Proposing Integration of Speech-Text Recognition with Animation Education to Support Deaf People

Abdulaziz Al Raisi[1] and Khoula Al Harthy[1#]

[1]Middle East College, Muscat, Oman
[#]Advisor

ABSTRACT

In Oman, the community needs to give more attention to deaf and hard-hearing people, specifically using American Sign Language (ASL). Yes, there are special condition schools that will host the anthem and teach them, but ASL people will not join the regular school and colleges. They will only attend standard classes and learn the lectures with ordinary people. This study proposes the integration of speech-text recognition technology with animation education as a means of supporting deaf individuals. The use of animation as a teaching tool has proven effective in engaging and educating diverse learners. By integrating speech-text recognition, deaf students can access the audio component of animation education and fully participate in the learning experience. Through a pilot study, we demonstrated the feasibility and benefits of this approach by providing animation education to a group of deaf students using speech-text recognition technology. Results showed improved understanding and engagement in the material and increased confidence and communication skills. The proposed integration has the potential to improve access to education for deaf individuals and enhance the effectiveness of animation as a teaching tool for all learners.

Introduction

Every human has the full right to be educated, either a normal one or for whom they have exceptional conditions. Yes, many states require a unique approach to teaching and letting them feel equal to normal ones. And it might take a lot of time, cost, and resources to provide for each exceptional condition. However, we were in 2022, and technologies are the key to building a solution customized for each one. Considering the schools and colleges in Oman are provided with technologies, it is only for normal ones. Since the environment is not yet ready, the Deaf and Hard-hearing-ones will not be able to join and learn with the normal equally. The technologies available now are much more capable, stable, and flexible for building an educational solution for deaf/hard hearing. No offense to the special schools for these people, but it is the time to let them feel equal in education and have their roadmap to graduation.

In Oman, the community needs to give more attention to deaf and hard-hearing people, specifically using American Sign Language (ASL). Yes, there are special condition schools that will host the anthem and teach them, but ASL people will not join the regular school and colleges. They will only attend standard classes and learn the lectures with ordinary people. They have the full right to graduate from any college with any degree. Still, the education environment in Oman is ready to educate them and make them feel the journey of study once they graduate. The below figure shows the values of disabilities in Oman. Stating the problem of the ASL people (Def & hard hearing) are kept aside and teaching them limited lecturers compared with regular they will mostly have a road until graduation. The education environment needs to be prepared with a solid solution to implement and put the ASL people on the regular roadmap. None of the colleges and schools in Oman have at least one auto ASL translator present and get them in to show them enjoinment and what the lecturers have been demanding from the regular people:

Configuration and integration are essential technical tasks that enable various systems and applications to work efficiently and seamlessly. One critical area in this field is AI and ML configuration and integration, which involves implementing artificial intelligence and machine learning algorithms into different systems and applications to enhance their capabilities. Additionally, animation configuration and integration enable the inclusion of animated graphics and visual effects into software programs, games, and multimedia content. Another important area is developing a translator from text or speech to American Sign Language (ASL) and integrating it into different applications to assist individuals with hearing disabilities. Cloud integration is another vital aspect that involves the incorporation of cloud-based services, platforms, and infrastructures into existing IT systems and applications. Lastly, network configuration plays a crucial role in setting up and optimizing network devices and protocols to facilitate seamless communication and connectivity between different computing devices and systems. The aim is: "Propose the Integration of Speech-Text recognition with animation in education to support Deaf people." To ensure the solution is sold and capable of providing the same education level for ASL people on the same campus as regular students.

The objectives of this solution are centered on addressing the challenges of educating students with disabilities in higher education institutions in Oman. Firstly, there is a need to evaluate the current issues and challenges that these students face to develop effective solutions. This involves assessing the current learning environment and identifying the various obstacles that hinder their education. Secondly, it is essential to analyze the current environmental requirements and possibilities of higher education institutions to determine the necessary modifications to support students with disabilities. This includes identifying the infrastructure, technologies, and support services required to ensure that they have equal access to educational opportunities and resources.The third objective is to design a comprehensive solution that translates from text or speech to American Sign Language (ASL) over cloud services. Fourthly, an AI & ML solution will be implemented in the education environment, and its database will be built over cloud computing to improve the learning outcomes for students with disabilities. Finally, testing the proposed solution is essential to ensure that it is easy to implement and can be used without difficulty. By focusing on these objectives, it is possible to develop an effective and sustainable solution that enhances the education and well-being of students with disabilities in higher education institutions in Oman.

The proposed solution is a comprehensive solution that can allow them to be integrated with regular students and benefit them greatly. The signs can be displayed over the display or as an animation using the robot. In addition, machine learning will be integrated to learn all the translations and feed the artificial intelligence, which can act independently with efficacy and do the translation smother. However, the solution will be cloud-based, taking the critical processes over the cloud and providing only the services over the hosted network.

Literature Review

Regarding understanding the proposed solution and supporting the main idea, a literature review is one of the best methods to apply. A literature review is an academic writing which will demonstrate and elaborate on the topic in context. In addition, it includes a significant evaluation of the subject. The literature review will cover the below technologies: Artificial intelligence, Machine learning, Cloud Computing, Voice recognition, and Text recognition. However, the demonstration with the literature review will explain: 1) How it is related to the proposed solution. 2)What are the fixed issues, and how to avoid or fix them in the proposed solution. AI is A technology concerned with developing the computer to have the capability to act human-like in learning, auto-correction, and logic. They have become the demand technology nowadays in various areas due to the ability and techniques used to solve the problem. Moreover, the AI can solve problems and self-act for a specific goal, which the computer will collect the data from the sensor and then take into process and action it. There are two types of AI. AI has become increasingly important in various

industries and domains. There are different types of AI, such as narrow AI and general AI, which have unique characteristics and applications. Narrow AI is designed to perform specific tasks and can be found in intelligent systems such as Siri on Apple devices and self-driving cars. It has many applications, including assisting with drone inspections, calendar preparation, customer service, hotel booking coordination, and medical diagnostics.

