Specification Document Of

# Final Production Project

Submitted by: Sameer Basnet

Student Id: 77356702

BSc (Hons) Computing

## Breaking the Slience: AI and Computer Vision Driven Sign Language Translation System

Project Aim:

The main aim of the project is to help AI developers to use computer vision and AI to help deaf and mute persons to communicate with people in English language in form of text or speech by translating their hand gestures into English language by using computer vision and machine learning technique’s.

Objectives:

The main objective is to develop a product which:

* **To develop a system that accurately detects and recognizes hand gestures in real-time.**
* **Translate recognizes hand gestures into English language for adequate communication.**
* **Simple UI that allows users to easily interact with the system and view output.**
* **Integrate Text-to-Speech (TTS) module into the product to convert output text to speech for enhance accessibility of the product for users.**
* **To improve the system robustness, a real-time error handling mechanism will be integrated that provides real time feedback when a gesture is not recognized or detected, ensuring a smooth and user-friendly experience.**

Product Specification:

The product specification is categorized into functional and non-functional requirements and is described below with MoSCoW method:

|  |  |
| --- | --- |
| **Functional Requirements** | **MoSCo W** |
| **Real- time gesture detection and recognition** | M |
| **Translation of recognized gesture into English text** | M |
| **Integration of hand detection and recognition models** | M |
| **User Friendly UI for user to use the product** | M |
| **Integration of Text-to-Speech (TTS) module** | S |
| **Real time Error Handling System in case of gesture not recognized or detected** | S |
| **Integration with mobile platform for portability** | C |
| **AR/VR integration** | W |
| **Complex Dynamic Sign Language Grammar Support** | W |
| **Multi-User real time collaboration for gesture translation.** | W |

|  |  |
| --- | --- |
| **Non-Functional Requirements** | **MoSCo W** |
| **90 percent accuracy in gesture recognition** | M |
| **Product is tested and validated** | M |
| **Product is user friendly** | S |
| **Product shall be platform independent** | C |
| **Product must maintain stable and consistent performance in diverse conditions.** | **M** |
| **Product UI to be accessible for users with little technical knowledge.** | **S** |
| **Provide support for external Camera** | **C** |
| **Development on cloud platform for large-scale use.** | **W** |
| **Real time integration with AR/VR glasses** | **W** |

Research:

Sign language Detection (SLR) refer to the process of recognizing and interpreting gestures, hand movements and body language used in sign language to accommodate communication between people who use sign language and who do not. The field of SLR has witnessed eloquent improvements over years, staring from neural network-based systems to deep learning based systems. Early work by (Kim, S. et. al., 1995) explored two stage neural network that uses a DataGlove for phoneme-level recognition, achieving 86% accuracy but was constrained by hardware reliance. Another research by (Cui, Y. et. al., 1995) used self-organizing frameworks like SHOSLIF-M, enhancing spatiotemporal recognition with a 96% accuracy rate. However, its reliance on handcrafted features and limited scalability to real-life scenarios posed challenges. Another Research by (Wadhawan, A. et. al., 2021) focus on static, isolated, single-handed signs using camera-based systems, highlighting a lack of standardized datasets and a need for dynamic sign recognition advancements. The use of CNNs by (Kumar, P. et. al., 2020) showed remarkable performance with 99.90% accuracy for static signs using extensive datasets, emphasizing the effectiveness of deep learning. Use of statistical methods for continuous sign language by (Koller, O. et. al., 2015) address real-life variability, employing multimodal features like facial landmarks and achieving significant word error rate reductions. Another Study by (Bauer, B. et. al., 2002) utilized statistical methods and fenones for efficient sign recognition but struggled with robust subunit definition for diverse gestures. Another study by (Ouhyoung, M. et. al., 1998) developed a real-time sign language interpreter using a data glove and HMMs but faced challenges with signer dependency and limited accuracy.

Evaluation:

The assessment of the Sign Language Recognition System will focus on its primary and extended functionalities as outlined in the product specifications and objectives. Initially, each component will be tested separately, including hand detection, gesture recognition, and text-to-speech integration, concentrating on their speed of response, accuracy, and reliability. Subsequently, the entire workflow will be assessed for real-time capability and system response time. The evaluation process will involve both quantitative metrics (such as recognition accuracy, response time, and error rates) and qualitative feedback (like user opinions on the system’s usability). Additionally, the system will undergo testing in diverse scenarios, considering factors such as varying lighting conditions, hand sizes, and positions, in order to validate its robustness and scalability.

