RESEARCH ARTICLE OPEN ACCESS

A Text to Speech and Speech to Text Application for Students with Hearing and Speaking Impairments

Olubukola D. Adekola*, Esther O. Adeoye**, Oyebola Akande***, Solomon S. Akpan****
* Department of Software Engineering, Babcock University, Ilisan-Remo, Ogun State, Nigeria
Email: adekolao@babcock.edu.ng
, *, ****Department of Computer Science, Babcock University, Ilisan-Remo, Ogun State, Nigeria
estheradeoye14@gmail.com, akandeo@babcock.edu.ng, adexbukene@gmail.com

----------**********----------

Abstract:

Hearing impairment (HI) is a defect in which a person cannot hear properly or can only hear partially and it is usually measured using the hearing threshold which has a unit in decibels (dB). Deafness can be caused by accidents, old age, or sicknesses. Speech impairment is a defect in which a person cannot speak properly or has deficiency in the power of speech. Many text to speech (TTS) and speech to text (STT) applications have been made to make communication easier but hardy with a feature to record previous conversations. This study employed iterative development model to develop a text to speech and speech to text mobile application with the capability to track and record previous conversations which would yield benefits as progress in the academic and social life of hearing and speech impaired students. Analysis of the existing systems was done and the proposed system was designed using unified modeling language (UML) tools and others such as Ionic Framework, Angular JS, Hypertext Markup Language (HTML), Cascading Style Sheets (CSS) and Firebase.

Keywords — Hearing impairment, Speech impairment, Speech recognition, Speech synthesis.

----------**********----------

I. INTRODUCTION

Deafness is the inability of an individual or person to hear due to the circumstance of birth, due to external factors after birth or due to old age. Being dumb is characterized by the inability to speak clearly or to be completely mute. Students with hearing and speaking impairment (SI) suffer sufficient communication. Parents who are not aware of schools for special students end up keeping their child away from the society and denying the child the right to education only for the misconception that educating the hearing impaired is a sheer waste. This makes the child feel inferior to the hearing child because of his/her disability. Some other parents who know about these schools do not have the funds to cater for the needs of their children. Furthermore, many of these special schools are not well funded by the government and the teachers do not have degrees or proper knowledge of the techniques needed to teach the students [1]. Deafness or Hearing loss (HL) can be explained as a condition when the hearing threshold is below normal for understanding of speech [2]. The unit of measurement of the hearing threshold is in decibels (dB). The normal hearing threshold for a human adult is 0dB (decibel) - 25dB. For children, it is 0dB - 15dB [3]. Hearing loss can be classified into the following categories [4], [5]:

(a) Mild Hearing Loss (26 – 40dB), (b) Moderate Hearing Loss (41-55dB), (c) Moderately Severe Hearing Loss (56 – 70dB), (d) Severe Hearing loss (71 – 90dB) and (e) Profound Hearing Loss (90dB - above). Countries with higher numbers of people

with HL are China (Hong-Kong), Korea, Indonesia, Malaysia, Philippines, Afghanistan, India, Poland, Congo, Kenya, South Africa, Nigeria, Togo, Mexico, Brazil and New Zealand [5].

According to [6], deaf people are categorized into three:

a) Those born or became deaf before language development (Pre-Lingual Deafness)
b) Those who became deaf after language development (Post-Lingual Deafness)
c) Those who became deaf in old age (Presbycusis)

Speech and language are not hereditary and are most times learnt from observation and hearing. The way a person can learn Korean from watching and listening to Korean Movies is the same way a baby learns to speak a language by listening to his/her parents. The condition where a person has the disability of speaking and hearing (Deaf Mutism) is caused by Pre-Lingual deafness due to the fact that the child has probably never heard a word being said and so, the child finds it hard to produce words and construct speech [7]. This study analysed various mobile technologies available to provide help for hearing impaired students in classroom environments. Many text to speech (TTS) and speech to text (STT) applications have been made to make communication easier but hardy with a feature to record previous conversations. The relevance of implementing on this capability is to further learning and to make students with these challenges cope better. This study employed iterative development model to develop a text to speech and speech to text mobile application with the capability to track and record previous conversations which would yield benefits such as progress in the academic and social life of hearing and speech impaired students.