In contrast, general AI is entirely different and has human-like thinking, flexibility, and the capability of learning to perform various tasks, such as hair cutting. While general AI is popular in movies, it has many practical applications, such as creating AI-generated faces. AI is software embedded in specific hardware, built by humans, and configured to perform human-like capabilities, goals, and processes based on the data collected from the environment. It has several properties, including autonomy, adaptability, and interactivity. With the increasing development of AI, it has become more efficient, accurate, and effective in various industries and domains. Several types of artificial intelligence (AI) could be suitable for a speech-to-text-to-sign language translation project. One possibility is to use machine learning, a kind of AI that involves training algorithms on data to enable them to make predictions or decisions without being explicitly programmed (Mitchell, 1997). Machine learning algorithms, such as decision trees, support vector machines, and neural networks, have been successfully applied to various natural language processing (NLP) tasks, including part-of-speech tagging, named entity recognition, and machine translation (Jurafsky & Martin, 2019).

Another option is to use deep learning, a type of machine learning that involves training neural networks with many layers of interconnected nodes on large datasets (LeCun, Bengio, & Hinton, 2015). Deep understanding has achieved state-of-the-art results on various NLP tasks, including language translation and text classification (Jurafsky & Martin, 2019). Ultimately, the most suitable type of AI for a speech-to-text-to-sign language translation project will depend on the specific requirements and constraints of the application, as well as the availability of data and resources (Mitchell, 1997; Jurafsky & Martin, 2019). It may be necessary to evaluate various AI technologies to find the most effective solution. Machine learning is a field of artificial intelligence that focuses on developing algorithms and models that can learn from data and improve their performance over time. It involves using statistical and computational techniques to analyze and understand patterns in data and make predictions or decisions based on that analysis. In machine learning, data is used to train a model, then to make predictions or decisions without being explicitly programmed to perform those tasks. The model learns from the information and can continue improving its performance as it is exposed to more data. Machine learning makes it so powerful and helpful – it allows computers to learn and adapt to new situations without being explicitly programmed.

Many machine learning types include supervised, unsupervised, semi-supervised, and reinforcement learning. Each approach has unique characteristics and is suited to different tasks and problems. (Gupta, 2022). A specific algorithm's study is essential when using the computer to perform a task and calls for machine learning. It is part of AI and designed on a mathematical setup based on sample data, known as Training data, to predict and decide independently.(Chhaya et al., 2020). In machine learning, data types refer to the kind of values that can be stored in a particular variable or feature. Different types of data require different types of processing and analysis, and it is essential to understand the characteristics of each class to work with them in a machine-learning model effectively. (Precisely, Inc., 2021) (Lish, n.d.)

There are several common data types in machine learning, including:

1. Numeric data: This includes continuous and discrete numerical values, such as ages, heights, and weights. Numeric data can be further divided into integer and float data types, depending on whether the values are whole numbers or include decimals.
2. Categorical data: This includes data that can be organized into distinct categories, such as gender, hair color, and eye color. Categorical data can be either ordinal, meaning that the classes have a specific order or ranking, or nominal, meaning that the categories do not have a particular order.

3. Text data: This includes data in the form of words or phrases, such as comments, reviews, and descriptions. Text data requires special processing, such as tokenization and steam out, to be used in a machine-learning model.
4. Binary data includes only two values, true or false, or 0 and 1. Binary data is often used in classification tasks, where the goal is to predict one of two possible outcomes.

Machine learning algorithms are a set of methods that allow a computer to learn from data without being explicitly programmed. These algorithms can be classified into three main categories: supervised, unsupervised, and reinforcement learning. Supervised learning algorithms are trained on labeled data, meaning that the data used to prepare the algorithm includes both input data and the corresponding correct output. The algorithm makes predictions based on this input-output mapping, and the goal is to learn a function that can generalize to new, unseen data. Examples of supervised learning algorithms include linear regression, logistic regression, and support vector machines. On the other hand, unsupervised learning algorithms are trained on unlabeled data. The goal, in this case, is to find patterns or relationships in the data rather than making specific predictions. Examples of unsupervised learning algorithms include k-means clustering and principal component analysis.

Reinforcement learning algorithms are trained through trial and error, receiving rewards or penalties for specific actions. The goal is to learn a policy that will maximize the cumulative bonus over time. These algorithms are often used in artificial intelligence applications, such as self-driving cars or game-playing agents.In addition to these main categories, there are semi-supervised learning algorithms, which are trained on a mix of labeled and unlabeled data, and active learning algorithms, which can interact with their environment and choose which data to mark next. The choice of which machine learning algorithm to use depends on the specific problem, the data's characteristics, and available resources. (Bravo News, 2022), (Asmussen, 2019). Several machine learning algorithms can be used for text-to-sign language recognition, including:

Hidden Markov models (HMMs) (Assistant, personal communication, January 6, 2023): These are probabilistic models that are particularly well-suited for tasks involving sequential data, such as sign language gestures. They can capture the dependencies between different gestures in a sequence and predict the correct gesture at each time step. However, they can be sensitive to the choice of model parameters and may require a large amount of training data to achieve good performance. Conditional random fields (CRFs) (Assistant, personal communication, January 6, 2023): These graphical models are also well-suited for tasks involving sequential data. They can capture the dependencies between different gestures in a sequence and can be trained to incorporate additional context or information about the input text. However, they can be more computationally expensive and may require much training data to achieve good performance. Support vector machines (SVMs) (Assistant, personal communication, January 6, 2023): These are supervised learning algorithms that classify sign language gestures based on features extracted from the input data. They can effectively use relatively small datasets and work well with high-dimensional feature spaces. However, they may not be as well-suited for tasks involving sequential data because they do not explicitly model the dependencies between different gestures in a sequence.

Deep learning models (Assistant, personal communication, January 6, 2023): These are a type of neural network that can be used to learn complex patterns in data and can be particularly effective for tasks like text-to-sign language recognition. They can learn to extract relevant features from the input data and model the dependencies between different gestures in a sequence. However, they can require a lot of data and computational resources to train and may be more challenging to debug and interpret than other models. Once the text-to-sign language recognition has been performed, the recognized sign language gestures can be transmitted and displayed via animation using computer graphics techniques (Assistant, personal communication, January 6, 2023). It can be done using software programs designed to create animations or

general-purpose graphics software, such as Blender or Maya (Assistant, personal communication, January 6, 2023).

