dialect of Malvi became known as Umadwadi.

**4) Sondhwadi** is spoken in Jhalawar district of Rajasthan and in northern Ujjain, north-western Shajapur, eastern Mandsaur, and western Rajgarh districts of MP. Since the region of Jhalawar district of Rajasthan and part of Mandsaur district of MP (between two Kalasindh rivers) was ruled by Sondhia kings (Sondhia being one of the Rajput

clans), and since Sondhias are in the majority in this area at present, this dialect is referred to as Sondhwadi.

### 1.1.3 Malvi Speakers

Malvi has been appearing in various Indian Censuses with considerable strength. As per the latest published data in Census of 2001, Malvi has been returned by 5565167 at all India level. The distribution of Malvi in different states of India including Rajasthan is as given below in the table:

| India/State | Total | | |
|---|---|---|---|
| | Person | Male | Female |
| India | 5565167 | 2851688 | 2713479 |
| Madhya Pradesh | 5175793 | 2652734 | 2523059 |
| Rajasthan | 385393 | 196876 | 188517 |
| Chhattisgarh | 2198 | 1146 | 1052 |
| Maharashtra | 950 | 486 | 464 |
| Gujarat | 675 | 348 | 327 |

Table 1.1.3.1 Total Malwa Speakers

| Rural | | |
|---|---|---|
| Person | Male | Female |
| 4979328 | 2549583 | 2429745 |
| 4614982 | 2363687 | 2251295 |
| 361533 | 184424 | 177109 |
| 1752 | 921 | 831 |
| 533 | 269 | 264 |
| 478 | 248 | 230 |

Table 1.1.3.2 Rural Speakers

| Urban | | |
|---|---|---|
| Person | Male | Female |
| 585839 | 302105 | 283734 |
| 560811 | 289047 | 271764 |
| 23860 | 12452 | 11408 |
| 446 | 225 | 221 |
| 417 | 217 | 200 |
| 197 | 100 | 91 |

Table 1.1.3.3 Urban Speakers

According to the 2011 Census of India, the Malvi language ranked 24th among the mother tongues spoken by over 1 million people in the country. With a total of 5,565,167 speakers, Malvi represents approximately 0.43% of India's total population. This ranking highlights the significant presence of Malvi, especially in the western regions of Madhya Pradesh and southeastern Rajasthan. Despite its large speaker base,

the language remains underrepresented in technological resources and linguistic research, which further emphasizes the need for initiatives like this project to develop automatic speech recognition systems and other language technologies for Malvi.

## 1.2 Motivatioin of work

The Malwa region, renowned for its rich cultural heritage and linguistic diversity, is home to the Malvi language, an Indo-Aryan dialect spoken by millions. As Malvi speakers ourselves, we are deeply connected to this language, which holds immense cultural and historical value. However, despite its widespread use, Malvi remains a low-resource language, lacking the technological tools and linguistic datasets necessary for its preservation and growth.Currently, there is no existing Automatic Speech Recognition (ASR) system or standardized corpus available for Malvi. This absence limits its representation in the digital world and hinders efforts to document, analyze, and integrate the language into modern technological applications. Motivated by this gap, our work aims to contribute a comprehensive ASR system and a curated corpus for Malvi. This effort will not only provide linguistic support for Malvi but also serve as a significant contribution to the open-source community, encouraging further research and innovation in regional languages. By bridging this gap, we aspire to bring recognition to Malvi and ensure its presence in the rapidly evolving digital landscape.

## 1.3 Objective of the work

### 1.2.1 Development of an Automatic Speech Recognition (ASR) System for Malvi

To create an ASR system for the Malvi language using advanced techniques like the Wav2Vec2.0 framework. This system will convert spoken Malvi language into text, enabling its use in various speech-to-text applications.

### 1.2.2 Corpus Creation and Data Collection

To curate a comprehensive Malvi corpus by collecting real-world speech data from native speakers in the Malwa region. This will involve extensive groundwork, engaging with speakers of different Malvi dialects (Ujjaini, Rajawadi, Umadwadi, Sondhwadi), ensuring a diverse representation of the language for training and evaluation.

### 1.2.3 Preservation and Recognition of Malvi

To contribute to the preservation and digital recognition of the Malvi language by making it accessible through the development of ASR technology and creating a publicly available dataset. This will promote its use in modern technology and ensure its survival in the digital era.

