NATURAL SPEECH RECONSTRUCTION SYSTEM WITH BANDWIDTH CONSTRAINT

BY

CHONG YEE XIANG

A REPORT

SUBMITTED TO

Universiti Tunku Abdul Rahman

in partial fulfillment of the requirements

for the degree of

BACHELOR OF INFORMATION SYSTEM (HONS)
INFORMATION SYSTEM ENGINEERING

Faculty of Information and Communication Technology
(Kampar Campus)

JAN 2019

UNIVERSITI TUNKU ABDUL RAHMAN

REPORT STATUS DECLARATION FORM

Title: ____

____

____

Academic Session: _____________

I ____

(CAPITAL LETTER)

declare that I allow this Final Year Project Report to be kept in Universiti Tunku Abdul Rahman Library subject to the regulations as follows:

1. The dissertation is a property of the Library.
2. The Library is allowed to make copies of this dissertation for academic purposes.

Verified by,

_________________________ _________________________

(Author's signature) (Supervisor's signature)

Address:

____

_________________________ _________________________

_________________________ Supervisor's name

Date: _____________________ **Date**: _____________________

NATURAL SPEECH RECONSTRUCTION SYSTEM WITH BANDWIDTH CONSTRAINT

BY

CHONG YEE XIANG

A REPORT

SUBMITTED TO

Universiti Tunku Abdul Rahman

in partial fulfillment of the requirements

for the degree of

BACHELOR OF INFORMATION SYSTEM (HONS)
INFORMATION SYSTEM ENGINEERING

Faculty of Information and Communication Technology
(Kampar Campus)

JAN 2019

DECLARATION OF ORIGINALITY

I declare that this report entitled "**METHODOLOGY, CONCEPT AND DESIGN OF NATURAL SPEECH RECONSTRUCTION SYSTEM WITH BANDWIDTH CONSTRAINT**" is my own work except as cited in the references. The report has not been accepted for any degree and is not being submitted concurrently in candidature for any degree or other award.

Signature : ___________________________

Name : ___________________________

Date : ___________________________

ACKNOWLEDGEMENTS

I would like to express my sincere thanks and appreciation to my supervisors, Dr. Aun Yichiet who has given me this bright opportunity to engage in a speech related project. It is my first step to establish a career in speech-related field. I learn a lot of knowledge on the speech-related field by the guidance of Dr. Aun. Besides, he always by my side when I was having trouble on the project. His guideline and solution of the problem are very helpful. Millions of thank Dr. Aun.

Next, I would like to thank my academic advisor, Mr Yong Tien Fui. He is the one that can give a consultation on mental health. When I was stressed, he can give me a motivation talk to pick me up. Thank you for helping me as my academic advisor, Mr Yong.

A very special thanks to a special person in my life, Chan Ling Hui, for her patience, unconditional support and love, and for standing by my side during hard times. She also tries to participate in the project so that she could do something to reduce my stress. She been become my responded for the system testing too. Thank you very much for being love and patience from my loved one.

Finally, I must say thanks to my parents and my family for their love, support and continuous encouragement throughout the course. If my parents are not supportive, I will not able to finish this project by myself. A great thanks to my parents.

ABSTRACT

Natural speech reconstruction system with bandwidth constraint is a system that enhancing existing video calling system on adaption for an emergency call. This system is mainly focused on the reconstruction of the speech but not focus on the video. The objective of the project is to develop a system that can convert the text to speech, convert back from speech to text based on the speaker's voice and to design a training using smaller dataset while achieving similar accuracy.

The system flow starts from a video is being affected by an emergency network, the system will capture the speech from the speaker. The speech will be converted into text and send to the listener device. After the listener device receives the data, the system will convert the text back to speech. This conversion involves the training voice model which is trained early and store inside the voice model database. After the conversion, the speech will be playing by the system to the listener.

The technique used in this project is speech recognition, speech synthesis and machine learning. Speech recognition is doing the speech to text conversion in the speaker device. Next, speech synthesis is the technique used to convert text to speech in the listener device. Lastly, machine learning is involved in training a voice model by using supervised machine learning.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Mbps	Megabits Per Second
SMS	Short Message Service
VoIP	Voice over Internet Protocol
SDN	Software Defined Networking
SVC	Scalable Video Coding
GDP	Greedy Display Power
API	Application Programming Interface
TTS	Text-To-Speech
IP	Internet Protocol
MP3	Moving Picture Experts Group Layer-3 Audio High-definition
HD	High-definition
SD	Standard-definition
4K	4 thousand resolution integrated development environment
IDE	Integrated development environment

Chapter 1 Introduction

1.1. Problem Statement and Motivation

Video calling system is a system that allows a user to communicate with others by video in a different location. This system uses commonly in any electronic devices with some basic software and internet connection in this generation. This system allows users to see other people on the other side on a display device. A camera is needed to capture the video of the users and show the video on the screen. Besides, for the sound that transmitted in the system, the voice from the user has recorded down by a microphone.

However, there are some network requirements for using video call. Bandwidth, a stable and minimum internet connection of 1Mbps download and upload speed is needed for a high definition video quality for video call. Therefore, a slow bandwidth will lead to a choppy video and sound. Moreover, video call might be terminated if the internet connection is unstable and this will influence the user experience.

For the existing bandwidth constraint video calling system, most of them need a higher speed of bandwidth to perform a good quality of video for video call. High bandwidth can transmit the data faster and less chance to lose a packet of the data. For example, Skype needs a minimum requirement that requires 1.2Mbps to ensure a good quality of the video call. The requirement scales up when the video call quality goes up to HD or 4K. In a slow network, the internet connection is bad and slow thus the data need more time to transmit or dropped. Current, there are some compensation methods such as auto quality scaling (from HD to SD) are used but they are not effective in an emergency network.

As mentioned above, most of the people prefer voice than video when the video call is not working well. There is a solution that solves this problem is to improve the bandwidth of the areas, but this is costly due to the price of the network devices and the price for upgrading the bandwidth. Other than this, the technology for network devices is not that advances in Malaysia too. Actually, the network service provider did a great job on the network coverage for the urban area. But for the rural area and forest area, the network coverage is bad or no coverage at all. The aim of the network service providers is to get more subscribers to their service and earn more money. They forget or selectively do not want to upgrade their device in those rural and jungle area for better network coverage. Those networks can be named as the emergency networks.

