| **FORM 2**  **THE PATENT ACT 1970**  **(39 OF 1970)**  **&**  **The Patents Rules, 2003**  **COMPLETE**  **SPECIFICATION** |
| --- |
| 1. **TITLE OF THE INVENTION**   Speech to text conversion for local ethnic languages |
| 1. **APPLICANT(S)** 2. NAME: Francis Xavier Engineering College, Tirunelveli 3. NATIONALITY: Indian 4. ADDRESS:   103/G2, Bypass Road,  Vannarpettai, Tirunelveli - 627003  Tamil Nadu, India   1. INVENTORS:   Dr. R.Ravi |
| **COMPLETE**  The following specification particularly describes the invention and the manner in which it is to be performed. |

**Field of invention**

This involves the development of systems that can accurately transcribe spoken language into written text. For local ethnic languages, it requires acoustic modeling, phonetic analysis, and tuning models to handle diverse accents and dialects.

**Objective of invention**

Make digital content and services accessible to speakers of local ethnic languages, enable users to interact with technology, access information, and perform tasks in their native languages, bridging the digital divide.

Achieving these objectives involves addressing the unique challenges posed by local languages, such as diverse dialects, limited linguistic resources, and variations in phonetics.

**Detailed description:**

Speech-to-text (STT) conversion technology, also known as automatic speech recognition (ASR), transforms spoken language into written text. For local ethnic languages of Madhya Pradesh (MP), such as Chhattisgarhi, Bhili, Gondi, and others

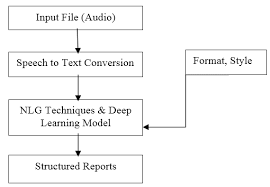
Identified problem:

Many local ethnic languages in MP lack substantial, high-quality datasets needed for training accurate STT models.

Proposed Solution:

Partner with local communities, universities, and linguistic organizations to collect diverse and representative speech data,conduct surveys and recording sessions in various settings (rural, urban, formal, informal) to capture the richness of language use.

 **Block diagram:**

****

**Block description 1:**

The system begins with capturing the user's speech through a microphone. The audio signal is preprocessed to remove noise and enhance quality for accurate conversion. The captured audio is analyzed to determine the specific ethnic language being spoken. This step ensures the correct linguistic model is applied for conversion.

**Hardware setup**

**Data Collection Hardware:**

**Audio Recording Devices**

* **High-Quality Microphones**: Use high-fidelity microphones to capture clear and accurate audio from speakers. Lavalier microphones, shotgun microphones, and USB condenser microphones are good choices.
* **Portable Recorders**: Use portable digital audio recorders for fieldwork and on-site data collection. Devices should support high-resolution audio recording (e.g., 24-bit/48kHz).

**Acoustic Environment**

* **Soundproofing**: Set up soundproof or acoustically treated rooms to minimize background noise and improve audio quality during recordings.
* **Noise-Canceling Equipment**: Use noise-canceling headphones for monitoring audio during recording sessions to ensure clarity.

**Real-Time Processing Hardware:**

**Edge Devices**

* **Local Servers**: For real-time or near-real-time applications, deploy local servers or edge devices with sufficient processing power to handle speech recognition tasks on-site.
* **Embedded Systems**: Use embedded systems or single-board computers (e.g., Raspberry Pi, NVIDIA Jetson) for specific applications where real-time processing is required at the edge.

**Network Infrastructure**

* **High-Speed Network**: Ensure robust and high-speed network infrastructure to handle large volumes of data transmission between devices and servers, especially if cloud-based processing is involved.
* **Redundant Connectivity**: Implement redundant network connections to minimize downtime and maintain system reliability.

**Network Interface:**

For cloud-based processing or updates, a reliable network interface is necessary. Devices for capturing user feedback, which can be a simple interface on the computer or additional hardware like a feedback button.

**Software setup**

Operating System**:**

The right operating system (OS) for developing a speech-to-text conversion system for local ethnic languages is crucial for ensuring compatibility with necessary tools and libraries, ease of use, and efficient performance. Here are some considerations and recommendations for selecting an OS:

**Acoustic Model Training Software :**

Speech recognition of a language is a key area in the field of pattern recognition. This paper presents a comprehensive survey on the speech recognition techniques for non-Indian and Indian languages, and compiled some of the computational models used for processing speech acoustics. An immense number of frameworks are available for speech processing and recognition for languages persisting around the globe. However, a limited number of automatic speech recognition systems are available for commercial use

**Data Collection:**

Data collection is a crucial step for developing an effective speech-to-text conversion system for local ethnic languages. High-quality and diverse data is essential to train robust models. Here’s a detailed guide on how to collect and prepare data Identify the specific local ethnic languages and dialects you want to cover.Determine the types of data needed (e.g., conversational speech, read speech, spontaneous speech). Estimate the amount of data required for training, validation, and testing.Ensure the inclusion of speakers of different ages, genders, and backgrounds to capture a variety of accents and speaking styles.Obtain consent from participants for recording and using their speech data.

**Claims:**

We claim

1. **Cultural Preservation and Promotion**

Camera

Cultural Heritage Microcontroller

1. Social Impact claimed in claim 1 as, Promotes social inclusion by ensuring that speakers of all languages have access to the same technologies and services. Empowers speakers of local languages by giving them tools to express themselves and access information in their native language.
2. **APPLICATIONS:**

* This device can be applied in various domains,.