II. PROBLEM STATEMENT

Findings reveal that many existing text to speech (TTS), speech to text (STT) applications and various other technologies apart from hearing aids that help hearing and speaking impaired people to communicate and take part in activities that require the use of the ears for hearing and the mouth for speaking do not have the feature to record previous conversations which could positively influence necessary revision and learning. Therefore, this study aims to develop a mobile application for Text to Speech and Speech to Text translation with a feature to save previous conversations to help students with hearing and speaking impairments communicate better in classrooms and in school environments. This development would leverage on the proliferation of smart phones in the present technological age. Thus, majority have the propensity of gaining easy and direct access to TTS and STT applications which would be available via Google Play store and Apple App store.

III. LITERATURE REVIEW

According to the World Health Organization [8] statistics, there are more adults than children suffering from hearing loss. The prevalence being highest in Southern Asia, Asian Pacific and sub-Saharan Africa. This section presents a review on hearing impairment and associated factors.

A. *Forms of Deafness*

The following are remarkable forms of deafness as found in the literature:

1) Pre-Lingual Deafness (PLD)

A baby in the womb already has the ability to hear. A child suffering from hearing loss in the womb or after being born i.e. before speech formation would not be able to speak because speech is acquired not inherited. This hearing loss that occurs before the formation of speech is known as Pre-Lingual Deafness. Causes of pre-lingual deafness can be classified into congenital causes (while the baby is in the womb until birth) and post congenital causes (after birth). The congenital causes are those that arise as a result of certain complications during the gestation period or at the birth of the baby. Some of these congenital causes can be classified as either syndromic or non-Syndromic. The syndromic are the ones with signs and symptoms such as loss of pigmentation in the skin, the eyes and the hair. The non-syndromic causes of deafness are infections like the Toxoplasmosis, Rubella, Cytomegalovirus, Herpes Virus and Syphilis

(TORCHS) Group of infections [9],[10]. For post congenital causes which is after birth, most of the aforementioned syndromic and non-Syndromic causes can still affect the child. Meningitis may also affect a child before the formation of speech. Chemotherapeutic Radiation (cancerous treatment radiation) may also cause hearing loss.

2) Post-Lingual Deafness (PoLD)

This form of deafness comes after speech formation. In such cases, the person can speak properly and may at times read the lips of the speaker to understand what is being said. PoLD may be caused by Noise (Noise Induced Hearing Loss) or infectious diseases like Jaundice, Measles etc. Noise Induced HL can be categorized into (a) Long Term Noise Exposure and (b) Short Term Noise Exposure. The long term noise exposure can come in form of exposure to noisy environments like factories, high volumes of music on headphones and earphones. The short term noise exposure can come from explosions, gun shots or single exposure to loud noises in the environment which may cause acoustic trauma. HL from short term noise exposure cannot be cured but technologies like hearing aids and cochlea implants may be used to help the person to hear a little better.

3) Presbyacusis

Presbyacusis is a hearing disorder that comes as a result of old age. People can start having a form of Hearing loss from 45 years and above. The main cause of Presbyacusis is unknown but it can be caused by problems with the nerve cells in the ear. With time, the nerve cells depreciate and stop functioning as they used to due to the aging process. Smoking over time may also cause Presbyacusis, exposure to some medications and harmful chemicals over the years [11], [12] .

B. Deaf Technologies

Over the years, a lot of studies have been in line with deafness and ways to improve hearing for people suffering with deafness and also to help them to communicate better in the hearing world. Different technologies like hearing aids, Cochlea Implants, Sign Language Interpreters, TTS and STT applications have been created to bridge the hearing gap to different extents.

1) Hearing Aids

A hearing aid is an electronic device usually worn in or on the ear of a hearing-impaired person for amplifying sound. Hearing aids can be analog or digital. The hearing aids are made up of 4 components namely:

a) The Microphone which is responsible for picking sounds from the environment
b) The Amplifier circuitry which is used to amplify the sound that has been received by the microphone.
c) The Receiver or loud speaker which sends the sound signals into the ear canal
d) The Battery which powers the device

Analog Hearing aids are created to amplify all the sound waves in the environment as they are. The hearing aids can be adjusted based on the environment whether it is a classroom environment, or a worship environment or a loud environment like a football field. This is possible due to the fact that these settings have been pre-preprogrammed into the hearing aids and so, when in that environment, the user just needs to press a button and the hearing aids will adjust to the environment. The problem about analog hearing aids is that it does not differentiate the noise from the actual sound. It picks up all sounds. For this reason, the Digital Hearing Aids were created. [13], [14].