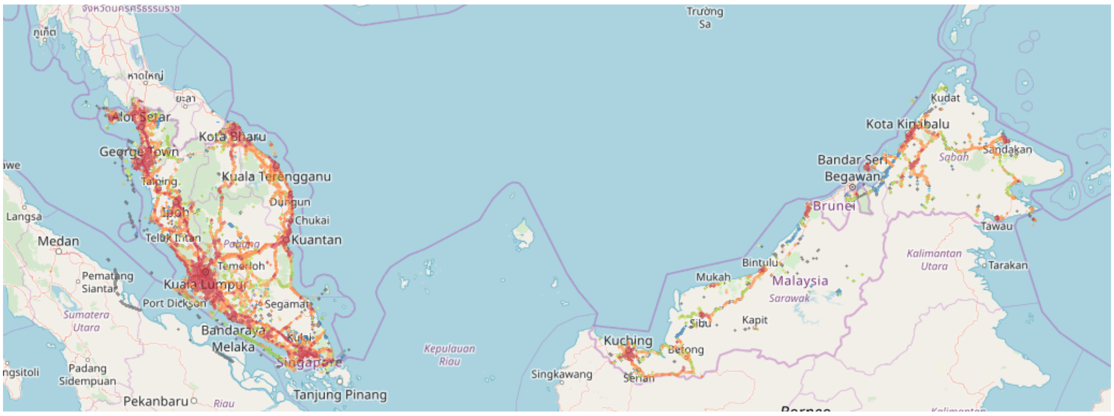


Figure 1.1: Network coverage on Malaysia

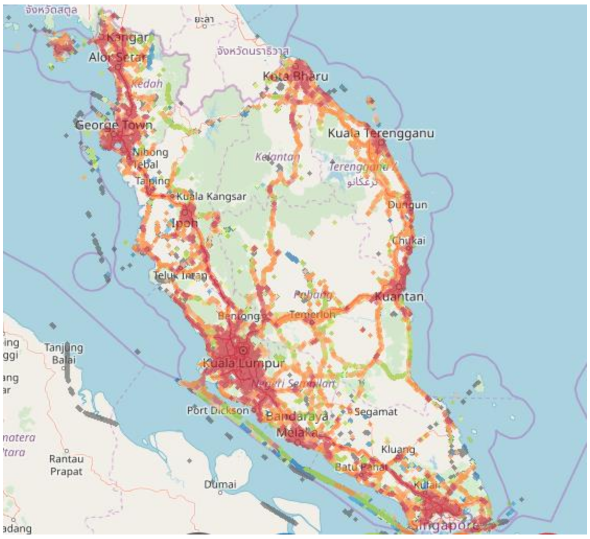


Figure 1.2: Network coverage on West Malaysia

Figure 1.1 and Figure 1.2 show the network coverage in Malaysia. The orange colour mark on the map represents 3G signal coverage while the red colour mark represents 4G signal coverage. All marked areas are the urban areas such as Kuala Lumpur, Johor, Perak and Kuantan. However, the area in Kelantan and Pahang, those are rural and jungle area has less network coverage and even no network coverage at all. A low cost solution to solve this problem is reconstructing the speech when bandwidth is too bad to transmit a video.

Text transmission uses less bandwidth than video and audio. By relying on artificial intelligence, the system can reconstruct the audio signal based on the transmitted text to recreate a similar experience while saving bandwidth. This project come out a system that supports the existing video calling system to prevent voice packet lost problem. When a sentence is transmitting on the internet, some of the words might be lost due to the bad connection of the internet. After that, the people from the other side might not hear some of the missing words and confusing what the speaker said. To solve this problem, this system will detect the lost word and reconstruct the sentence by using the voice of the speaker to pronouns the lost word. By using some technique to converting the speech to a text to reduce the size of the data and send to listener device. After that, using the voice of the speaker to pronouns the word in the text retrieved from the speaker device.

1.2. Project Scope & Limitation

This project is aimed to enhance the video calling system to cope with bandwidth constrain more effectively. Most of the existing system will automatically reduce the quality of the video to encounter bandwidth constraint. This project will more focus on the audio instead of video because converting video to a smaller file is still larger than a text file and may not surely solve the problem. The concept of this project is the video may pause but the audio is still running when video calling with bandwidth constraint. The outcome of this project is able to reconstruct the speech with a natural voice which can be compatible with a human voice. Furthermore, the project will prove that the bandwidth requirement is less than the existing video calling system.

There are a few limitations founded in this project. The level of study of this technique is high and take time to achieve it. Besides, this project is a concept development demonstration, some of the results may not perfect and few assumptions have to make to support this system. Below are the limitations of the proposed natural speech reconstruction system:

Language and Slang

The user has to speak in English and the slang of the user may not recognize due to the recognition library is not support. The recognition library has various language to recognize but the library must state a language statically before setup the system. The recognition library cannot perform a language detection and change to the preferred language dynamically. This issues also apply in the synthesis library too. The language of the voice after the reconstruction must be defined before the setup.

Noise

The noise in the environment may affect the accuracy of the speech recognizing. Basic noise adaption is developed in this system. Thus, this system can cope with some noise but most of the noises are surely affect the accuracy of the system. In the end, the performance and accuracy of the system are affected. Creating a voice recognition that can perfectly cope with noise is a high level of study and require more time to develop it. An assumption is made for this limitation, which is speaking in a quiet room or situation to ensure the correct text is being recognized.

Dynamic voice recognition

To train the machine to learn about a human voice, the machine needs a lot of voice data and keyword to work on it. The requirement to train the model is very difficult to achieve to produce a good quality voice too. The used API only can store less than 2 people's voice for the speech synthesis part. So only less than 2 people data will be collected to show the system recognize these two people and reconstruct the speech base on who is the speaker in the current state.

1.3. Project Objective

Every project and system has its own objective and ultimate goal to achieve. For this project, the ultimate goal is to reconstruct the speech using speaker natural voice in an emergency network.

- To develop a speech to text module to encode voice to text for data transmission.
- To develop a text to speech module to reconstruct text to audio.
- To design a training using smaller dataset while achieving similar accuracy.
- To simulate an emergency network to evaluate the system.
- To conduct a perception test to measure user experience versus bandwidth usage.

1.4. Impact, Significance and Contribution

Nowadays the existing video calling systems are using dynamic video quality reduction system to cope with bandwidth constraint. This system help video calling systems to adjust the video quality based on the bandwidth condition. But when the bandwidth is too bad thus unable to stream a video anymore, the video will pause at the moment including sound too. This will affect the user experience and message is wrongly delivered.

By using natural speech reconstruction system, the message is still delivered correctly even video is pausing. This system will convert the speech to text and send to listener device for converting back to speech. The user may still listen to the speaker while fixing the bandwidth connection. Users need not changing from video call to audio call if the user still can hear what the speaker is saying. This will improve the user experience since they no need to changing anything in order to continue to communicate.

User or consumer nowadays is focused on the performance on the system, but there is not mean that a system need not focus on problem and situation handling. This system helps a video calling system to handle an existing problem, which is bandwidth constraint, and solve by a new solution. More problem handling solution, more powerful of the system is. Even a high performance system with low problem handling will create a lot of problems that affect user experience, thus problem handling is important to develop a good system.

1.5. Background Information

During the old times, people were using traditional ways such as sending a letter to contact a person that is far from us. It takes a long time to reach each other when technology is not advance as now. However, nowadays, we use more convenient ways to contact each other like phone calls, sending SMS, emailing or faxing. On top of that, a video call is considered as the most popular ways of communication that employ by people.

Video calling system not only uses for interpersonal communication between friends and families but also contribute to the globalization of communication industry. For instance, an international meeting and a cooking lesson conducted by a chef to his student via online. There are some software provide this service such as Skype, FaceTime and Wechat.

Figure 1.3: Logo of Wechat, FaceTime, Skype

Using this system is a cost-effective way to stay connected with each other. Firstly, the user does not need to travel around just for a meeting as travel is costly and take time if the distances are too long. For example, a big company has a meeting with other overseas branches to discuss a new project. People are not from the main branches has to travel to the main branches to has this meeting. However, this is costly because the price of the flight is not cheap and a hotel or homestay is needed to stay there for one night. Next, a video calling system enables online coaching services for people to acquire some knowledge or skill through the internet. A teacher or lecturer can conduct a lesson to a few students by group video call. This can reduce the inconvenience of distances and time constraint. By using video call, people able to see each other face but not only listening to their sound. Using video call is similar to face to face communication in real life situation. People can observe each other's facial expression and non-verbal language and this definitely makes communication more effective than old times.