### 1.2.4 Contribution to the Open-Source Community

To provide open-source access to the Malvi ASR system and corpus, fostering further research, development, and innovation in the field of low-resource languages. This will enable other researchers and developers to build upon the foundation laid by this work, benefiting the linguistic and technological community.

# Chapter 2

## 2. Literature review

[1] Abhayjeet Singh, Arjun Singh Mehta, 2023 proposed "An ASR Corpus in Chhattisgarhi." In their paper, they discussed effective methods for creating an ASR corpus for Chhattisgarhi, a low-resource language. Findings of the research paper include effective corpus creation strategies for low-resource languages. However, the limitations mentioned in the paper are the high cost associated with data collection and corpus creation.

[2] Zhanibek Kozhirbayev, Aiden Williams, Andrea Demarco, 2023 proposed "Kazakh Speech Recognition: Wav2Vec2.0 vs. Whisper." In their paper, they compared the performance of Wav2Vec2.0 and Whisper models for Kazakh ASR. Key findings of the research include that Wav2Vec2.0 showed better results for low-resource languages. However, the limitations identified in the study are the challenges posed by unscripted speech in recognition accuracy.

[3] Claudia Borg, Prashant Upadhyaya, Omar Farooq, Musiur Raza Abidi, Yash Vardhan Varshney, 2023 proposed "The Applicability of Wav2Vec2 and Whisper for Low-Resource Maltese ASR." In their work, they focus on the effectiveness of both Wav2Vec2 and Whisper models for Maltese ASR in a low-resource setting. The findings highlight that both models are effective for low-resource languages like Maltese. The limitations pointed out in the research include the limited dataset size and complexities of low-resource language data.

[4] Prashant Upadhyaya, Omar Farooq, Musiur Raza Abidi, Yash Vardhan Varshney, 2017 proposed "Continuous Hindi Speech Recognition Using Kaldi ASR." In their paper, they describe the application of Kaldi for continuous Hindi speech recognition using deep neural networks. Findings of the research indicate the successful application of Kaldi for Hindi ASR. Limitations mentioned in the paper include being limited by the size and quality of the available Hindi speech dataset.

# Chapter 3

# 3.Problem formulation and proposed methodology

## 3.1 Problem Formulation

The primary challenge addressed in Phase 1 of this research is the absence of an Automatic Speech Recognition (ASR) model for the Malvi language, a low-resource language predominantly spoken in western Madhya Pradesh and southeastern Rajasthan. The Malvi language faces several obstacles in developing ASR systems, including:

**No ASR Model**: Currently, there is no ASR model for Malvi, which impedes automatic speech data processing and the creation of effective speech recognition systems.

**Low Resource**: Malvi is a low-resource language, characterized by limited digital presence and a scarcity of available annotated data, which hinders the development of an ASR model.

**Minimal Digital Presence**: Technological tools and online resources for Malvi are scarce, limiting the availability of language processing tools and data for researchers.

**No Language Models**: Due to the dialectical diversity of Malvi, no pre-existing language models exist, making it more challenging to generalize the language for ASR applications across different regions and speakers.

These challenges underscore the need to create a Malvi speech corpus as the foundation for developing future ASR systems. The corpus will help represent Malvi in the field of speech technology and enable the development of recognition systems that can accurately process the language.

## 3.2 Proposed Methodology (Phase 1: Corpus Creation)

In this phase, the main objective is to create a speech corpus for the Malvi language, which will later serve as a resource for training ASR models. The following methodology is being implemented:

### 3.2.1 Data Collection

**Speaker Diversity**: The corpus includes speech data from 34 native speakers of Malvi, covering a wide range of regions, accents, and dialects across the Malwa region. This diversity ensures the inclusion of linguistic variation.

**Recording Setup**: High-quality mono-channel microphones are used, with controlled environments to minimize background noise and ensure clarity.

**Speech Segments**: The corpus includes varied speech types, such as common phrases, conversational speech, and formal sentences, to represent different communication contexts.

**Demographics**: The 34 speakers consist of 16 females and 18 males. The age range of speakers is between 18 to 60 years. These speakers come from 10-15 distinct locations, including villages and districts within the Malwa region.