Digital Hearing Aids convert sound waves in the environment to digital signals but it also reduces the surrounding noise that might not be relevant to the person with the hearing loss. It exactly clones the sound from the environment and sends this sound to the ear which then does the rest. Also, the aids can be adjusted to the environment and the sound transmission can be fine-tuned for different levels of hearing loss.

Hearing aids may also be classified by their design as follows:

a) Behind-the-Ear (BTE) Hearing Aids
b) Mini BTE Hearing Aids
c) In-The-Ear (ITE) Hearing Aids
d) In-The-Canal (ITC) Hearing Aids
e) Completely-in-Canal (CIC) Hearing Aids

Some hearing aids like Directional Microphones and T-Coil (Telephone switch) have also been created to add special features that normal hearing aids may not be able to perform. Some of them are;

2) Cochlea Implants

The human cochlea is the main organ of hearing. It is a snail-shell structured organ in the ear that has various hairs called cilia and a fluid which gets agitated and moves around when it catches the vibrations from the eardrum (Tympanic Membrane). These vibrations are then sent through the nerve endings to the brain for interpretation. When the cochlea loses some of these cilia or gets damaged, the ear would not be able to transmit the sound signals to the brain and so, the person would not be able to hear. To solve this problem, a small electronic device called the cochlea implant was created to do the job of the cochlea. It consists of an auditory (acoustic) component that works like a hearing aid to amplify the surrounding sounds and sends it to the hearing canal. It also has a processor which converts the sound signals to digital signals which is then sent to an implant underneath the skin. The digital signal is sent from the implant to the cochlea via the electrode that passes through the cochlea. This electrode then simulates the nerve fibers and then the brain processes the sound. Cochlea implants are one of the very effective technologies that have been created to bridge the hearing gap [15], [16] .

3) Digital Sign Language Interpreters

Sign Language is a language that uses hand gestures and signs to portray information; it is widely used by the deaf as a form of communication. Sign languages may differ based on location but the most widely used sign language is the American Sign Language (ASL). Digital Sign Language interpreters may be applications or devices that can translate the sign Language to English (or any other language) or translate a spoken Language to Sign Language (Sign Language Recognition). This technology is implemented using motion sensors, motion mapping from videos etc. Challenges with using sign language interpreters might be the inability of the device or application to capture the exact facial expressions that go with the hand gestures and thus, make the translations ambiguous [6].

C. Text-To-Speech (TTS) and Speech-To-Text (STT) Applications.

This technology is useful in helping people with hearing and speaking impairments to read from the environment and speak back to the environment. Simple speech to text applications can be used when the user is only hearing impaired and can speak while text to speech is good for people who only have speech loss but can hear. But in situations where the user is hearing and speaking impaired, only TTS or only STT applications may not work. What can handle both TTS and STT is the main goal of this study. A number of the existing applications are stated as follows:

(a) The "Smart Voice Assistant" Application: This application was created to help users to translate speech commands to actual application actions and to text [17]. This STT application might not be suitable for people with both hearing and speaking impairments.

(b) The "Deaf Communicator" Application: This application was created to help deaf people to communicate without sign language. It is of dual functionality i.e. can convert text to speech and speech to text. This application only supports English and Indonesian Languages.

(c) The "Deaf Assistant" Application: This is a dual functionality application created to convert the speech of the hearing person to text to be read by the person with the hearing loss and to capture text from the hearing-Impaired person and translate to speech to be understood by the hearing person. It supports a number of languages. Just like the deaf communicator, it also has a slow translation time and the application has to be kept close to the speaker to capture the words properly.

In summary, the proposed application named ***Comuno app*** coined from the word "Communication" is expected to have additional feature that would allow user to refer back to previous conversations and save conversations.

IV. MATERIALS AND METHODS

These studies embody case studies, systematic literature reviews on deaf learning technologies, challenges of students with hearing loss (HL) in hearing environments, current mobile applications for speech synthesis and speech recognition. The

relevant documents obtained were qualitatively analyzed for convergence, and relevant details were extracted using inductive approach. This study employed iterative development model to ensure requirements are well captured and duly refined through iterations in order to produce the desired system [18]. The proposed system was designed using unified modeling language (UML) tools and others such as Ionic Framework, Angular JS, Hypertext Markup Language (HTML), Cascading Style Sheets (CSS) and Firebase.