Chapter 2 Literature Review

2.1. Introduction of Video Calling System

In order to develop this speech reconstruction system to support the existing video calling system, review the existing system and the related job is a step to be done. "Video calling over wireless networks have become increasingly popular because it is more reliable" (Ghag, et al., 2014). The statement above provides the motivation to improve the existing video calling system. Before this project, there are many project or work that improve video calling system too.

The most popular video conferencing platform is Skype. According to Baset and Schulzrinne (nd.), Skype is a peer-to-peer VoIP client developed in 2003. Skype has many features other than video calling such as instant messages, audio call and buddy/contact list. The main feature of Skype is video calling, a connection of network is a must for this feature. The camera capture the video and display on the other's device. For the sound of the user, the microphone will record and play it on the other's device. The user also can switch between a video call and audio call when the stability of the connection is not stable. A user can always mute the sound and the device only displaying the other side on the screen. There is a button click to disable the microphone when a user in a video call. Besides, the user also can switch between the front camera and rear camera to show the different view to the other side. After the video call is done, the user may click the decline button to terminate the video call. The strength of Skype is the interface is simple and user friendly. For video calling system, the buttons are all located below the screen to avoid blocking the view of the screen. If the friends or family are not online, the system will auto generate and send a message to the friends to notice them call back. For the weakness of Skype, it is highly depending on high speed and a stable connection in order to have a video call. For the video part, the quality of the video will automatically decrease when the connection is not stable and slow. However, the sound from the other side is missing when the connection is bad.

Wechat, known as Weixin in China, was launched in the year 2011 and developed by China's largest listed Internet company, Tencent (Wu, 2014). It is a social media with video calling system and other communication systems such as instant message, friends circle and Wechat pay. In Wechat, the video call is one of the most popular ways to contact others. It requires an account login to the system to use the feature in Wechat. When a user is using a video call to contact someone, he is able to see the people face and hear the voice from the device. Besides, the user may switch to voice call by clicking a button to switch when the connection is not stable. The system also provides a feature that able user to switch between the front and rear camera to show the view of the surrounding. "Floating Window" is the speciality of video calling system on Wechat. The user is able to use the device to perform other tasks while video calling to others. This is the factor that leads to doing task concurrently for the user. User wish to end the call by clicking the decline button and the system will be terminated. The weakness of the video system in Wechat is similar to Skype that is high speed demand on bandwidth. If the connection is unstable, the quality of the video will be decreased to perform a smooth flow of video call. However, when the data passing process are not fully done and some of the packet loss in the process, it may lead to the video pause and some of the words are missing. This may decrease the good experience in a video call for the user. At the end of the call, the message spoke by the speaker cannot clearly convey to the listener. After that, the user may switch from a video call to voice call and the popularity of video calls may decrease.

2.2. Comparisons between the Existing System and Enhanced System

In Table 2.1, there is a comparison between the existing system (Skype and Wechat) and the new system in this project. Firstly, the minimum bandwidth needed for video call for Skype and Wechat needs more to perform a good quality of video call. However, the new system needs less bandwidth to perform a video call and a reconstruction system is introduced to show the bad connection problem for video call. Thus, the reliability of the bandwidth of the new system is less than the existing system. For this three video calling system, and authentication is needed for a user to ensure the personally of a user is protected but Wechat has more than 2 variable to verify the user. The existing system will only lower down the quality of video when the connection is unstable, but the new system will also reconstruct the speech to ensure the messages is convey to the listener. The purpose to develop the new system is to ensure the messages can convey to the listener, so the speciality and other feature is none for the new system.

System Name Factor to compare	Skype	Wechat	Enhanced Video Calling System
Minimum bandwidth needed for video call	128kbps	128kbps	<128kbps
Reliability on bandwidth	Highly rely on bandwidth	Highly rely on bandwidth	Less rely on bandwidth compare with existing system
Action when the connection is unstable	Lower down the video quality to ensure smooth video call	Lower down the video quality to ensure smooth video call	Perform same action with existing system and reconstruct the speech
Specialities on video calling systems	Auto send message if other which no pick up the call	Floating window allows user to perform task concurrently	-
Other feature	Voice call, Buddy list, Moment and etc	Friends circle, Instant message, Wechat pay and etc	Basic video calling system

Table 2.1: Comparison of the existing systems and enhanced system

2.3. Related Work

Besides the video calling system, there are some work is done to solve the problem of highly rely on bandwidth or improving existing video calling system. Since the video calling system is so famous and reliable, there are still many people try to make it better to improve the user experiences.

When the video call is started, there are some techniques to ensure that you are talking to the right person. Before that, video calling was used for teleconferencing in the corporate world (Akhil, et al., 2016). An authentication system had been introduced to support the video call system in the year 2016. This authentication system is using id and password, fingerprint scanning and face recognition system to verify the authorized user. The user first creates an account for the system. After that, the id and the password of the account are storing into a database. For the fingerprint and user's image of the face also store in the database. Each time the user login to the system, the system will retrieve the data from the database and match it. The data entries must be matched with the data store in the database, then the user login to the system successfully. If both of the user successful sign in to the system, the video call may start and the users may communicate with each other. However, if the authentication of any one of the side fail, the system will not go through the video call session until the authentication success.

One of the projects to improve video calling system is using Scalable Video Coding and Software Defined Networking techniques on video conferencing. As stated by (Hasrouty, et al., 2017), they propose an algorithm capable of reducing the bandwidth consumed by video conferencing using Software Defined Networking (SDN) to compute and deploy multicast trees, as well as taking advantage of the Scalable Video Coding (SVC) layering feature that allows sacrificing video quality for usability purposes. SVC is a layer-based video compressing technique that removes some video layers without preventing the stream to be decoded. This help the video stream bitrate reduced. Besides, SDN is define the video during streaming and identify the layer to drop to optimize network usage. Both solutions combine together to make the video conferencing to be better in term of bandwidth constraint.

Next, Tamar, et al. (2010) designed a dynamic bandwidth estimation and adaption algorithm to improve video conferencing. Bandwidth sometime might be unstable and delay the video packet transmit to other end devices or packet loss. Their project is focusing on two main part, bandwidth/delay detection and bandwidth adaption. All the process are doing in real time, so the function must be dynamic and running all the time when video conferencing. The module will analysis the bandwidth usage and generate the result for further action. The result includes the latency of the bandwidth and the speed of the internet. Over-utilization of bandwidth will be detected when delaying of the receive video packet. Upon delay detection, the module will estimate the available bandwidth and the quality of the video is adapted to fit the environment. In short, the video quality will automatically change when the bandwidth speed and stability is changed.

Other than solving high rely on bandwidth problem, there was a project that reducing display power consumption for real-time video call in mobile devices. The display subsystem of a mobile device usually consumes 38%-68% of the total battery power in video streaming (Xiao, et al., 2015). Since video calling is famous now, many people will use it for several purposes like calling from overseas, meeting and etc. This demand for video call will dream up the power of the device rapidly. So Xiao and his inventors decided to reduce the display power of video streaming but retaining the quality of the video. They designed a saving scheme call LCD-GDP (Greedy Display Power). Basically, this scheme is utilizing the available GPU without any additional support. The experiment shows that the scheme can save up to 33% of the power consumption during video call without affecting the quality of the video.