Figure 3.2.1.1 Ground Work

Figure 3.2.1.2 Ground Work

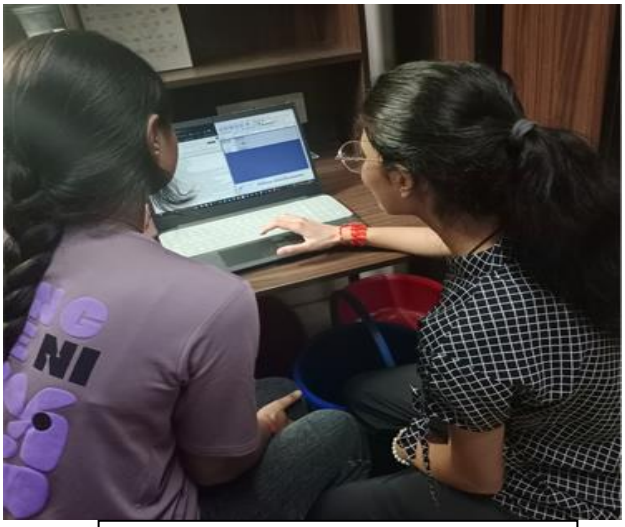


Figure 3.2.1.3 Ground Work

**Data Annotation :**Manual Transcription: Each audio sample is transcribed manually by native Malvi speakers to ensure transcription accuracy. The transcription includes text and phonetic representations where applicable.

**Linguistic Annotation**: In addition to transcription, the data is annotated with phonetic and semantic labels, allowing for a deeper linguistic analysis of the speech data.

**Data Augmentation**

**Artificial Expansion**: To overcome the challenge of limited data, various data augmentation techniques are applied:

**Noise Addition**: Background noise is introduced to simulate real-world conditions and enhance the robustness of the corpus.

**Pitch Shifting**: The pitch of the recordings is adjusted to represent different speakers and voice variations.

**Time Stretching**: The speed of the audio is altered without affecting its pitch, increasing the variability in the dataset.

## 3.2.2 Corpus Format and Storage

**Audio Format**: All audio files are stored in WAV format to ensure high-quality sound suitable for ASR training.

**Transcription Format:** The transcriptions are stored in a consistent text format for ease of use in training and model evaluation.

**Cloud Storage**: The corpus is stored on secure cloud-based platforms, enabling easy access and future scalability as the corpus expands.

**2.5 Corpus Size and Scope**

**Initial Corpus Size:** The initial corpus contains 8 hours of speech, with a word count of 19,566 words, including 4,274 unique words. It consists of 2,200 sentences, covering diverse speech patterns and accents.

**Future Expansion**: The corpus is planned to be expanded further by including more speech samples from diverse speakers, especially those from remote regions.

2.6 Tools and Technologies

**Recording and Chunking Software:** For audio recording and segmentation, Audacity software is used. This tool helps in clear recording and chunking of speech data into smaller segments.