A. Functional Requirements of Comuno app

a) The user should be able to type text to be translated by the system
b) The user should be able to click to project the text to speech signals
c) The user should be able to click to listen for speech
d) The user should be able to view translated speech
e) The user should be able to view previous conversations

B. System Analysis and Design of Comuno app

(a) **Use Case:** The key actors in the system are the user (student) and the Comuno App. Figure 1 represents the Comuno App use case diagram revealing the interactions between the user and the system.

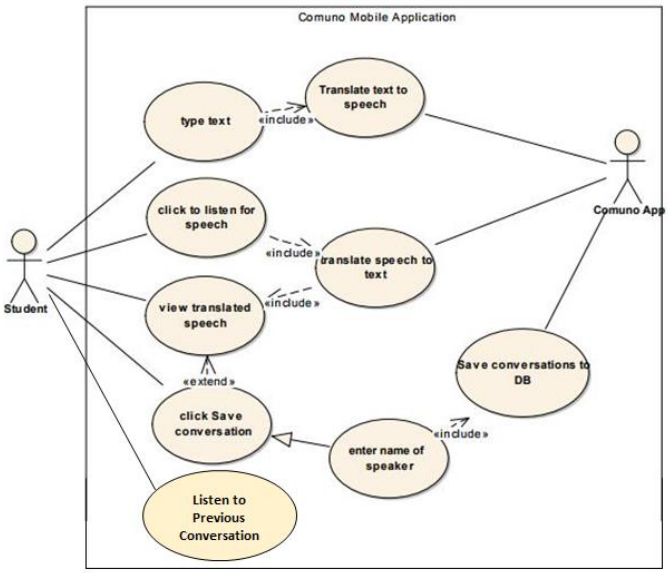

Figure 1: Comuno App use case diagram

(b) Scenario Analysis: This is a design tool used to analyze a working scenario of the system under a single use case. The use case analyzed here with scenario analysis tool is the “Listen for Speech”. The following Table 1 shows the Comuno App Scenario Analysis.

TABLE I
COMUNO APP SCENARIO ANALYSIS

Use Case Name	Listen for Speech	
Primary Actor	User	
Summary	The User can click on a button to enable the device to capture voice signals from the environment and the system would translate these voice signals into text for the user to read.	
Pre-condition(s)	1. The Comuno Application has to be installed on the device 2. The device should have native speech recognition functionality	
Flow of Events	User	Comuno App
	1. Click on the listen button. 6. User Views the translated speech	2. App captures the voice signals. 3. App processes Voice 4. App translates the voice to text 5. App displays the translated speech as text
Extensions	System takes a long time to translate speech	
Post-condition(s)	User reads text and provide feedback where necessary	

(c) **UML Class Diagram:** The following UML class diagram shows the static relationship between the following Typescript class components of the system design:

1) The HomePage Class: Contains one attribute and four operations.
2) The ListenPage Class: This is responsible for the speech recognition functionality.
3) The SpeakPage Class: This is responsible for the speech synthesis functionality.
4) The ConversationPage Class: Which is responsible for displaying previously saved data.

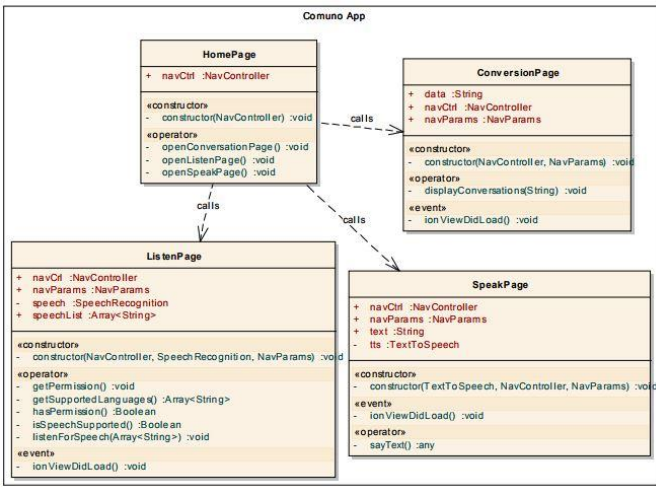

Figure 2: Comuno App Class Diagram

(d) **UML State Machine Diagram**: This diagram is to complement the Class diagram by showing the dynamic view of the various states the system would be in a single process of operation. Figure 3 reveals Idle state, Listening-for-speech state, Capturing-Text-from-the-keyboard state, Translating-Speech-to-text state and Translating-text-to-speech state.