2.4. Introduction of Technique Used

The following section is talking about the technique used in this project. There are three main technique will be discussed in this section. There are speech recognition, speech synthesis and machine learning.

Speech recognition is a technique that converts from speech data to text data. It has been named as Text to Speech (TTS) technique too. It lets the user control computer functions and dictates text by voice (Das, et al., 2015). This technique uses a lot of artificial intelligent and algorithm to recognize what the speaker is said in a word form. It separates into two part, the first part is using a microphone to capture the acoustic signal and process it. The second part is analysing the processed signal and map the signal with the word. This technique needs to consider some issues also, the biggest issue is the noise of the environment. Mel Frequency Cepstral Coefficients (MFCC) is the acoustic observation for this technique, but it is easily affected by noise (Mitra, et al., 2014). So while using the system, try to stay in a quiet area and use it.

Next, speech synthesis is the opposite part of speech recognition. Speech synthesis is a process sends a text to a speech synthesizer, which creates a spoken version that can be output through the audio hardware or saved to a file (Hande, 2014). In short sentence, it converts text data to speech data, is an opposite technique compare with speech recognition. It basically using concatenating pieces of recorded speech that are stored in a database to create the speech. To measure the quality of the converted speech, there are two things can be measured that is naturalness and intelligibility. Naturalness is how close the voice of the speech is similar to the human voice, while intelligibility is the speech is easy to understand.

Lastly, machine learning plays an important role in this project. Machine learning is a field of computer science that evolved from studying pattern recognition and computational learning theory in artificial intelligence (Simon, et al., 2015). It is able to predict a result by learning from some given data sets. It just like a human being, learn from the past and predict the future, the concept is the same. To train a model, previous experience data sets with the result is being executed. The machine will look through all the data sets and the result. After many of the data is being read, some testing data sets will pass to the machine for prediction. This technique is used on training the custom voice for speech synthesis. Currently, many libraries that provide speech synthesis service are using their own voice but not custom voice. Thus a process to train a custom voice using machine learning is needed in this project.

Chapter 3 System Design

At this chapter, the system design of the project is delivered briefly and the technologies involved will be included also. This chapter contains a few diagrams to describe the system clearly. Those diagrams are including block diagram and some UML diagram to clarify the flow of the system. The UML diagrams that being used in this project is use case diagram and activity diagram. Those diagrams will make the reader more understanding of the system and able to rebuild the system.

3.1. System Component

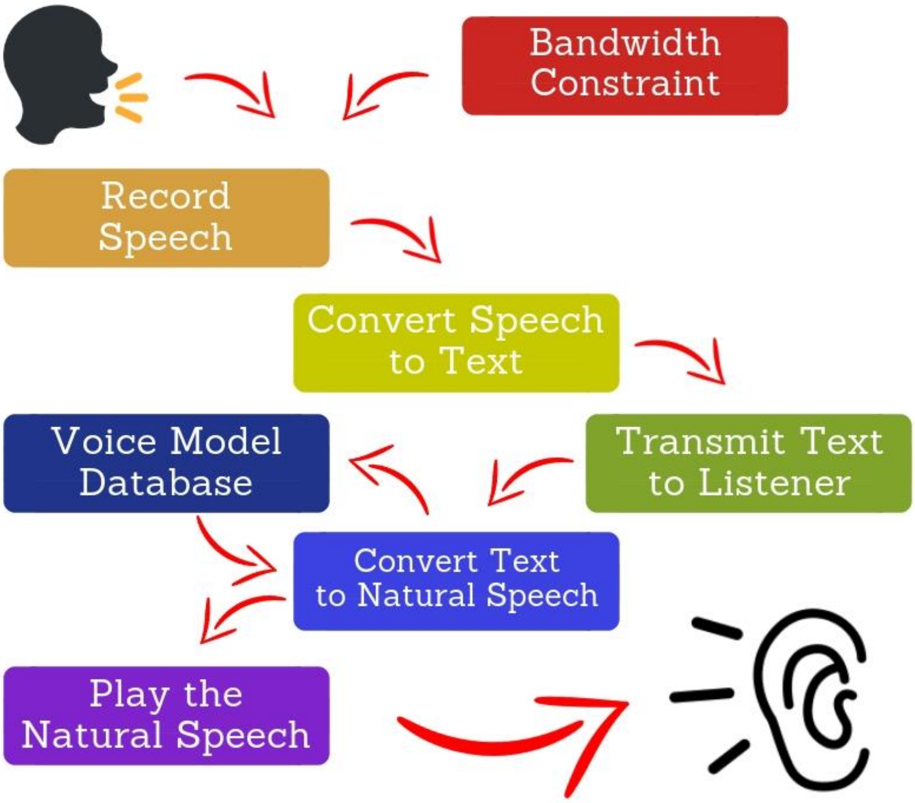


Figure 3.1: Block Diagram for Natural Speech Reconstruction System

The block diagram above shows the overview of the system. Assuming the network is in bad condition, but the user is still speaking and stream the video to another device. First, the device will record the speech and convert the speech to text and write into a file in the speaker's device. After converting to a text file, the file will transmit to the listener's device. The text file is small and easier to transmit compare with video and audio file. After the file has been sent, the device will use the trained voice model which is the speaker voice as the voice and language use in the speech synthesis part from the voice model database. After that, the text will convert to the speech with the speaker natural voice. After reconstruction the speech, the speech will play on the device, thus the information and message are delivered to the listener in a bandwidth constraint.

3.2. UML Diagrams

UML (Unified Modelling Language) diagram is a useful visual representation of a software system design. It displays all the information and flow of the implementation of the system. All the detail and information will be present in a visual model or in diagram form to make it simple and easy to understand.

3.2.1. Use Case Diagram

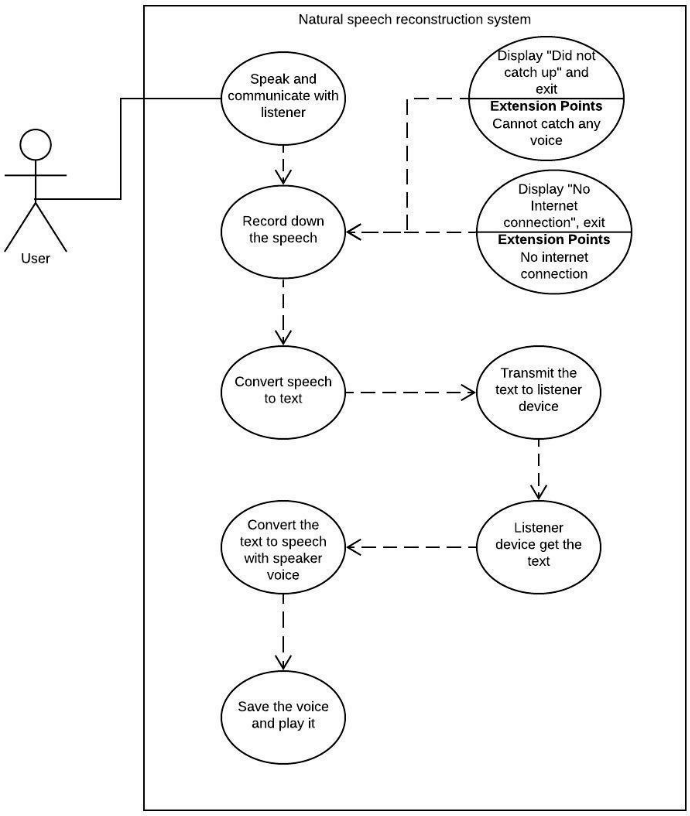

Natural speech reconstruction system
User
Speak and communicate with listener
Display "Did not catch up" and exit
Extension Points
Cannot catch any voice
Record down the speech
Display "No Internet connection", exit
Extension Points
No internet connection
Convert speech to text
Transmit the text to listener device
Listener device get the text
Convert the text to speech with speaker voice
Save the voice and play it