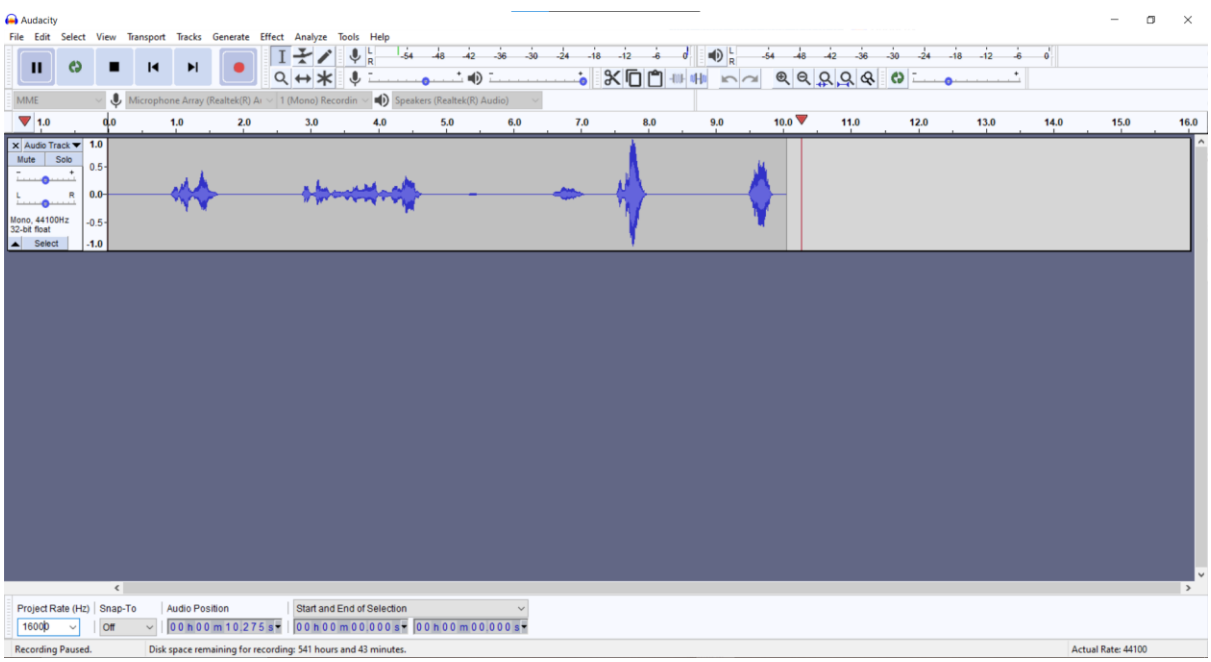



Figure 3.2.2.1 Audacity

## 3.3 Implementation

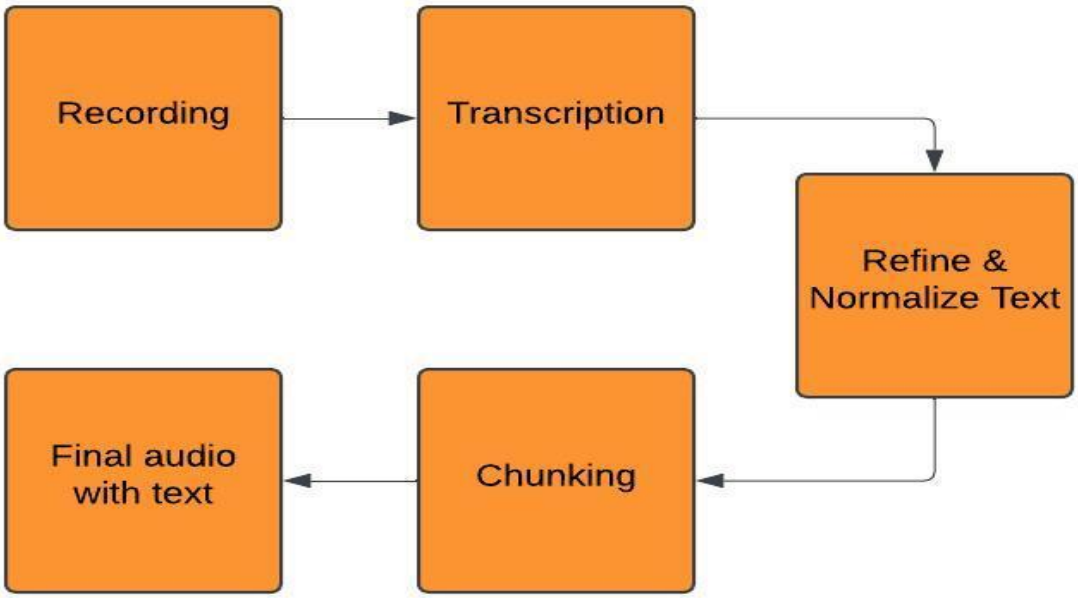



Figure 3.3.1 Flowchart

The flowchart illustrates the implementation process for creating the Malvi speech corpus, which serves as the foundation for developing an Automatic Speech Recognition (ASR) system for the language. Here's a step-by-step explanation of the process:

**Recording**: The first step involves capturing high-quality audio samples from native Malvi speakers. These recordings are collected in a controlled environment to minimize background noise. The speakers, chosen from diverse regions within the Malwa region, provide a variety of accents and speech patterns. The audio recordings, each lasting 5-10 minutes, are stored in WAV format for clarity and consistency.

**Transcription**: After recording, the audio files are transcribed into text. This step is critical for converting spoken Malvi into written form. The transcription is manually done by native speakers to ensure accuracy, capturing both the phonetic and semantic aspects of the language. The transcribed text is aligned with the corresponding audio for later use in training ASR models.

**Refine and Normalize Text:** Once the transcription is complete, the text undergoes refinement and normalization. This step involves correcting any spelling or grammatical inconsistencies and ensuring that the transcriptions are standardized. For instance, variations in regional dialects or informal language might be normalized to

ensure consistency across the corpus. This process helps maintain high-quality data that can be reliably used for model training.

Chunking: In this step, the text and audio are divided into smaller, manageable segments or "chunks." The speech recordings are split based on logical units, such as phrases or sentences. This chunking ensures that each segment of audio corresponds to a specific segment of text, which is necessary for effective training of the ASR model. The chunking also helps in aligning audio with transcription at a granular level, making the dataset more precise and useful.