Natural language processing (NLP) is a subfield of artificial intelligence that focuses on enabling computers to understand, interpret, and generate human language (Jurafsky & Martin, 2019). NLP has many applications, including language translation, text classification, information extraction, and question-answering (Manning, Raghavan, & Schütze, 2019). Over the past decades, NLP has made significant progress thanks to advances in machine learning and computational linguistics (Jurafsky & Martin, 2019). Early NLP systems relied on rule-based approaches, which were programmed with specific rules and heuristics for analyzing and processing language (Manning et al., 2019). However, these systems needed to be improved in handling the complexity and variability of natural language and could not learn from data (Jurafsky & Martin, 2019).

With the advent of machine learning, NLP has made significant strides in improving the accuracy and robustness of language processing tasks (Manning et al., 2019). Machine learning algorithms, such as decision trees, support vector machines, and neural networks, have been successfully applied to various NLP tasks, including part-of-speech tagging, named entity recognition, and machine translation (Jurafsky & Martin, 2019). One up-and-coming area of NLP research is deep learning, which involves training neural networks with many layers of interconnected nodes on large datasets (Manning et al., 2019). Deep understanding has achieved state-of-the-art results on various NLP tasks, including language translation, language modeling, and text classification (Jurafsky & Martin, 2019). Despite these advances, such as handling rare and out-of-vocabulary words, the ambiguity and variability of natural language, and the lack of labeled data for many languages (Manning et al., 2019). Researchers are actively working on addressing these challenges to improve further the accuracy and efficiency of NLP systems (Jurafsky & Martin, 2019).

In conclusion, NLP has made significant progress over the past decade thanks to advances in machine learning and computational linguistics (Jurafsky & Martin, 2019). While there are still challenges to be addressed, NLP has the potential to revolutionize the way we interact with computers and make access to information more widely available (Manning et al., 2019). Natural language processing (NLP) is well-suited for a speech-to-text-to-sign language translation project because it enables computers to understand and generate human language (Jurafsky & Martin, 2019). NLP techniques, such as text classification and information extraction, can transcribe spoken language into text, which can then be translated into sign language using machine learning algorithms or other techniques (Manning et al., 2019). One of the main benefits of using NLP in this project is that it allows for a more natural and intuitive way of interacting with the translation system. Rather than having to input text or other data manually, users can speak into a microphone and receive a sign language translation in real time (Jurafsky & Martin, 2019). It can benefit deaf individuals who cannot communicate easily using written or typed language (Manning et al., 2019).

In addition, NLP can help improve the translations' accuracy and fluency by considering the spoken language's context and meaning (Jurafsky & Martin, 2019). It is especially essential to gauge translation, as a sign's meaning can depend on the context in which it is used and the signer's facial expressions and body language (Manning et al., 2019). By using NLP to analyze and understand the meaning and context of the spoken language, the translation system can generate more accurate and nuanced sign language translations (Jurafsky & Martin, 2019). Overall, using NLP in a speech-to-text-to-sign language translation project can help improve the accessibility and effectiveness of the translation system for deaf individuals and make it easier for them to communicate and access information. It is essential to carefully consider the data types of the features when building a machine learning model, as the choice of data type can affect the performance and accuracy of the model. (Antoniou, 2022).

There are several ways to classify machine learning, but one of the most common is based on the amount and type of input data used to train the model. Here are three main categories:

1. Supervised learning: This machine learning involves training a model on labeled data, where the correct output is provided for each example in the training set. The goal is to make predictions on new, unseen examples drawn from the same distribution as the training set. Examples of supervised learning tasks include regression (predicting a continuous value) and classification (indicating a categorical label).
2. Unsupervised learning: The model has not given any labeled training examples in unsupervised learning. Instead, it must discover patterns and relationships in the data independently. The Common unsupervised learning techniques include clustering (similar grouping examples) and dimensionality reduction (finding a lower-dimensional representation of the data).
3. Reinforcement learning: Machine learning involves an agent interacting with an environment and learning to maximize a reward by choosing the best actions. In this case, the agent receives feedback in the form of bonuses or penalties for its activities, and it learns through trial and error to choose actions that will lead to the highest overall reward.

Another way to classify machine learning algorithms is based on how they are trained. Some algorithms are "lazy" and do not start the training process until they are given new data to predict. These are called lazy learners. Other algorithms are "eager" and build a model using the training data and are called eager learners.

- Hidden Markov models (HMMs) are probabilistic models that are particularly well-suited for tasks involving sequential data, such as sign language gestures. They capture dependencies between different gestures in a sequence but can be sensitive to model parameters and may require a large amount of training data.
- Conditional random fields (CRFs) are graphical models well-suited for sequential data tasks. They capture dependencies between different gestures in a sequence and can incorporate additional context, but they can be computationally expensive and require a lot of training data.
- Support vector machines (SVMs) are supervised learning algorithms that classify data based on features extracted from the input data. They work well with small datasets and high-dimensional feature spaces, but they may be less effective for sequential data tasks.
- Deep learning models are neural networks that can learn complex patterns in data and are adequate for tasks like image and speech recognition. They can extract relevant features from the input data and model dependencies between items in a sequence. However, they require a lot of data and computational resources to train and may be challenging to debug and interpret.
- Natural Language Processing (NLP) is a field of artificial intelligence that deals with the interaction between computers and humans using natural language. It involves tasks such as text classification, language translation, and text generation and can be implemented using machine learning techniques, including HMMs, CRFs, SVMs, and deep learning.