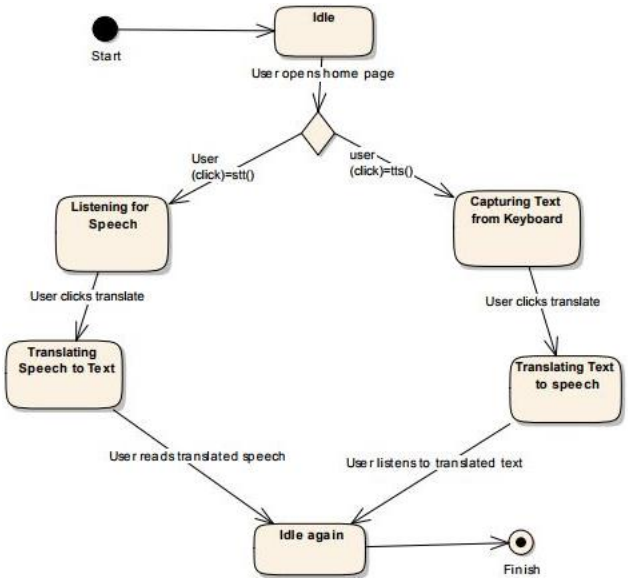

Figure 3: Comuno App State Diagram

(e) **Comuno App Sequence Diagram**: This shows the sequence of actions and events that occur in the Comuno App. Figure 4 is the Comuno App Sequence Diagram which efficiently modeled the sequence of operation within the system.

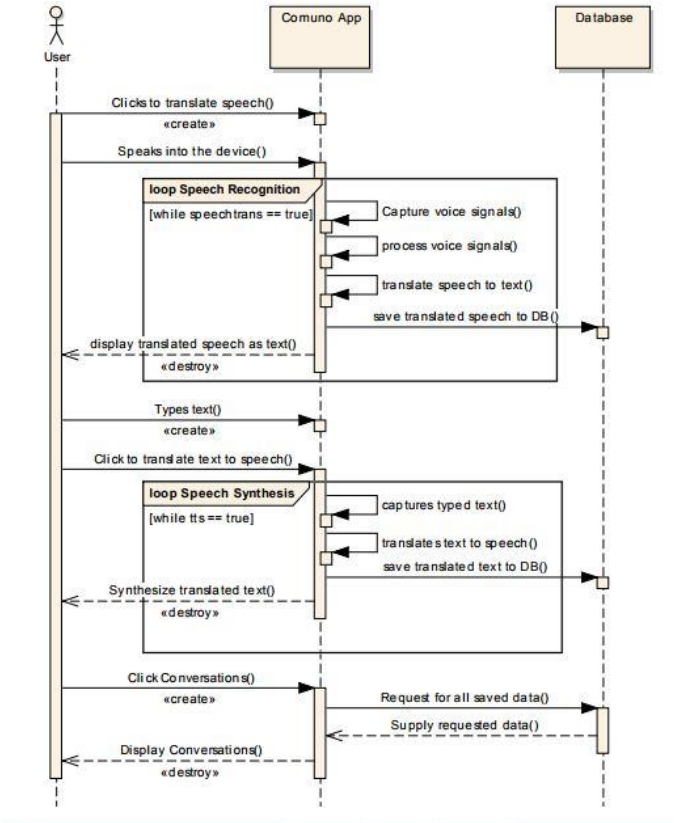

Figure 4: Comuno App Sequence Diagram

V. SYSTEM DEVELOPMENT

Visual Studio Code, a text editor from Microsoft Corporation was used to implement the design. Node JS, a Server-Side JavaScript runtime Environment was also used for executing JavaScript and other codes written in JavaScript Frameworks. Google Firebase Cloud database management System was introduced to make the app have real time data storage and easy authentication. Firebase is a NoSQL Database which uses an Object-oriented method of data entry and retrieval. Language tools and frameworks employed include HTML (Hyper Text Markup Language) with CSS, SASS (Syntactically Awesome Style Sheets), TypeScript (an open source programming language managed by the Microsoft Corporation), Angular (an open source scripting language used for building cross platform/cross-browser web applications) and Ionic framework (an open source software development kit used for developing hybrid multi-platform mobile applications).