**Final Audio and Text**: After chunking, the final audio and text data are prepared for use in ASR model training. The audio files, along with their corresponding transcriptions, are organized and stored in a consistent format. This final corpus is then ready to be used for training speech recognition models, providing the foundation for future advancements in ASR systems for the Malvi language.

# Chapter 4

# 4. Results and discussion

## Text Data Analysis

In this phase of the project, we performed an in-depth analysis of the refined text data collected for the Malvi speech corpus. The data collected from native speakers has been carefully processed and refined to ensure comprehensive coverage of the Malvi language, including its various dialects, tones, and accents. Below are the key statistics and insights from the collected and refined dataset:

| Age Group | Male | Female | Total |
|---|---|---|---|
| 18-30 | 8 | 6 | 14 |
| 30-45 | 6 | 6 | 12 |
| 45-60 | 4 | 4 | 8 |

Table 4.1 Age Wise Speakers

**4.1. Total Data Collected :**Total Audio Duration: The refined corpus consists of 8.54 hours of recorded speech, providing a comprehensive dataset for ASR model development.

**Speakers:** A total of 34 native speakers of Malvi contributed to the corpus. This included:
16 female speakers
18 male speakers
This balanced gender representation ensures that both male and female speech characteristics are captured, offering a more robust dataset for training the ASR model.

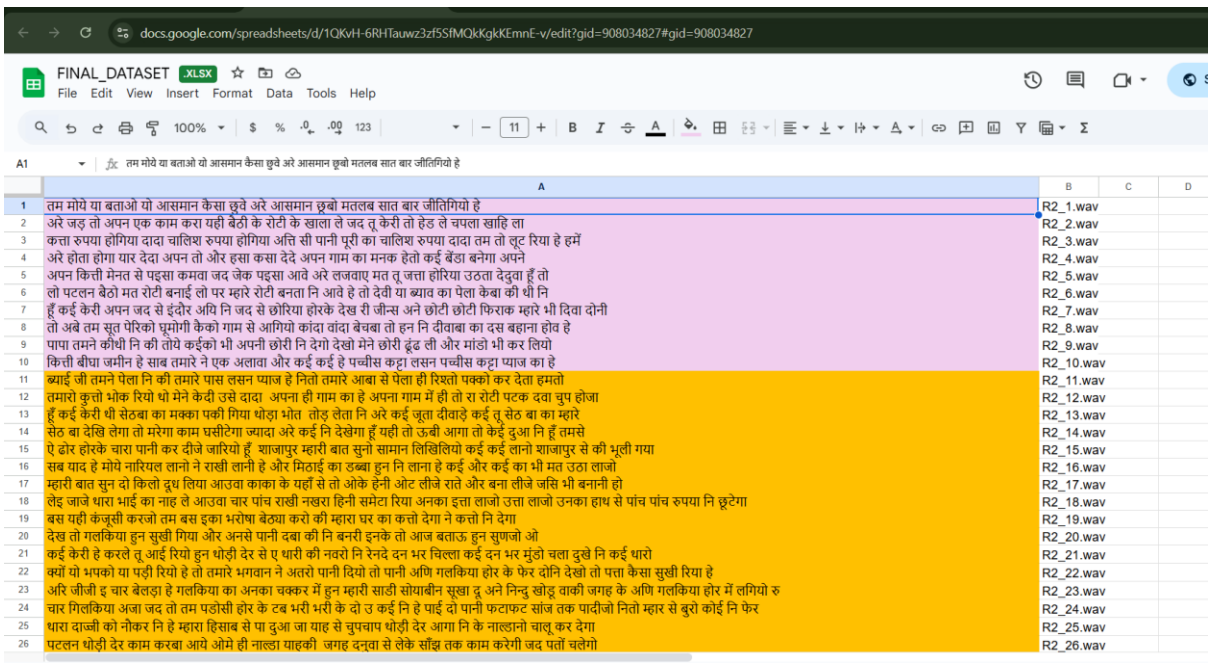
**4.2. Text Data Overview: Total Sentences**: The refined corpus includes 2200 sentences, which cover a wide array of linguistic features and speech contexts, including everyday conversations, formal sentences, and region-specific phrases.

**Total Words**: The corpus contains 19,566 words, which represent a diverse set of linguistic patterns, vocabulary, and commonly used phrases in the Malvi language.

**Unique Words**: There are 4,274 unique words in the corpus, reflecting the linguistic diversity and richness of the Malvi language. This large vocabulary set ensures that the ASR model will be trained to recognize a wide array of words and expressions.