Finally, machine learning algorithms can also be classified based on the type of output they produce. Some algorithms have a continuous output, while others make a discrete output. Algorithms that create a consistent result are called regression algorithms, while those that produce discrete work are called classification algorithms. Several machine learning algorithms could be used to translate speech-to-text to sign language (Jurafsky & Martin, 2019). One possibility is to use a sequence-to-sequence model, a deep-learning neural network designed to map input data to a sequence (Vinyals & Le, 2014). This model could be trained on a large dataset of audio recordings of spoken language and the corresponding sign language translations and then be used to generate sign language translations for new audio input (Jurafsky & Martin, 2019). Another option is to use a combination of machine learning algorithms, such as a classifier to transcribe the spoken language into text and a separate model to translate the text into sign language (Jurafsky & Martin, 2019). This approach would require different training models for each task but may be more effective for handling language translation's complex and nuanced nature (Manning et al., 2019). Ultimately,

the best machine learning approach for speech-to-text-to-sign language translation will depend on the specific requirements and constraints of the application, as well as the availability of data and resources (Jurafsky & Martin, 2019). It may be necessary to experiment with different algorithms and approaches to find the most effective solution (Manning et al., 2019). Cloud computing is one of the environments covered in this project. In recent years, most companies have been aiming to get resources from a trust sourced with high performance, low cost, and avoiding maintenance and security issues, which encourage the stakeholder to migrate to cloud computing, which has trustworthy.Cloud computing is a comprehensive model for sharing services across the network that is provisioning to the user with minimal management interaction needs. In addition, it has:

Fundamental Characteristics Model of a Cloud

- Self-service on demand, the user can have the Service as wish in pay-and-go and without any delays.
- Broad Network-Access: The Service and resources are always available everywhere, which can be provided quickly for users.
- Rapid elasticity: The user can quickly scale the resources without management interaction.
- Measured Service: the cloud aspect should monitor, optimized, well-controlled, and reported to the vendor and user.
- Multi-Tenacity is a solid policy that gives each user its segment, isolation, service level, and governance.

Below are the types of Deployment models of cloud computing:

- Public Cloud: The service owner will provide the benefits over the internet based on defined rules, policies, and cost modes.
- Private Cloud: Designed and prepared for the internal benefits of the consumer or organization; therefore, having this setup will decrease security. Due to the Private cloud, the management will manage the paid cost of the system.
- Community cloud: cloud based on community requirements, policies, and the number of organizations sharing the cloud with their consumers. The 3rd SP will manage the infrastructure, reducing costs by dividing the consumers with high security. (Dillon, Chen, & Chang, 2010).
- Hybrid cloud combines public and county cloud computing with different modes, keeping the same constative-infrastructure and creating a function to give corresponding allowing operate applications and data.

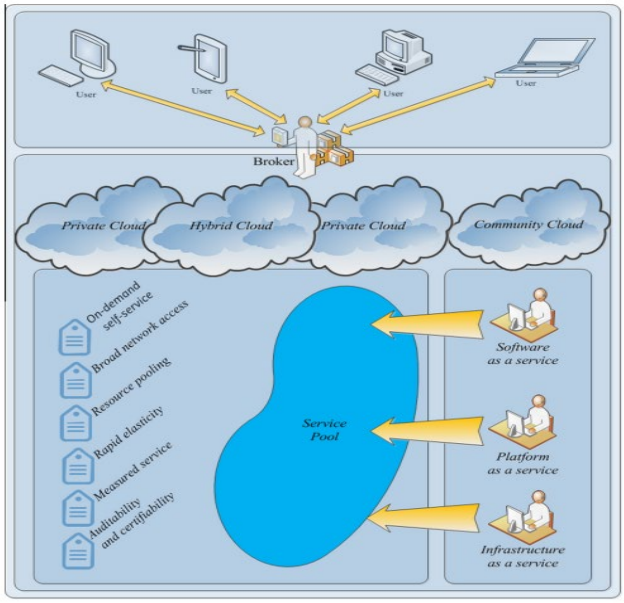


Figure 1. Cloud architecture Chhaya et al., 2020)

Cloud Model

- Service is a method that will obtain the capability to give a function used to compile with corresponding restrictions and rules on the interface. (Ellinger, 2013)
- Platform: A layer that hosts the hardware and application and will have a user interface that the user can execute and deploy.
- Infrastructure is the layer in which the physical device and components are underlying it to perform the functions. These devices and components can have Storage, processors, Networks, DB system management, and OS.

Several types of cloud computing could be suitable for storing a database for a speech-to-text-to-sign language translation project. One possibility is to use a database-as-a-service (DBaaS) cloud, which provides a fully managed database service that handles tasks such as provisioning, scaling, and backup (Mell & Grance, 2011). DBaaS clouds can benefit projects requiring high reliability and uptime, offering features such as automatic failover and data replication (Mell & Grance, 2011). Another option is to use an infrastructure-as-a-service (IaaS) cloud, which provides computing resources, such as storage, networking, and computing power, on a pay-per-use basis (Mell & Grance, 2011). IaaS clouds can be beneficial for projects that require a high level of customization and control, as they allow developers to install and configure their database software on the cloud infrastructure (Mell & Grance, 2011).Ultimately, the most suitable type of cloud computing for storing a database for a speech-to-text-to-sign language translation project will depend on the specific requirements and constraints of the application, as well as the availability of data and resources (Mell & Grance, 2011). Evaluating various cloud computing options may be necessary to find the most effective solution.

Voice/ Text recognition: This part is also covering the second domain of this project. Recently, multiple developments in applications to improve the interaction between people using gestures, text, and human-communication sign language. Moreover, the variable is one of the leading mediums used in talking

and interacting with machines and found as limited between them. Speech recognition setup is converting from speech to text to improve the interface between people and devices. Speech recognition depends upon the objective of recognizing the spleen, like "accents." Any element should not interpret the voice source to learn the words with a clear voice. Therefore, most recognition development recently aimed to improve the tools and make them independent.(Kinkiri et al., 2017)

The following are chrematistics aspects of sound:

1)The frequency of wave sound is tested as tone, which counts the number of vibrations in hertz.2) Determines the volume of the sound on how loud or soft.3) The tone is a particular pattern for the device generating the sound, meaning each device has its particular tone, which is available with the fundamental frequency.(Kinkiri et al., 2017).

In regards of voice control implementation, it divides into the following:

- Speech
- Recognition
- Translation
- Execution of the commands.