Project Planning and Methodology:

After research, the project will kick off by selecting a suitable dataset for gesture recognition, which could include the ASL dataset or a newly created dataset comprising both static and dynamic signs. Pretrained hand tracking models, such as Mediapipe, as well as TensorFlow or PyTorch models for gesture detection, will be evaluated. The research will adopt a modular strategy, starting with an evaluation of the performance of each model in tasks such as hand detection and gesture classification. These models will be integrated into a pipeline designed to recognize gestures in real time. Metrics such as accuracy, latency, and robustness in varying conditions will be measured for each component prior to being used within the complete system. The Agile methodology will emphasize iterative development and ongoing testing. The planning process will be clarified using a project timeline and Gantt chart to ensure clear milestones and efficient time management.

Project Timeline

A screenshot of a computer

Description automatically generated

Fig: *The above Project Timeline shows the project planning along with the dates.*

Gantt Chart:

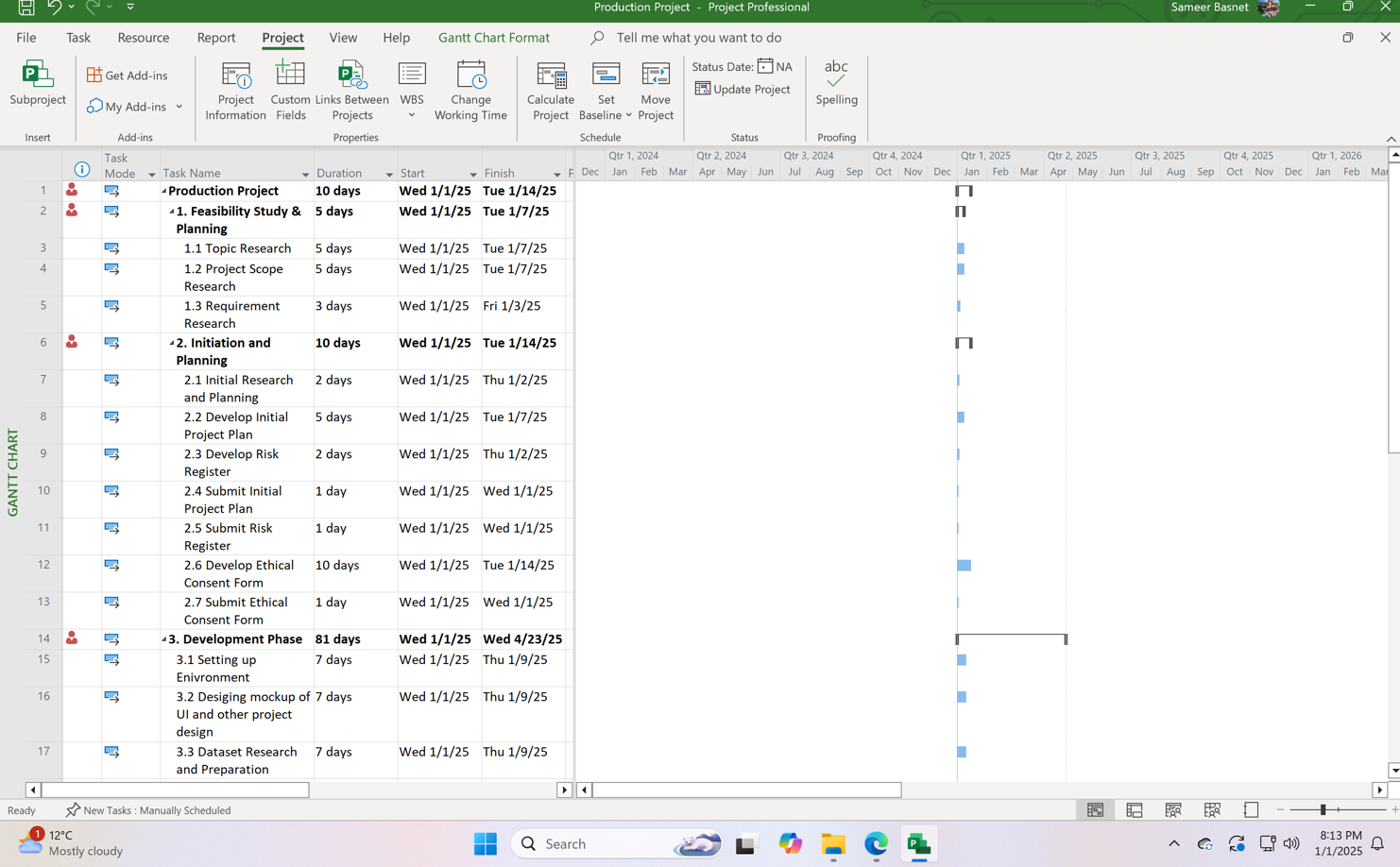


Fig: *The above Gantt chart shows the project planning along with the key dates.*

Resources:

The list of resources (hardware and software) is listed below. All the software is free open source products and can be easily downloaded online.

**List of Software:**

* Python
* PyCharm
* OpenCv
* Mediapipe
* MS Word
* MS Powerpoint
* MS Project
* MS Excel
* TensorFlow/Keras
* PyTorch
* Scikit-learn
* Pandas
* NumPy
* pyttsx3
* Matplotlib
* Seaborn
* TensorFlow Lite
* PyInstaller
* Tkinter
* HTML
* CSS
* JavaScript
* PHP
* Django
* Convolutional Neural Networks (CNNs)
* Recurrent Neural Networks (RNNs)
* **Transformers**
* Hidden Markov Models (HMMs)
* Discord
* Draw.io
* Google Drive
* Github
* Google doc
* Google Slide
* Google Sheets
* Gmail
* Google Scholar

|  |  |