A. The Comuno App Home Page

The Home Page was created to have aesthetic and colorful view and the icons are well-labeled with metaphors that reflect functionality purposes to enhance easy navigation.

The **Listen** button represents the speech to text functionality which gives the Hearing-Impaired user the ability to listen to the environment by reading what has been translated by the application.

The **Speak** button represents the text to speech functionality which gives the Speech-Impaired user the ability to speak and interact with the environment by typing what they want to say and letting the app translate and speak for them.

The **Conversation** button opens up the conversation page which displays previous conversations and translations for reference.

Figure 5 and Figure 6 show the home page of Comuno App developed to support cross platform-Android and IOS examples.

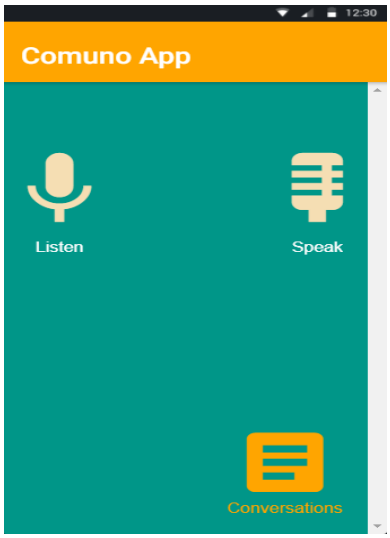

Figure 5: Comuno App Home Page Android view

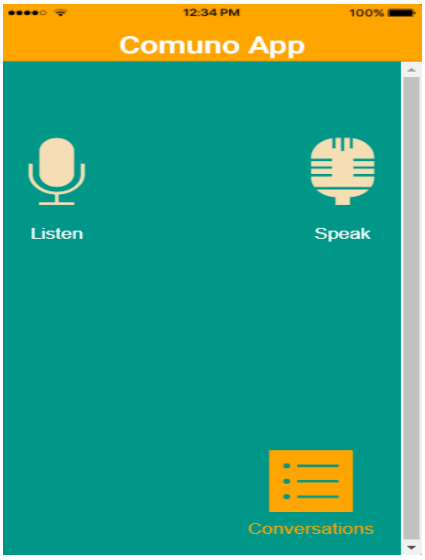

Figure 6: Comuno App Home Page IOS view

Figure 7 shows the interface conversation page with saved messages.

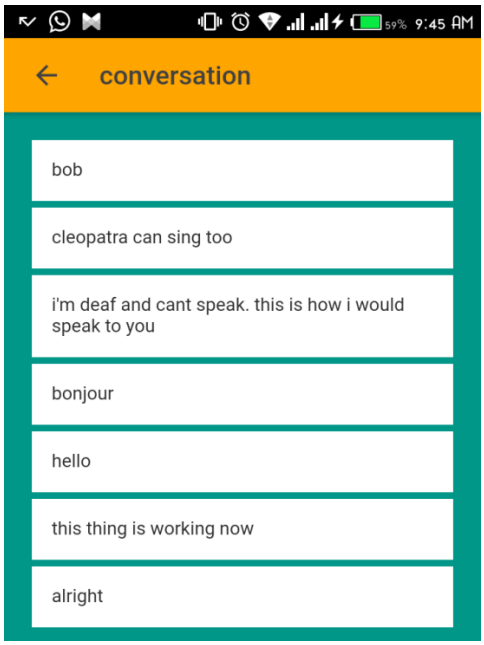

Figure 7: Android Interface Conversation Page with saved messages

VI. RESULT AND DISCUSSION

Students with hearing and speaking impairment are usually misconstrued as people who do not have the ability to learn or communicate well. The implementation of text to speech and speech to text application is a way of improving their learning and communication via computing technology. The Comuno system is an automated tool that makes learning and communication of these categories of students more effective and productive as it would be able to save conversations thereby allowing students to be able to do revision of their study. Learning and communication are done better with the TTS and STT mobile application as it helps to know exactly what message the teacher is trying to pass across

VII. CONCLUSION AND RECOMMENDATION

A standard TTS and STT mobile application is one that can translate written text to speech and also translate spoken words into text. The Comuno app overcomes challenges in previous Speech Recognition and Speech Synthesis mobile applications by making provision for the ability to save previous conversations. Future direction could be in correction of grammatical error, that is, when a person makes a grammatical error the mobile application should automatically correct it. The design could also be improved to perform location and language synchronization. That is, making the application language and natural language processing of the system change as the device changes location to another geographical area, for example, from Nigeria to Ukraine.