Figure 3.2: Use Case Diagram for Natural Speech Reconstruction System

Figure 3.2 shows that the use case diagram has one actor which is user or speaker. The interaction within the user and the system are speaking and communicate with the listener. The speaker can say something on the system and the system will perform some action after it. Besides, the use case diagram consists of 9 use cases which are record down the speech, transmit the text to listener and etc. Convert speech to text use case and convert the text to speech with speaker voice use case is core use cases in this project. A description of each use case will be shown following this.

Use Case 1: Speak and communicate with listener

Actor: User

Goal: To initial the system and capture the voice by microphone

Overview: The user will speak and communicate with the listener, the system will capture the speech using a microphone when the video calling is ongoing.

Use Case 2: Record down the speech

Actor: User

Goal: To record down the captured speech from the speaker

Overview: All the captured speech from the speaker will be recorded down to the buffer for the use of data conversion. There is some condition checking in the following extension use case.

Use Case 3: Display "Did not catch up" and exit

Actor: User

Goal: To detect the sound signal from the speaker

Overview: This purpose of this step is a condition checking to check whether the speaker is speaking or not. Unclear voice and no sound signal is detected, the system will display “Did not catch up” and exit the program.

Use Case 4: Display “No Internet connection” and exit

Actor: User

Goal: To detect Internet connection availability

Overview: This is a condition checking that check whether there is an Internet connection or not. If there is no Internet connection, the system will display “No Internet connection” and exit.

Use Case 5: Convert speech to text

Actor: User

Goal: To convert the speech data to text data.

Overview: The process of the speech data converting to text data. After the speech is been recorded down, the speech will convert to the text data. The text data will write into a file in the speaker device for transmitting purpose.

Use Case 6: Transmit the text to listener device

Actor: User

Goal: To transmit the text data to the listener device.

Overview: The process of transmitting the data. Listener device will listen to the request coming from the speaker device. When the data is ready, it will send the data and a request to the listener device.

Use Case 7: Listener device get the text

Actor: User

Goal: To receive the data from the speaker device.

Overview: The process of receiving the data from the speaker device. There will be a request message come from the speaker device with the data that send to the listener device. After the request processing is done, the speaker device may close the connection.

Use Case 8: Convert the text to speech with speaker voice

Actor: User

Goal: To generate a speech using the speaker's voice.

Overview: The process of converting a text data back to a speech with speaker natural voice. After receiving the text from the speaker device, the system will map the text data with the trained voice model and generate a natural speech as an output. The voice and language come from the voice model database.

Use Case 9: Save the voice and play it

Actor: User

Goal: To keep the speech in a file and play it to the listener.

Overview: The process of writing the speech data into a file and playing it to the listener. After generating the speech, the system will write the data into an mp3 file for playing and keeping as a recording purpose.

3.2.2. Activity Diagram

The activity diagram shows the flow of each use case in detail. All the flow and technique used will be discussed base on the activity diagram.

Activity 1: Speak and communicate with listener

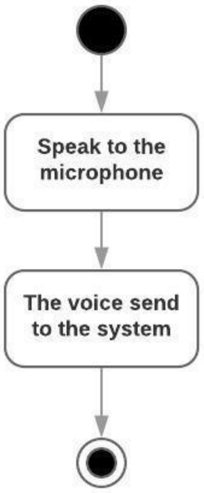


Figure 3.3: Speak and communicate with listener activity diagram

This activity diagram showed the initial state of the system. The speaker is communicating with the listener in a video call with a device on each side. Meanwhile, the microphone is capturing the speech and voice from the speaker. All the voice and speech is used in the natural speech reconstruction system when a bandwidth become very badly until the video call cannot proceed.

Activity 2: Record down the speech

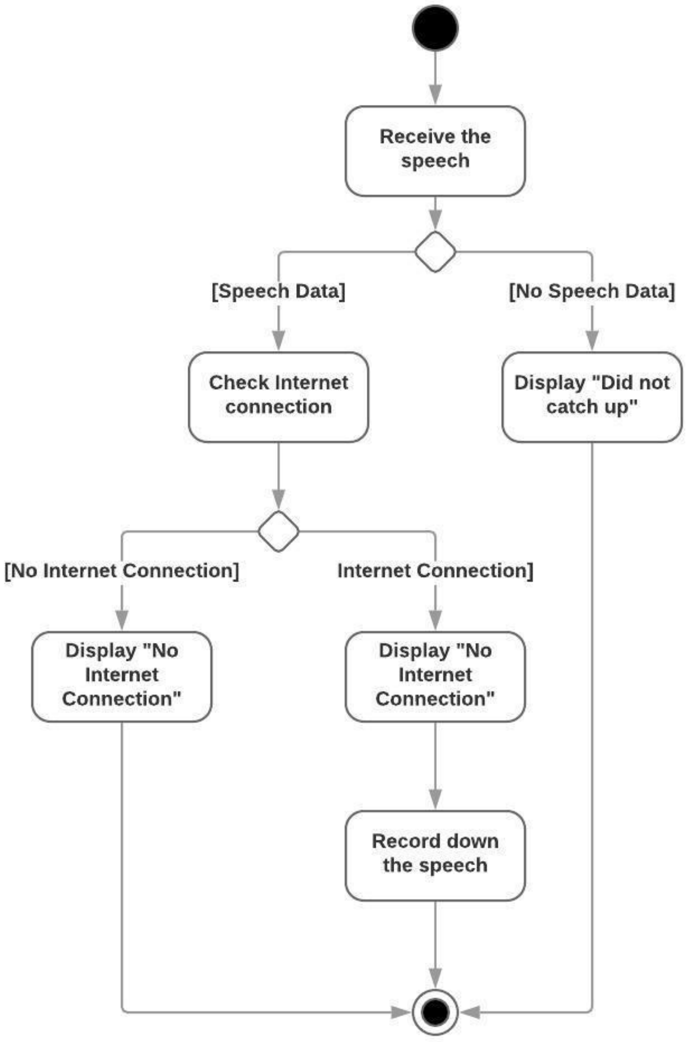

Receive the speech
[Speech Data]
[No Speech Data]
Check Internet connection
Display "Did not catch up"
[No Internet Connection]
Internet Connection]
Display "No Internet Connection"
Display "No Internet Connection"
Record down the speech

Figure 3.4: Record down the speech activity diagram

This activity combines the detail of 3 use case because there are 2 extension cases for this activity. The system will receive the speech from the microphone which is spoken by the speaker. The first condition checking is to check whether there are speech data and signal or not. If there is no speech data catch from the microphone, the system will display "Did not catch up" and exiting the program. If there is a speech data from the microphone, the flow will proceed to the second condition checking. This condition checking check for the availability of an Internet connection. If there is no Internet connection, the system will display "No Internet connection" and exit the program. If all the condition is given a positive result, the flow continues to record down the speech.