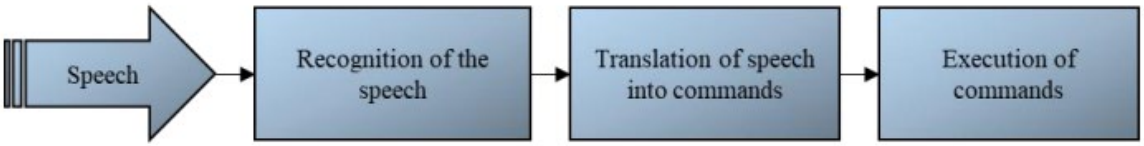


Figure 2. Voice recognition module (Peter et al., 2020)

In addition, the Voice System has two types:

- Open-source code has two algorithms:
  - The first is: The processing will happen on the local device.
  - The second is: It sends it to the remote server and processes it. (Google, Siri.

Closed System-Open-Source Code

Limited access to program source code/software is only a binary version to avoid modification. In addition, a granted access to source code while singing.

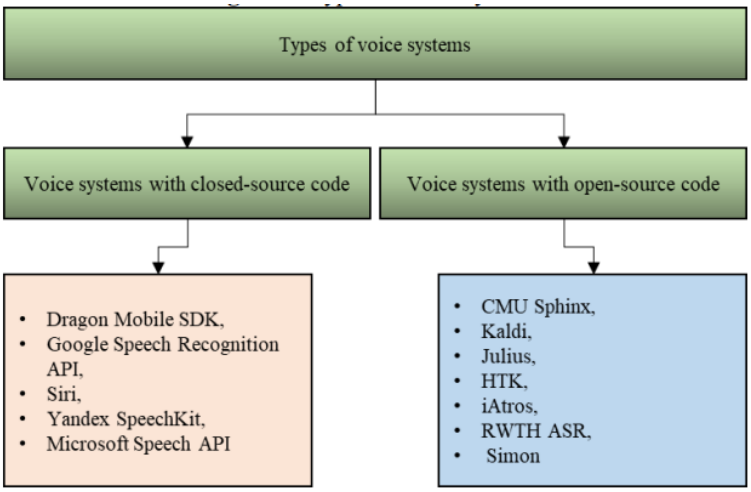


Figure 3. Types of voice systems (Peter et al., 2020)

Several voice and text recognition technologies could be suitable for a speech-to-text-to-sign language translation project. One possibility is to use an automatic speech recognition (ASR) system, a type of software that can transcribe spoken language into text (Bourlard & Huang, 2012). ASR systems use machine learning algorithms and acoustic models trained on significant audio recordings and transcriptions datasets to recognize and transcribe spoken language (Bourlard & Huang, 2012). Another option is to use a natural language processing (NLP) system, a type of software that can analyze and understand the meaning and context of human language (Manning et al., 2019). NLP systems can transcribe spoken language into text and perform tasks such as named entity recognition and text classification (Manning et al., 2019).

Ultimately, the most suitable voice and text recognition technology for a speech-to-text-to-sign language translation project will depend on the specific requirements and constraints of the application, as well as the availability of data and resources (Bourlard & Huang, 2012; Manning et al., 2019). It may be necessary to evaluate a variety of voice and text recognition technologies to find the most effective solution. Regarding problems, they are multiple issues were highlighted in chapter one of the overview, and this part will explain the difficulties which the researchers face and how they proposed a solution or workaround:

- Improper solutions not implemented by the educational management.
- Technologies at schools and colleges are implemented but not ready to add an intelligent answer.
- Lack of awareness and knowledge to deal with new technologies.

In addition, there are many problems were observed in the literature review:

- Problems: identify the gaps found in South Africa during the COVID-19 pandemic using the AI & ML methods in speech-to-text to Sing language for the real-time translation provided for hard-hearing and deaf ones. (Peter et al., 2020)

  And the research highlighted that 11 South African languages are officially announced, and more are yet to be t formally announced. Moreover, the accent of Africans is challenging to understand. And it was observed by the researchers that the translation processing has noise and eco and unfiltered quality. Involve the concept of AI and ML integrated with speech-to-txt to sign language recognition to learn the 11 Official languages in South Africa. s (Lanham, 1996). chat Applications installed on deaf devices like deaf chat researched by: Shezi and Ade-Ibijola (2020). Sign real-time Language Avatar, used for end application to display the signs for deaf ones. (Ezhumalai et al., 2021; Harkude et al., 2020; Papastratis et al., 2021; Shinde & Dandona 2020). Overcome noise,

echo, and unfiltered quality due to translation processing by applying speaker recognition, which uses MFCC and CORDIC algorithms. The below flow:

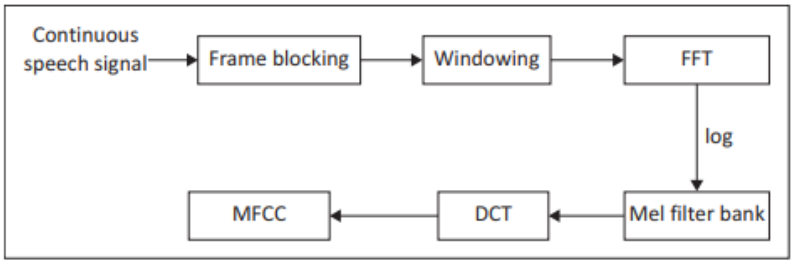


Figure 4. communication process (Peter et al., 2020)

As Shown above, this is the flow of extracting of MFCC feature (Mel-frequency Cepstral Coefficients); MAFCC is the natural prescription of the human voice ear based. In addition, it makes them fit and well used for speech recognition. The above flows show that the frame passes a hamming window before the processing, then distributing.