**4.3. Chunking Analysis** Average Chunk Length: To facilitate efficient processing, the audio has been divided into smaller chunks. The average chunk length is 14 seconds, allowing for accurate recognition of speech while preserving context in each segment.

**4.4 Chunk Count**: Given the total audio duration of 8.54 hours, and an average chunk length of 14 seconds, the corpus is divided into approximately X chunks (calculated by dividing the total duration by the average chunk length).



FINAL_DATASET .XLSX

File Edit View Insert Format Data Tools Help

A1: तम मोये या बताओ यो आसमान कैसा छुवे अरे आसमान छूबो मतलब सात बार जीतिगियो हे

| | A | B |
|---|---|---|
| 1 | तम मोये या बताओ यो आसमान कैसा छुवे अरे आसमान छूबो मतलब सात बार जीतिगियो हे | R2_1.wav |
| 2 | अरे जड़ तो अपन एक काम करा यही बैठी के रोटी के खाला ले जद तू केरी तो हेड ले चपला खाहि ला | R2_2.wav |
| 3 | कत्ता रुपया होगिया दादा चालिश रुपया होगिया अति सी पानी पूरी का चालिश रुपया दादा तम तो लूट रिया हे हमें | R2_3.wav |
| 4 | अरे होता होगा यार देदा अपन तो और हसा कसा देदे अपन गाम का मनक हेतो कई बेंठा बनेगा अपने | R2_4.wav |
| 5 | अपन कित्ती मेनत से पइसा कमवा जद जेक पइसा आवे अरे लजवाए मत तू जत्ता होरिया उठता देदुवा हूँ तो | R2_5.wav |
| 6 | लो पटलन बैठो मत रोटी बनाई लो पर म्हारे रोटी बनता नि आवे हे तो देवी या ब्याव का पेला केबा की थी नि | R2_6.wav |
| 7 | हूँ कई केरी अपन जद से इंदौर अयि नि जद से छोरिया होरके देख री जीन्स अने छोटी छोटी फिराक म्हारे भी दिवा दोनी | R2_7.wav |
| 8 | तो अबे तम सूत पेरिको घूमोगी कैको गाम से आगियो कांदा वांदा बेचबा तो हन नि दीवाबा का दस बहाना होव हे | R2_8.wav |
| 9 | पापा तमने कीधी नि की तोये कईको भी अपनी छोरी नि देगो देखो मेने छोरी ढूंढ ली ओर मांडो भी कर लियो | R2_9.wav |
| 10 | कित्ती बीघा जमीन हे साब तमारे ने एक अलावा ओर कई कई हे पच्चीस कट्टा लसन पच्चीस कट्टा प्याज का हे | R2_10.wav |
| 11 | ब्याई जी तमने पेला नि की तमारे पास लसन प्याज हे नितो तमारे आबा से पेला ही रिश्तो पक्को कर देता हमतो | R2_11.wav |
| 12 | तमारो कुत्तो भोक रियो थो मेने केदी उसे दादा अपना ही गाम का हे अपना गाम में ही तो रा रोटी पटक दवा चुप होजा | R2_12.wav |
| 13 | हूँ कई केरी थी सेठबा का मक्का पकी गिया थोड़ा भोत तोड़ लेता नि अरे कई जूता दीवाड़े कई तू सेठ बा का म्हारे | R2_13.wav |
| 14 | सेठ बा देखि लेगा तो मरेगा काम घसीटेगा ज्यादा अरे कई नि देखेगा हूँ यही तो ऊबी आगा तो केई दुआ नि हूँ तमसे | R2_14.wav |
| 15 | ऐ ढोर होरके चारा पानी कर दीजे जारियो हूँ शाजापुर म्हारी बात सुनो सामान लिखिलियो कई कई लानो शाजापुर से की भूली गया | R2_15.wav |
| 16 | सब याद हे मोये नारियल लानो ने राखी लानी हे और मिठाई का डब्बा हुन नि लाना हे कई ओर कई का भी मत उठा लाजो | R2_16.wav |
| 17 | म्हारी बात सुन दो किलो दूध लिया आउवा काका के यहाँ से तो ओके हेनी ओट लीजे राते ओर बना लीजे जसि भी बनानी हो | R2_17.wav |
| 18 | लेइ जाजे धारा भाई का नाह ले आउवा चार पांच राखी नखरा हिनी समेटा रिया अनका इत्ता लाजो उत्ता लाजो उनका हाथ से पांच पांच रुपया नि छूटेगा | R2_18.wav |
| 19 | बस यही कंजूसी करजो तम बस इका भरोषा बेठया करो की म्हारा घर का कत्तो देगा ने कत्तो नि देगा | R2_19.wav |
| 20 | देख तो गलकिया हुन सुखी गिया ओर अनसे पानी दबा की नि बनरी इनके तो आज बताऊ हुन सुणजो ओ | R2_20.wav |
| 21 | कई केरी हे करले तू आई रियो हुन थोड़ी देर से ए थारी की नवरो नि रेनदे दन भर चिल्ला कई दन भर मुंडो चला दुखे नि कई थारो | R2_21.wav |
| 22 | क्यों यो भपको या पड़ी रियो हे तो तमारे भगवान ने अतरो पानी दियो तो पानी अणि गलकिया होर के फेर दोनि देखो तो पत्ता कैसा सुखी रिया हे | R2_22.wav |
| 23 | अरि जीजी इ चार बेलड़ा हे गलकिया का अनका चक्कर में हुन म्हारी साडी सोयाबीन सूखा दू अने निन्दु खोदू वाकी जगह के अणि गलकिया होर में लगियो रु | R2_23.wav |
| 24 | चार गिलकिया अजा जद तो तम पडोसी होर के टब भरी भरी के दो उ कई नि हे पाई दो पानी फटाफट सांज तक पादीजो नितो म्हार से बुरो कोई नि फेर | R2_24.wav |
| 25 | थारा दाज्जी को नौकर नि हे म्हारा हिसाब से पा दुआ जा याह से चुपचाप थोड़ी देर आगा नि के नाल्डानो चालू कर देगा | R2_25.wav |
| 26 | पटलन थोड़ी देर काम करबा आये ओमे ही नाल्डा याहकी जगह दन्वा से लेके साँझ तक काम करेगी जद पतों चलेगो | R2_26.wav |