Activity 3: Convert speech to text

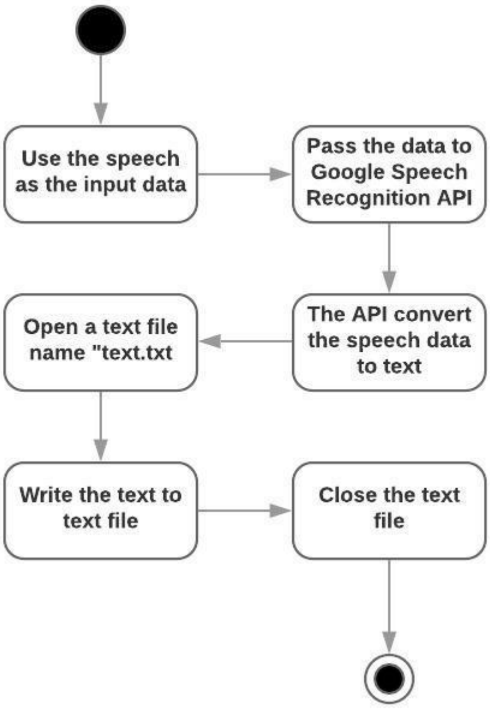

Use the speech as the input data
Pass the data to Google Speech Recognition API
The API convert the speech data to text
Open a text file name "text.txt
Write the text to text file
Close the text file

Figure 3.5: Convert speech to text activity diagram

Figure 3.5 shows the activity diagram for conversion of speech to text activity. After recording the speech, the system will use the recorded speech as the input data for the conversion. The system will pass the speech data to Google Speech Recognition API for converting the speech to text. After the conversion, the API will return a text data and the system needs to store it. Thus, the system opens a text file name "text.txt" and write the result into the text file. After using the text file, the system closes the text file to avoid missing any changes made on the file. After writing the data, the process will proceed to the next activity.

Activity 4: Transmit the text to listener device

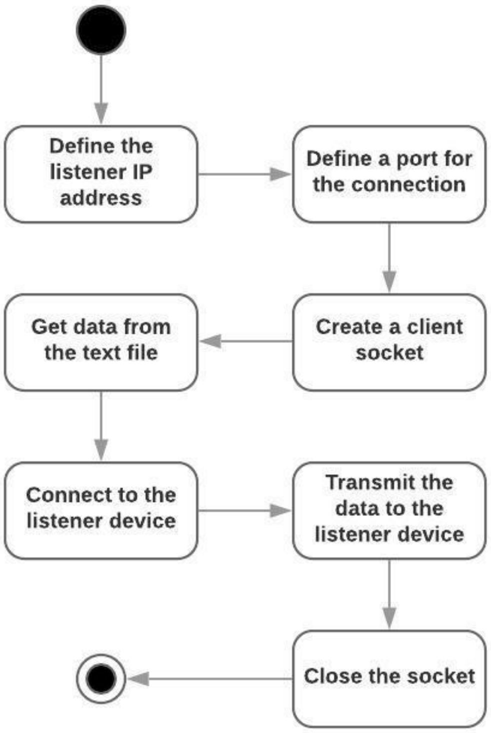

Define the listener IP address
Define a port for the connection
Create a client socket
Get data from the text file
Connect to the listener device
Transmit the data to the listener device
Close the socket

Figure 3.6: Transmit the text to listener device activity diagram

This activity showed the transmission of the data in the system. This process involved socket and networking programming to create a socket for both sides of the device. This activity diagram is discussing the configuration and process on the speaker device. First, the system in the speaker device will define the listener IP address. Next, the system may define a port number for the connection. A server may have different service but only have one IP address. Thus, the port number is important because it can determine service and the request type. After configuring the IP address and port number, the system will create a client socket to start the connection. After that, create a packet to hold the text data that need to be transmitted later. If the listener device is listening to get some request, the connection between two devices exists and the text data will transmit through the connection. After the transmission, the system will close the socket to stop the connection between both devices.

Activity 5: Listener device get the text

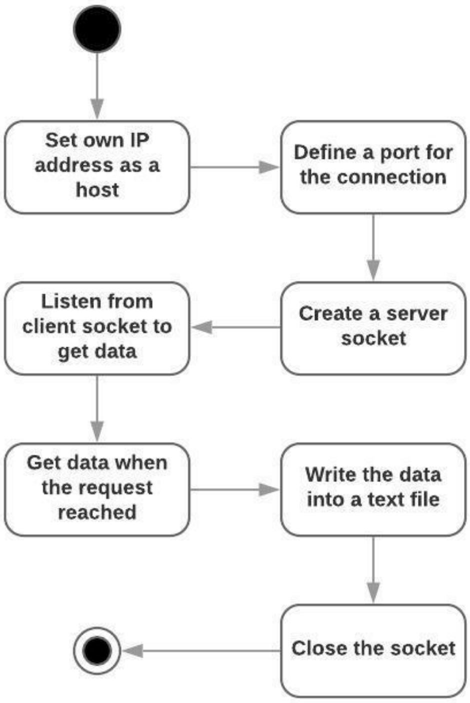

Set own IP address as a host
Define a port for the connection
Create a server socket
Listen from client socket to get data
Get data when the request reached
Write the data into a text file
Close the socket

Figure 3.7: Listener device get the text activity diagram

In Figure 3.7, the detail of flow for listener device gets the text activity is shown. This process has a similar configuration compare with previous activity. It involves socket and networking programming to transmit and receive the data. This activity is mainly focused on the system in the listener device. Firstly, the system will set its own IP address as the host IP address. This is representing the device as a server device to get a request from another device. Next, the system will define a port for the connection. It just like create a service with a specified number to differentiate with other services. After defining the port number, the system will create the server socket and bind the IP address and the port number to create the connection. After all the configuration is completed, the server side device is ready to listen for a request. It keeps listening to another device for request and data come in. When a client side sends a request with the text data to the server side, it will display connected by the client side with the IP address. Besides, the data is sent to the listener device and have to find something to store it. Thus, after getting the data from the speaker device, the system will write the data into a text file for the conversion. When all the request has been processed, the connection may end and proceed to the next activity, which is the text to speech conversion process.