Methodology

Based on the project objective and the requirements, Agile approaches place a strong emphasis on flexibility and adaptability, which could be particularly important given the inclusion of AI, ML, and cloud technologies, which may require the project to adapt and respond to changes in requirements and priorities as the project progresses. Agile methodologies also prioritize collaboration and transparency, which could be necessary given the need for close communication and cooperation with stakeholders.
An agile project plan typically consists of the following elements:

1. The scope includes the project's goals, objectives, and deliverables.
2. Sprints: The project is divided into smaller chunks of work called "sprints," typically two to four weeks.
3. Prioritized work: The work is prioritized based on importance and dependencies.
4. Schedule: A schedule created that outlines each sprint's start and end dates and the specific tasks that need to be completed during each sprint.
5. Tasks: Tasks are assigned to team members and tracked using a project management tool.
6. Progress monitoring: The progress of the work is regularly monitored and tracked.
7. Review and adjustment: At the end of each sprint, a review of completed work and apply necessary changes only.

The WBS, Gantt chart, and network diagram are mandatory to create them regarding the scope of work, resources, deadlines, and requirements as shown below:

A work breakdown structure: A method used for project management to decompose a project into smaller, more manageable components. It is a hierarchical representation of the work that needs to be done to complete a task. To define the project's scope and help plan, execute, and control the position, which position supports the project to be more managed efficiently. Below is a draft design which might be restructured and redesigned slightly based on the best practice in terms of delivering the project properly:

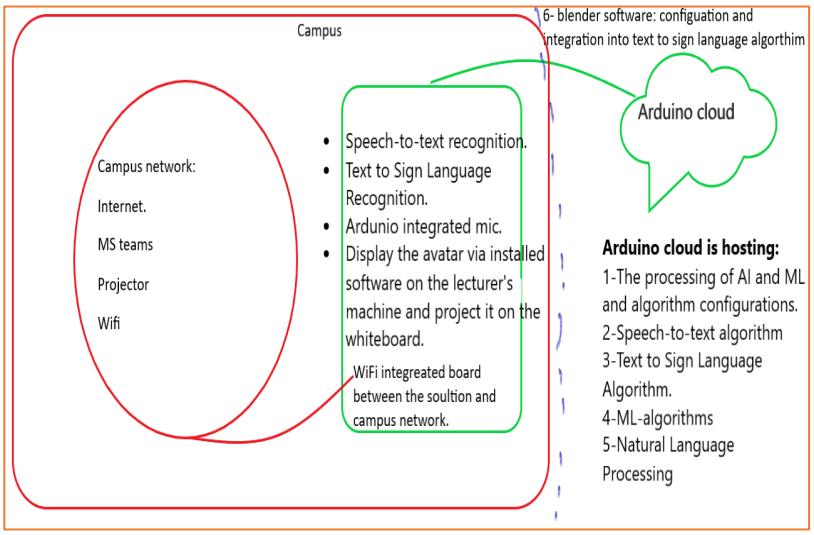


Figure 5. Break down structure.

In the general workflow of speech-to-text recognition projects with animation to display sign language:

1. Gather a dataset of speech audio and the corresponding text and sign language gestures: It will need a dataset of speech audio and the corresponding text and sign language gestures to train the machine learning model. It can create this dataset manually or find a pre-existing dataset online.
2. Preprocess the data: It will need to preprocess the data to extract relevant features from the speech audio and sign language gestures.
    - It may involve converting the audio to a numerical representation (e.g., using a spectrogram or Mel-frequency cepstral coefficients (MFCCs))
    - converting the text to a numerical model (e.g., using word embeddings),
    - and extracting relevant features from the sign language gestures (e.g., using computer vision techniques).
3. Train a machine learning model: Once it has preprocessed the data, it can train a machine learning model using the preprocessed data. It can use various algorithms for this task, including HMMs, CRFs, SVMs, and deep learning models.
4. Test the model: After training the model, it will need to test its performance on a separate dataset to see how well it generalizes to unseen data. Then it can use metrics such as accuracy, precision, and recall to evaluate the model's performance.
5. Implement the animation: Once it is trained and tested by a machine learning model, it performs speech-to-text recognition in real time. Can then use the recognized text to trigger the display of the corresponding sign language gestures using computer graphics techniques.

This process allows deaf individuals to communicate using speech, which is then converted to text and processed into sign language that can be understood by the avatar and displayed to the user. Integrating these technologies makes it possible to create a system that enables deaf individuals to communicate more effectively with hearing individuals and other deaf individuals.

The specific details of this process may vary depending on the technologies and approaches used, but this is a general outline of the project flow.

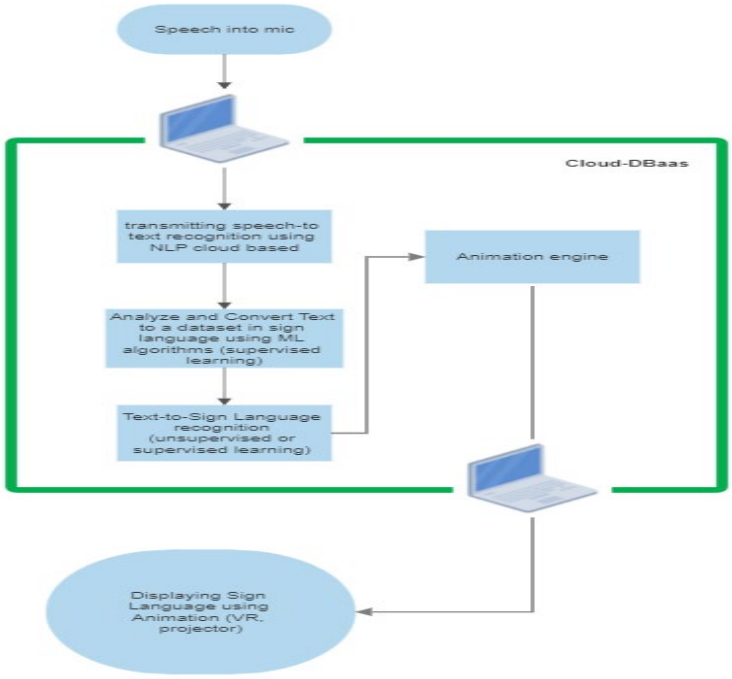


Figure 1. Workflow diagram

Design Model

The physical design is a critical aspect of the project, and it involves setting up and highlighting all the components required to achieve the solution. The design solution is in physical form and will involve hardware, software, and other elements required for integrating speech-text recognition with animation for displaying the signs for Deaf individuals in the education environment. This includes a speech-input device, cloud-computing platform, text-animation output device, API, and database for storing and managing data. Additionally, the solution will be integrated into new or existing networks, which may include access points, switches, routers, and firewalls. The data flow on the network backbone involves the teacher's API collecting speech data, which is processed over Wi-Fi and flows through the network based on its configuration. The data then goes out to the internet and reaches the cloud service provider, where it is received, stored, and processed using NLP and supervised ML algorithms. The text is then sent to the animation engine, where it is transformed into sign language and projected on the student's API.