REFERENCES

[1] L. Darling-Hammond, L. Flook, C. Cook-Harvey, B. Barron and D. Osher, Implications for educational practice of the science of learning and development. *Applied Developmental Science*, *24*(2), 97-140, 2020.

[2] M. Bance, Hearing and aging. *CMAJ*, *176*(7): 925-927, 2007. doi: 10.1503/cmaj.070007.

[3] W. B. Alshuaib, J. M. Al-Kandari, and S. M. Hasan, Classification of Hearing Loss, Update on Hearing Loss, Fayez Bahmad Jr., IntechOpen, 2015. doi: 10.5772/61835. [Online]. Available from: https://www.intechopen.com/chapters/49574.

[4] M. Agmon, L. Lavie, and M. Doumas, The Association between Hearing Loss, Postural Control, and Mobility in Older Adults: A Systematic Review. *Journal of the American Academy of Audiology*. *28*(6), 575–588, 2017.

[5] WHO, (2015), Deafness and hearing loss Fact sheet N°300. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs300/en/

[6] R. Ladner. Technology for Deaf People. 2010. [Online]. Available: https://courses.cs.washington.edu/courses/cse590w/10sp/deaf-tech10.pdf

[7] CDC (June 08,2020), "Types of Hearing Loss". Centers for Disease Control and Prevention. [Online]. Available: https://www.cdc.gov/ncbddd/hearingloss/types.html. Retrieved 17-08-2021.

[8] WHO (2021), World Health Statistics 2021 from World Health Organization. Retrieved from https://www.who.int/data/stories/world-health-statistics-2021-a-visual-summary

[9] Centers for Disease Control and Prevention. (2017, June 5). Cytomegalovirus (CMV) and Congenital CMV Infection. Retrieved from Centers for Disease Control and Prevention: https://www.cdc.gov/cmv/index.html

[10] Mayo Clinic. (2017). Toxoplasmosis. [Online: October 3, 2017]. Retrieved: https://www.mayoclinic.org/diseases-conditions/toxoplasmosis/symptoms-causes/syc-20356249

[11] M. Hardling, and H. Huins, Hearing Loss of Older People (Presbyacusis), (2015, February 23). Retrieved from: https://patient.info/health/hearing-loss-of-older-people-presbyacusis

[12] H. N. Kim, S. G. Kim, H. K. Lee, H. Ohrr, S. K. Moon, J. Chi, et al. Incidence of presbycusis of Korean populations in Seoul, Kyunggi and Kangwon provinces. *Journal of Korean Medical Science*, *15*(5):580-4, 2000.

[13] S. Arlinger, and E. Billermark, One year follow-up of users of a digital hearing aid. *British Journal of Audiology*, *33*(4), 223–232, 1999.

[14] Y. Wu, E. Stangl, O. Chipara, S. S. Hasan, S. DeVries, and J. Oleson, Efficacy and Effectiveness of Advanced Hearing Aid Directional and Noise Reduction Technologies for Older Adults With Mild to Moderate Hearing Loss. *Ear and Hearing*, *40*(4), 805-822, 2019.

[15] C. F. Hwang, Y. Chen, H. C. Lin, P. Narayanan, S. H. Oh, and E. Truy, Cochlear Implant and Its Related Science. *BioMed Research International*, 2015, 1–2, 2015.

[16] NIH, (2021). Cochlear Implants. National Institute on Deafness and Other Communication Disorders. Retrieved from: https://www.nidcd.nih.gov/health/cochlear-implants

[17] G. Terzopoulos, and M. Satratzemi, Voice Assistants and Smart Speakers in Everyday Life and in Education. *Informatics in Education, 19*(3), 473–490, 2020.

[18] A. Dennis, B. H. Wixom, and D. Tegarden, Systems Analysis & Design With UML Version 2.0: An Object-Oriented Approach (3rd. ed.) USA: John Wiley & Sons Inc. USA, 2009.