Activity 6: Convert the text to speech with speaker voice

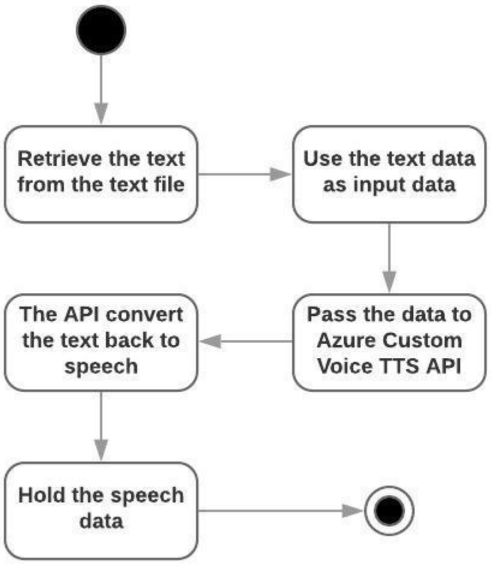

Retrieve the text from the text file
Use the text data as input data
Pass the data to Azure Custom Voice TTS API
The API convert the text back to speech
Hold the speech data

Figure 3.8: Convert the text to speech with speaker voice activity diagram

As mention in the previous activity diagram description, Figure 3.8 is the activity diagram that discusses the detail of converting the text to back speech with speaker voice activity. It starts from retrieving the text data from the text file that previously stored by previous activity. The text file contains the word speak by the speaker and send from the speaker device. Thus, the text data is using as the input data for the conversion. After that, the system will pass the data to Azure Custom Voice Text to Speech API to do the conversion. The API will do the conversion from text to speech, the voice of the speaker has to be defined before this process. Thus, a voice model has been trained early and store in the database of the API. The API can use the preferred custom voice model from the database and convert the text back to speak with that voice. At the end of the process of the API, a speech is generated with the speaker natural voice. Lastly, the API passes the speech data back to the system and the system holds it for later use.

Activity 7: Save the voice and play it

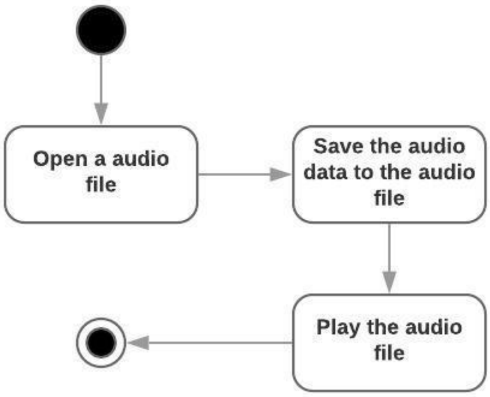


Figure 3.9: Save the voice and play it activity diagram

This is the last activity of the system, save the voice and play it. This process is mainly focused on storing and playing part. After the conversion from text to speech, the data has to store it somewhere. The system will open an audio file to store the data. Next, the system will store the audio data into the audio file. Each of time a conversion is done, an audio file will be created with a new name to record purpose. After the file has been saved, the system will play the audio file automatically.

Chapter 4 Design Specification

4.1. Methodology

In this section, the methodology will be discussed in detail. A methodology describes the way the system is being developed thought out the whole development process. The example of the methodology can be used is waterfall model, rapid development and prototyping model. In the development process for this system, the used methodology is evolutionary prototyping model.

Evolutionary prototype life cycle considers to first deliver an initial fielded prototype subsequent modifications and enhancements result in delivery of further more mature prototypes (Kumar.Arikepudi, 2003). Why this life cycle is being chosen because this system needs more and frequent feedback from the user to enhance and develop a better system. Hence, the process of developing the prototype of the system keeps repeating until the user feels satisfaction on the system function and performance.

The advantages of this life cycle are the system may match with the requirement very closely because the developing phase keeps on repeat and repeat until the system meet the requirement. The system also will be better performance because every error found on the prototype is being focused and solve after developing a new prototype.

However, there are some disadvantages to this life cycle too. The time usage and the cost of the project are difficult to determine since the prototyping phase is kept going. Furthermore, the approach for this life cycle will have less project planning and system analysis. Most of the time and resources are focusing on developing the system prototype again and again.

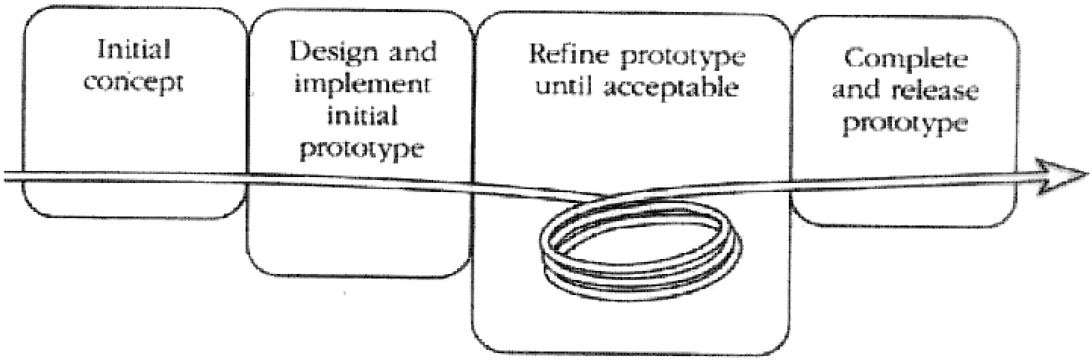

Initial concept
Design and implement initial prototype
Refine prototype until acceptable
Complete and release prototype

Figure 4.1: Evolutionary Prototype Life Cycle

Initial concept

This is the phase that gathers more information about or related to the system. All this information is coming from existing literature such as journal and reliable article. This process helps to clarify the requirement and non-functional requirement of the system. It initials a lot of ideas to develop the system and understand how the system work and the flow. The technique like speech synthesis, speech recognition is the information get from the journal and they are core development skill for this system. At the end of this phase, a project plan is the outcome of the phase. The project plan consists of an overview of the system and the overview development process.

Design and implement initial prototype

This phase is to specify the design and implementation of the initial prototype. The information gathered previously will use as a reference to design the natural reconstruction speech system. Google Speech Recognition and Azure Custom Voice Synthesis are being chosen as the core services for this system after the research done in the initial concept phase. After planning for the technique used, some diagram like UML diagram is being drawn to specify the system design. UML diagram is able to clarify the action of the user and the activity done by the system in a visual presentation.

After all system design and specification part, an initial version of the prototype is being developed next. Before developing the prototype, some software and programming library have to install first. Besides, the development process has to follow the project planning which is created in the previous phase. After the first prototype finishes developing, it is being tested and evaluated by the user and the user will give feedback and suggestion for the improvement of the next prototype. At the end of the phase, the developer will need to develop a new prototype base on the system requirement, feedback and suggestion.

Refine prototype until acceptable

The refinement and modification of prototype are being done in this phase. The developer has to come out with a new prototype by refining and modifying the previous prototype. After developing a new prototype, the user will test and evaluate the prototype again and give the developer some feedback. This process keeps on repeating until the final prototype is developed. Each time a prototype is developing, the system requirement and the quality of the system is a consideration for the developer.

Complete and release prototype

This is the last phase of the evolutionary prototyping life cycle. At the end of this phase, a complete natural speech reconstruction system is developed and able to deliver to the user. The user is satisfied with the functional requirement of the system after the last evaluate on the last prototype.