Figure 4.1 Text Data

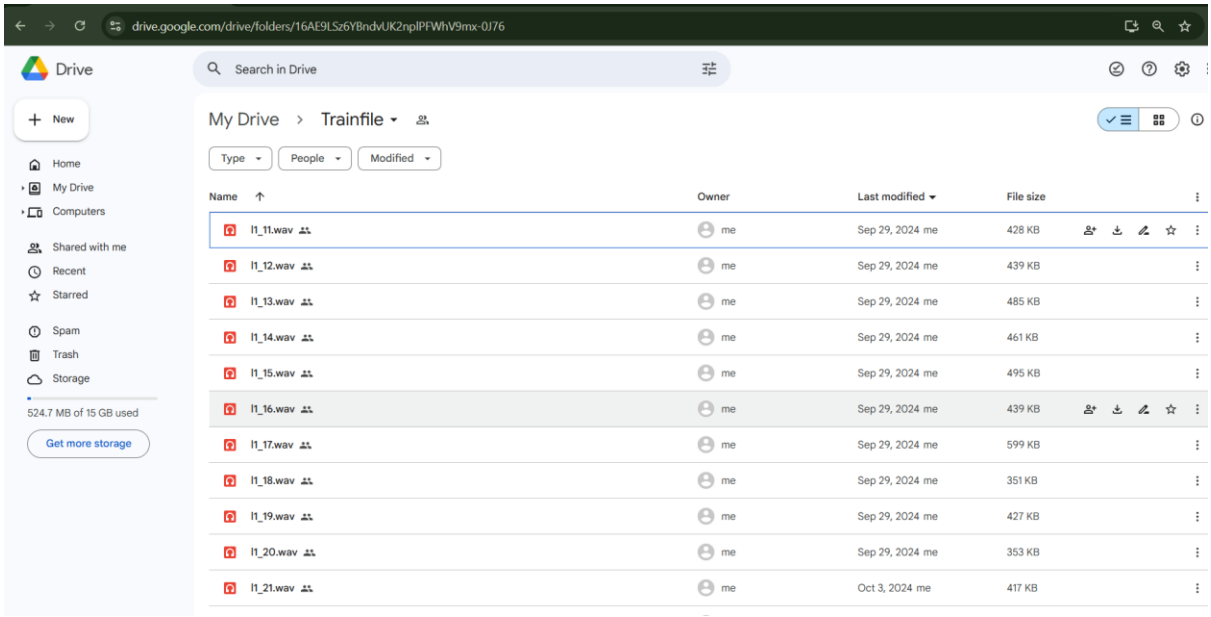



Figure 4.2 Audio Data

# Chapter 5

## 5. Conclusion and future scope

**5.1 Conclusion:** we have successfully completed the data collection phase for the Malvi language, which includes 8.54 hours of speech data recorded from 34 native speakers representing various dialects and accents. This data has been meticulously transcribed, annotated, and refined to ensure it accurately reflects the linguistic diversity of the language.