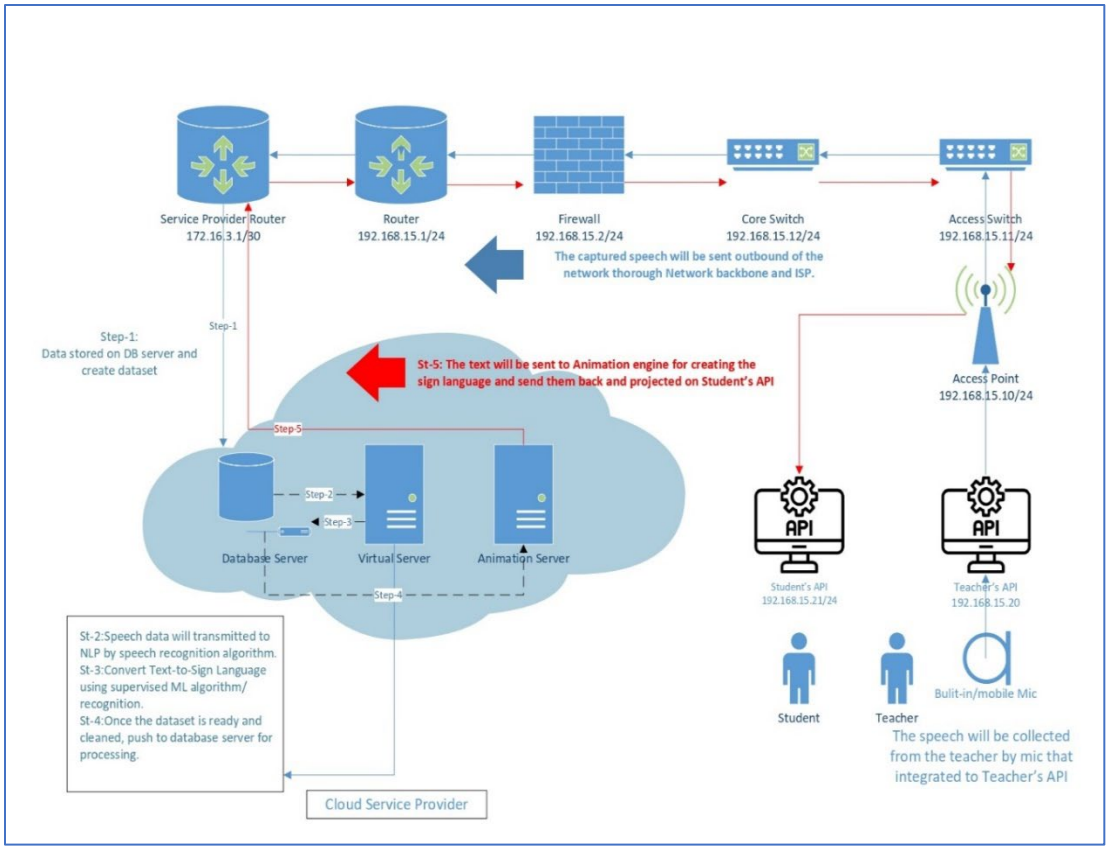
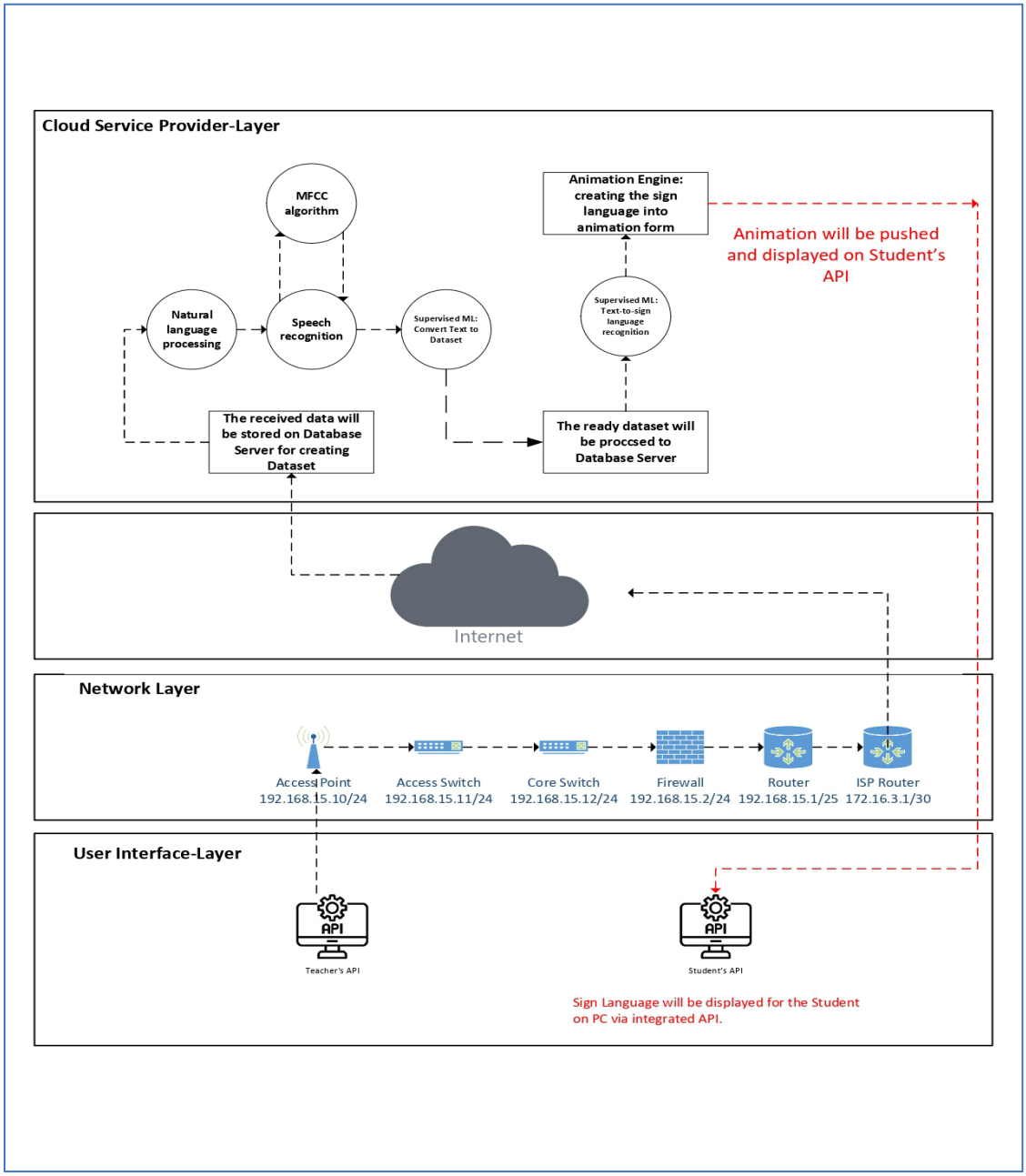