| --- | --- |
| **List of Hardware:**   * MacBook Air M1 13 inch * External Webcam |  |

Bibliography:

* **Bauer, B., Swets, D.L., Weng, J.J. Friedrich, K.K., (2002) Towards an Automatic Sign Language Recognition System Using Subunits. Springer Nature Link. [Online]. 2298, Available from: <https://link.springer.com/chapter/10.1007/3-540-47873-6\_7> [Accessed 31 Dec 2024].**
* **Cui, Y. , (1995) Learning-based hand sign recognition using SHOSLIF-M. IEEE. [Online]. 0 (10.1109/ICCV.1995.466879), pp. 0. Available from: <https://ieeexplore.ieee.org/abstract/document/466879> [Accessed 31 Dec 2024].**
* **Kim, S., Waldrom, M.B. (1995) Isolated ASL sign recognition system for deaf persons. IEEE. [Online]. 3 (3), pp. 261 - 271. Available from: <https://ieeexplore.ieee.org/abstract/document/413199> [Accessed 31 Dec 2024].**
* **Koller, O. (2015) Continuous sign language recognition: Towards large vocabulary statistical recognition systems handling multiple signers. Computer Vision and Image Understanding. [Online]. 141, pp. 108-125. Available from: <https://www.sciencedirect.com/science/article/abs/pii/S1077314215002088> [Accessed 31 Dec 2024].**
* **Kumar, P.** Wadhawan, A.**(2020) Deep learning-based sign language recognition system for static signs. S.I. : Hybrid Artificial Intelligence and Machine Learning Tech. [Online]. 32 , pp. 7957–7968. Available from: <https://link.springer.com/article/10.1007/s00521-019-04691-y#citeas> [Accessed 31 Dec 2024].**
* **Ouhyoung, M., Ling, R.H., (1998) A real-time continuous gesture recognition system for sign language. IEEE. [Online]. pp. 10.1109/AFGR.1998.671007. Available from: <https://ieeexplore.ieee.org/abstract/document/671007/authors#citations> [Accessed 31 Dec 2024].**
* **Wadhawan, A., Kumar, P. (2019) Sign Language Recognition Systems: A Decade Systematic Literature Review. Springer Nature Link. [Online]. 28 , pp. 785–813. Available from: <https://link.springer.com/article/10.1007/s11831-019-09384-2> [Accessed 31 Dec 2024].**