**5.2 Future Scope**: The next phase of the project will focus on building the Automatic Speech Recognition (ASR) model using this dataset. We plan to employ advanced techniques such as Wav2Vec2.0, where the model will process the raw speech audio, extract latent audio representations, and use a transformer-based approach for improved recognition accuracy. This will enable the development of an ASR system capable of recognizing and transcribing Malvi speech.

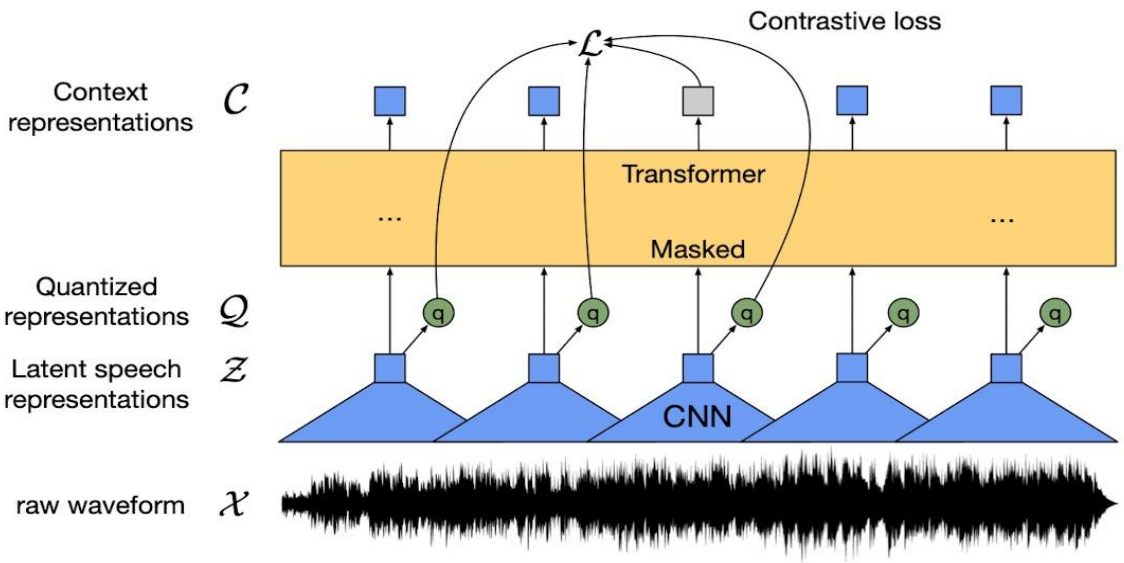



Figure 5.1 Wav2.vec2.0 model

The project focuses on developing an end-to-end Automatic Speech Recognition (ASR) system using Wav2Vec 2.0. At this stage, the dataset collection process has been completed, which provides a solid foundation for the upcoming steps in the project.

# References

[1] A Singh, AS Mehta, KS Ashish Khuraishi, G Deekshitha, G Date, J Nanavati, J Bandekar, "**An ASR Corpus in Chhattisgarhi, a Low Resource Indian Language",** *International Conference on Speech and Computer, 2023*

[2] **Kozhirbayev, Zhanibek**. "Kazakh Speech Recognition: Wav2vec2. 0 vs. Whisper." *Journal of Advances in Information Technology* 14.6 (2023): 1382-1389.

[3] **Williams, Aiden, Andrea Demarco, and Claudia Borg**. "The Applicability of Wav2Vec2 and Whisper for Low-Resource Maltese ASR." *Proc. 2nd Annual Meeting of the ELRA/ISCA SIG on Under-resourced Languages (SIGUL 2023)*. 2023.

[4] **Upadhyaya, Prashant,** et al. "Continuous hindi speech recognition using kaldi asr based on deep neural network." *Machine Intelligence and Signal Analysis*. Springer Singapore, 2019.

Census of India, 2011. *Linguistic Survey of India (Volume 5)*. Retrieved from https://language.census.gov.in/eLanguageDivision_VirtualPath/LSI_Reports/pdf/5.pdf

Wikipedia