|  |
| --- |
| **FORM 2**  **THE PATENTS ACT, 1970**  **(39 of 1970)**  **&**  **The Patents Rules, 2003**  **PROVISIONAL/COMPLETE SPECIFICATION**  **(See section 10 and rule 13)** |
| **1. TITLE OF THE INVENTION:** **Speak2Summarize : Daily Task Recap** |
| **2. APPLICANT(S)**  (a) NAME:  (b) NATIONALITY: Indian  (c) ADDRESS:. Department of Computer Science Engineering, Y.C.C.E., Hingna Road, Wanadongri, Nagpur-441110 |
| **3. PREAMBLE TO THE DESCRIPTION:**  The following specification particularly describes the invention and the manner in which it is to be performed. |
| **COMPLETE SPECIFICATION:**  **1. FIELD OF THE INVENTION:**  This invention relates to the Machine Learning and Natural Language Processing fields concerning Speech-To-Text (STT) transformations and Text Summarization. It includes the recent techniques of Automatic Speech Recognition (ASR) in the context of deep learning architectural models and open source software . Also, it incorporates methods of summarization by use of these open source models and techniques aimed at enhancing data process automation, accessibility and content creation. The invention is useful for many sectors such as systems for transcription of protocols of meetings, automatic text summarization systems, and voice writing systems and daily voice based journalling .  **2. BACKGROUND OF THE INVENTION AND PRIOR ART:**  The improvement of AI and NLP technologies has created a new demand for efficient speech-to-text transformation and text summarization. Earlier speech recognition systems used statistical methods like Hidden Markov Models (HMM) and Gaussian Mixture Models (GMM) because of their effectiveness. These systems, however, failed in dealing with accent variations and noisy background systems. The accuracy and efficiency of speech to text systems with deep learning methods has advanced with the introduction of various forms of RNNs, LSTMs and transformer architectures such as BERT and GPT. These systems now recognize speech with greater speed and precision in multiple languages and dialects.  At the same time, large volumes of spontaneously generated digital content requires advanced text summarization tools that can efficiently extract relevant information. The previously existing rule-based and statistical methods for extracting summaries of a text have slowly been transitioned to incorporate machine learning and transformer techniques which are far more capable of producing accurate and concise summaries.  The enhancement of speech to text tools through the combination of speech recognition and automated summarization has benefits for virtual assistants, meeting transcription software, podasts, content analysis, and speech accessibility applications. Multi-speaker environments, background noise, context understanding, and handling specific vocabularies still pose a challenge.  This invention seeks to address these issues through the application of advanced deep learning and NLP models aimed at improving the power and precision of speech recognition, text comprehension, and, ultimately, human-machine communication, information retrieval, and content consumption  **Prior Arts:**  **Multi-Modal Voice Recognition System and Method for Conversation Summarization (Patent No. 18/540,594, Filed: 2024)**  The patent deals with the methodologies and systems for summarizing any conversation in conjunction with multi-modality voice recognition inputs, which may include the following issues in real-time performance, bias, and sensitivity to context.  Patent No. 18/540,594 outlines a multi-modal voice recognition system designed to summarize conversations by leveraging both media (audio, video, images) and text inputs. The invention segments conversation data into media and text components, applying specialized machine learning models to each for enhanced topic modeling and key element extraction. These elements are then grouped to create concise headline banners and context-rich summaries. Particularly effective in real-time, multi-user environments like messaging platforms, the system ensures accurate, unbiased, and holistic summarization by treating all input modalities fairly and context-sensitively.  **3. SUMMARY OF THE INVENTION**  **4. OBJECTIVE OF THE INVENTION:**   * + To make speech-to-text and summarization tools more accessible to **visually impaired users, professionals, students, and organizations** requiring fast and reliable transcription and summarization services.   + To enable individuals with memory related problems to **record, transcribe, and summarize conversations**, providing a textual reference for later recall.   **5. BRIEF DESCRIPTION OF THE DRAWINGS:**  **A diagram of a computer  AI-generated content may be incorrect.**  Fig 1. Circuit diagram of the audio recording apparatus  1. **Setup Phase**   * RTC is initialized. If invalid, it's set to the current time from the computer (\_\_DATE\_\_ & \_\_TIME\_\_ macros). * SD Card is initialized. * Buttons and LEDs are configured.   2. **Recording Phase**   * When the Start button is pressed:   + The RTC fetches the current date and time.   + Unique .mp3 filename is generated using the format: YYYYMMDD\_HHMMSS.mp3   + SD card creates a new file with that name.   + LED1 turns ON, indicating recording has started. (You would need to interface the MAX9814 with a proper ADC or external recorder for actual audio.)   3. **Stop Phase**   * When the Stop button is pressed:   + The file is closed to ensure it's saved correctly.   + LED1 turns OFF, and LED2 turns ON briefly to indicate successful saving of the audio |