Figure 7. Design flow

The logical design of the proposed project involves the conceptual and functional organization of the speech-to-text and animation generation components using a layered architecture. The layered approach divides the system into functional layers that communicate only with the layer above and below, creating a clear separation of concerns and improving overall modularity, scalability, and maintainability. The proposed logical design is structured as follows: The User Interface layer allows the teacher to access the integrated API and collect speech through the mobile or built-in mic, while the student's API displays the sign language on a smart device or laptop. The Network layer allows data to flow across the network until it reaches the ISP router, and from there, it goes to the Internet layer, where it reaches the Cloud Service Provider layer. At this layer, the data is received, stored, and processed into NLP, utilizing ML Speech recognition algorithms fetched from the library at the DB server. The data is then converted using supervised ML algorithms for text-to-dataset and text-to-sign language recognition. Once the dataset is ready, it is stored on the database server, and the Animation engine creates the signs in real time, which are then pushed to the Student's API to display the sign language. The production costs are anticipated to be $2 per kilogram of green hydrogen (based on current US prices). In India, brown hydrogen costs $2 per kg and is created from cheap coal. Depending on supply, the cost of green hydrogen produced through electrolysis using renewable energy ranges from US$10 to US$15 per kg. Hydrogen may substitute for and decarbonize natural gas in all applications, beginning right away at up to a 20% blend and increasing that amount as pipes are updated. While one of California's top utilities plans to blend 20% of hydrogen into its natural gas power plants, other large industrial companies in Japan are already blending up to 90% hydrogen into their natural gas units that are currently in operation.

- Just 20% hydrogen added to natural gas pipelines in the UK would eliminate 2.5 million cars' worth of carbon dioxide yearly, or 6 million tons.
- A transition approach for decarbonizing natural gas grid assets that doesn't require significant infrastructure expenditures is hydrogen blending.

Table 1. (SG H2 ENERGY, 2023)



References

Papastratis I, Chatzikonstantinou C, […] Daras PSensors (2021) 21(17)

Han W, Chan C, […] Pun KProceedings - IEEE International Symposium on Circuits and Systems (2006).

Matarneh R, Maksymova S, […] Belova NIOSR Journal of Computer Engineering (IOSR-JCE) (2017) 19(5).

Peter V, Paken J, Joseph Ljournals.co.za (2020) 67(1).

Kinkiri S, Melis W, Keates S The Second Medway Engineering Conference on Systems: Efficiency, Sustainability, and M. Chhaya K, Khanzode A, Sarode RInternational Journal of Library & Information Science (IJLIS) (2020) 9(1).

Dignum VLondon Review of Education (2021) 19(1).

Chakladar D, Kumar P, […] Kim BApplied Sciences (Switzerland) (2021) 11(8).

Brvl News. (2022, December 20). "What is AI and How Does it Work?" Brvl News.

Asmussen, C. B. (2019, October 19). Smart literature review: a practical topic modeling approach to exploratory literature review - Journal of Big Data. SpringerOpen.

Antoniou, N. (2022, December 15). Alternating Objectives Generates Stronger PGD-Based Adversarial Attacks. arXiv.org.

Chapter 2 Flashcards by Andrea DiTullio | Brainscape. (n.d.).

Precisely, Inc. (2021, October 20). Modern Data Architecture with Legacy Data - How to Build One. Precisely

Lish, T. (n.d.). What Can Affect A Sensor's Accuracy

Gupta, S. (2022, December 7). AI vs. Machine Learning vs. Deep Learning: What's the Difference? Springboard Blog.

Jurafsky, D., & Martin, J. H. (2019). Speech and language processing (3rd ed.). Pearson.

Manning, C. D., Raghavan, P., & Schütze, H. (2019). Introduction to information retrieval (3rd ed.). Cambridge University Press.

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. In Advances in neural information processing systems (pp. 3104-3112).

Mell, P., & Grance, T. (2011). The NIST definition of cloud computing. National Institute of Standards and Technology.

Bourlard, H., & Huang, X. (2012). Connectionist speech recognition: A comprehensive foundation. Cambridge University Press.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.
Mitchell, T. M. (1997). Machine learning. McGraw-Hill.

Ozturk, O., & Kuntman, H. (2010). Field-programmable gate array (FPGA)-based designs for real-time image and video processing. In Field-Programmable Gate Array (FPGA) Based System Design (pp. 3-26). Springer, Boston, MA.

Bryson, J. (2020). Social and ethical considerations in technology development and deployment. In The Cambridge Handbook of Human-Computer Interaction (pp. 89–107). Cambridge University Press.

DevTeam.Space. (n.d.). What is an Agile Methodology? DevTeam.Space Blog.

Agile Alliance. (n.d.). What is Agile?