MINOR PROJECT - (SRS)

1. Introduction

1.1 **Purpose**

This Software Requirements Specification (SRS) document outlines the requirements for developing a system that assists visually impaired users by providing detailed auditory descriptions of their environment. The system captures video input, processes it through image detection and segmentation, utilizes a Large Language Model (LLM) to generate comprehensive textual descriptions, and converts these descriptions into speech using text-to-speech (TTS) technology.

1.2 **Scope**

The system aims to enhance the autonomy of visually impaired individuals by offering real-time, detailed auditory scene descriptions. Key functionalities include:

- Capturing video input from the user's environment.
- Performing image detection and segmentation to identify and delineate objects and features.
- Utilizing an LLM to generate detailed textual descriptions based on the identified features.
- Converting the textual descriptions into natural-sounding speech using TTS libraries in Python.

1.3 **Definitions, Acronyms, and Abbreviations**

- **LLM**: Large Language Model
- **TTS**: Text-to-Speech
- **API**: Application Programming Interface

1.4 **References**

- Asana's Software Requirement Document Template citeturn0search0
- IEEE Recommended Practice for Software Requirements Specifications citeturn0search10

2. Overall Description

2.1 **Product Perspective**

The proposed system integrates multiple technologies to transform visual information into auditory descriptions, enhancing environmental awareness for visually impaired users.

2.2 **Product Functions**

- **Video Capture**: Obtain real-time video input from the user's environment.
- **Image Detection and Segmentation**: Identify and segment key components within the captured video frames.
- **Text Generation**: Use an LLM to create detailed textual descriptions based on the segmented image data.
- **Speech Output**: Convert the generated text into speech using Python-based TTS libraries.

2.3 **User Characteristics**

Primary users are visually impaired individuals seeking tools to better understand and navigate their surroundings.

2.4 **Constraints**

- Real-time processing requirements to ensure timely feedback.
- Accuracy of image detection and segmentation to provide reliable descriptions.
- Naturalness and clarity of the synthesized speech.

3. Specific Requirements

3.1 Functional Requirements

3.1.1 Video Capture

- The system shall capture continuous video input from an integrated camera.

3.1.2 Image Detection and Segmentation

- The system shall detect and segment objects and features within each video frame.

3.1.3 Text Generation

- The system shall utilize an LLM to generate textual descriptions based on the segmented image data.

3.1.4 Speech Output

- The system shall convert textual descriptions into speech using Python-based TTS libraries.

3.2 Performance Requirements

- The system shall process video input and provide auditory descriptions with minimal latency to ensure real-time feedback.

3.3 Design Constraints

- The system shall be implemented using Python and compatible libraries for image processing, LLM integration, and TTS functionalities.

4. External Interface Requirements

4.1 User Interfaces

- The system shall provide auditory feedback through headphones or speakers.

4.2 **Hardware Interfaces**

- The system shall interface with a camera to capture video input.
- The system shall interface with audio output devices for speech delivery.

4.3 **Software Interfaces**

- The system shall utilize APIs for image processing, LLM integration, and TTS functionalities.

5. Other Nonfunctional Requirements

5.1 **Performance**

- The system shall process and describe scenes in real-time to ensure immediate feedback.

5.2 **Usability**

- The system shall provide clear and natural-sounding speech output to ensure user comprehension.

5.3 **Reliability**

- The system shall accurately detect and describe environmental features to provide reliable information to the user.

6. Appendices

6.1 **Assumptions and Dependencies**

- The system assumes access to a stable power source and functional hardware components (camera and audio output devices).

Note: This SRS is a foundational document and may require updates as the project progresses and new requirements emerge.