| **5. DESCRIPTION**  After observing these problems faced by individuals with memory and the individuals like professials and students who has to consumer large amount of audio based data everyday, we decided to develop a system for effective processing of the audio data and generate text summaries so that the individuals and easily get gist of the happenings of the day. The Arduino serves as the central processing unit of the recording system. It receives the audio signals from the **MAX9814 Electret Microphone Module**. The Arduino processes this audio signal and saves the audio file along with the timestamp generated by the RTC into the Micro SD Card which is accessed thorough Micro SD Card Reader module connected to the **Arduino Nano (ATmega328P)**. The user then put this Micro SD Card into their system and the software we have developed, using Open Source Maching learning and NLP model processes the audio present in the Micro SD Card. The Micro SD Card is FAT32 formatted which give it abilty to work with any operating system.  The software as mentioned consists of three main models which are Speaker Diarization, Speech Recognition Model and Text Summarization Model. These models works sequentially in the aforementioned order. The Speaker Diarization Model Breaks the audio into the segments if there are more than one speaker present in audio data (this is possible in sceanario such as meetings and interviews); the process is done in memory. The Speech Recognition Model then transcribe this segmented data in the sequential format based on timestamping. The transcribed data which is currently in text format is stored for the user to read in detail, along with this the text data is also fed into the Text Summarization Model to generate a concise and meaningful summaries while maitaining the context awareness. This summarised data is then presented to user with the aim of helping them remember their daily conversations to improve their productivity. |
| **6. CLAIMS**   1. **Independent Claims**    * A Audio Recording and Storage System, comprising:      + A microphone module (MAX9814) for capturing speech input with automatic gain control (AGC).      + A microcontroller (Arduino Nano) for processing audio signals and managing data acquisition.      + A real-time clock (DS1302 RTC) for timestamping speech recordings.      + A storage module (Micro SD card) for storing transcriptions and summaries.    * A method for converting speech to text and generating text summaries, comprising:      + Capturing speech input via a microphone module.      + Preprocessing audio using noise reduction and segmentation techniques.      + Converting speech to text using an ASR model based on Transformers.      + Applying text summarization models to generate summaries. 2. **Dependent Claims**    * The system of claim 1, wherein the system processes audio signals to digital audio and stores the audio file in Micro SD Card.    * The system of claim 1, wherein users can manually start and stop recording along with storing the timestamp of when the recording has started.    * The system of claim 2, wherein the summarization process includes abstractive techniques, utilizing pre-trained models fine-tuned for speech-based text summarization.    * The system of claim 2, wherein transcriptions are automatically segmented and summarized based on number of speakers.    * The system of claim 2, wherein the audio data is processed to a .wav file format so that pre-processing can be performed efficiently.    * The system of claim 2, wherein the speech recognition system filters background noise and enhances speech clarity using noise reduction techniques. |
| **7. ABSTRACT OF THE INVENTION**  The present invention relates to an **Speech-to-Text (STT) and Text Summarization System** that integrates **hardware and software** to efficiently convert speech into accurate text and generate concise, context-aware summaries. The system is designed to assist individuals to get a record of their daily summaries as well as individuals with **memory-related challenges as** by providing **transcription, conversation recall, and automated summarization**. The invention comprises a **hardware module** that includes a **MAX9814 electret microphone for speech input**, an **Arduino Nano (ATmega328P) for audio signal processing**, a **DS1302 Real-Time Clock (RTC) for timestamping**, and an **SD card storage unit for saving transcriptions and summaries**. The **power management system** ensures seamless operation, making it suitable for both **portable and embedded applications**. On the software side, the system utilizes **Natural Language Processing (NLP) and Deep Learning models**, such as **Transformer-based Automatic Speech Recognition (ASR) models (WhisperX, Wav2Vec 2.0)** for **highly accurate speech-to-text conversion**, and a**bstractive summarization techniques** to generate meaningful summaries. The system is currently adaptable to Englishmaking it suitable for applications in **assistive technologies, virtual assistants, business transcription services, and education**. By combining **AI-driven text summarization, and memory-assistive features**, this invention provides a **scalable, accessible, and efficient solution for speech recognition and content retrieval**, enhancing accessibility and information retention for a wide range of users. The system aims to help with memory retention by keeping a log of summarised day to day interactions of an individual. |
| **8. DATE AND SIGNATURE** (to be given at the end of last page of specification). |

**Note:-**

**\*Repeat boxes in case of more than one entry.**

**\*To be signed by the applicant(s) or by authorized registered patent agent.**

**\*Name of the applicant should be given in full, family name in the beginning.**

**\*Complete address of the applicant should be given stating the postal index no./code, state and country.**

**\*Strike out the column(s) which is/are not applicable.**