4.2. Technology Involved

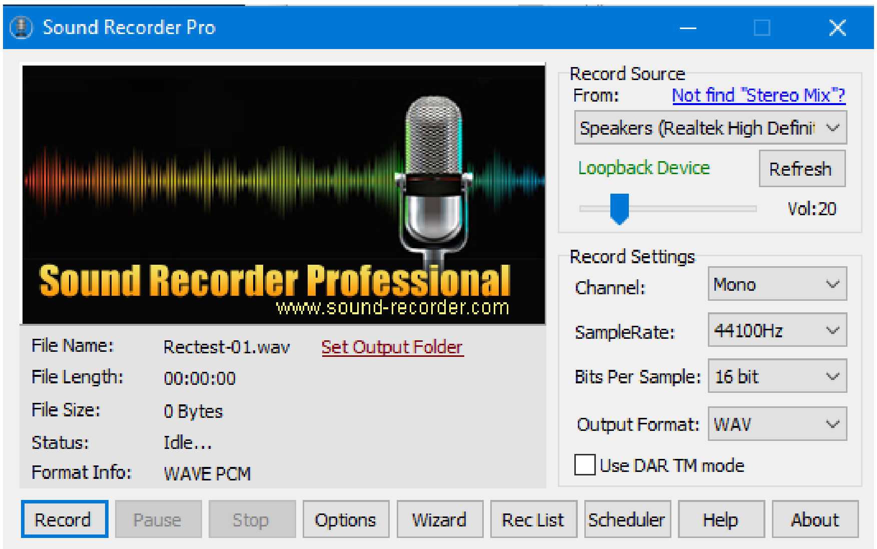
4.2.1. Software

Atom

Figure 4.2: Logo of Atom

Atom is an open source IDE provided by GitHub. This IDE is a basic text editor that can install any package to suit the project language. For this project, a Python debugger is installed to read Python file and code in Atom. An autocomplete package for Python is available in Atom too, it helps to correct the syntax error. Besides, Atom is has a simple interface compare with other professional IDE, this helps to reduce the level of confusing while developing the system. Besides, it is easier to find out the error because the error is underlined by a colour line.

Sound Recorder



Record Source
From: Not find "Stereo Mix"?
Speakers (Realtek High Defini
Loopback Device Refresh
Vol:20

Record Settings
Channel: Mono
SampleRate: 44100Hz
Bits Per Sample: 16 bit
Output Format: WAV
□ Use DAR TM mode

File Name: Rectest-01.wav Set Output Folder
File Length: 00:00:00
File Size: 0 Bytes
Status: Idle...
Format Info: WAVE PCM

Figure 4.3: Interface of Sound Recorder

The main purpose of sound recorder software is to record the voice and speech from the speaker. All the voice and speech is recorded for the training of the voice model. The reason why this software is chosen because this software provides the feature that can adjust the sample rate and bits per sample of the recorded audio. The service that trains the voice model has a requirement and format for the audio.

NetLimiter 4

Figure 4.4: Logo of NetLimiter 4

NetLimiter 4 is very useful software for monitoring the bandwidth usage of a computer. The services provided by NetLimiter 4 including graph for bandwidth usage per second, limiting the bandwidth speed and software bandwidth usage checking. This software help to evaluate the system by creating an emergency network and measure the bandwidth usage of the system that needs to compare with.

4.2.2. Hardware

Customization Asus Desktop Computer

Figure 4.5: Asus Desktop Computer

The main hardware to develop this system is using the developer own customization desktop computer with Asus motherboard. It runs all the required software to develop the system. The specifications of the computer are shown below:

• Processor	: Intel® Core™ i5-4400H CPU @ 3.10GHz
• RAM	: 8GB
• Operating System	: Window 10 Pro 64-bit Operating System
• Graphic Card	: NVIDIA GeForce GTX660
• Storage	: 1TB SSHD
• Speaker	: Vinnfier Ether 121X

Microphone

Figure 4.6: Microphone

A microphone is a must in this system to record down the speech spoken by the speaker. The main purpose of the microphone is just recording down the speech to perform the speech to text process. The brand of the microphone is original equipment manufacturer (OEM).

4.2.3. Programming Language

Python

Figure 4.7: Logo of Python

There are some tools needed in this project. This project is using Python programming language to develop because Python is using more simple syntax and more suitable for machine learning as there are many machine learning libraries can be used in Python. In order to train the machine, this project needs a lot of voice data and keyword to train the machine to generate the expected output. Besides, Python extensively supports many libraries. The libraries are easy to import and used too.

4.2.4. Technique and Skill

Speech recognition

In order to generate a text based on the speech from the speaker, speech recognition is needed in this section. A microphone is a must for this system to record down the speaker's voice and send to the buffer in the listener's device. Before sending to the receiver side, the system will apply speech recognition technique to generate a text file based on the speech of the speaker. Each of the word said by the speaker will record down and write inside a .txt text file to ensure the size of the file is small.

Speech synthesis

Speech synthesis is the skill needed in this project to regenerate the speech and play the speech using the speaker's voice. This system may apply text-to-speech (TTS) system to generate the speech using the speaker voice and the input text together. When there is bad connectivity for internet, the speech reconstruction system converts the speech to text and using speech synthesis to generate the back speech from text. After reconstruction and generate the speech, the system plays the speech.

Machine learning

If using existing text to speech system to reconstruction the speech, the voice may not natural and not the speaker own voice. By using machine learning, the machine may remember the speaker voice and able to reconstruct the speech using the speaker voice naturally. Machine learning needs a lot of data to train a voice model. This data including the voice of the speaker, some utterance and etc. This component makes up about 50% of this system because collecting the data and training section take a longer time to complete.

Optimisation

Every new system that wants to improve some existing system must be optimised the system to have a better performance with lower cost. To maintain a low cost in this project, by using a free source of software to develop the system is one of the ways. Besides, testing the system using the asset of school such as the computer in the computer lab. Using a shorter and simple coding style to develop the system to ensure the performance of the system is faster and stable. A text file has a smaller size compared with audio and video file, so sending text file is faster than sending the audio file and this also ensures the performance of the system.

Socket Programming

Natural reconstruction system is a system that enhances a video calling system, data transmitting is a consideration on this project. To create a similar environment of video calling system, the text data converted in the speaker device has to transmit to the listener device. This process has to use socket programming in order to perform it. The listener device act as a server, keep listening to the request come from another device, while the speaker device act as a client, send a request with the data to the server.

4.2.5. Library and Services

Google Speech Recognition Library

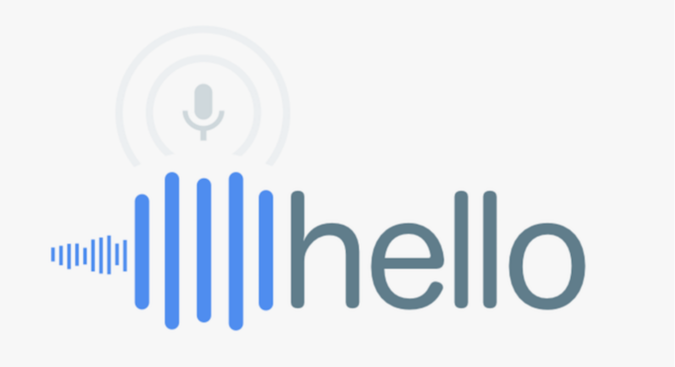


Figure 4.8: Logo of Google Speech Recognition Library

Google Speech Recognition Library playing an important role in this project. It is a library created by Google and Python able to use very easily. To install this library, just tap "sudo pip3 install SpeechRecognition" after installing Python language in the computer. After that, it able to use as a library by just import the library, call the method and use it. Besides easy and simple library, it is a reliable library that it has very high accuracy on converting speech to text. In short, Google Speech Recognition Library is an easy to use and with high accuracy, reliability library for this system.

Azure Custom Voice Service

Azure Custom Voice Service is a service that provides a machine learning training environment for custom voice. It is able to train a voice model and use in the speech synthesis part of the system. In other to train a good and natural voice model, it needs to follow the requirement and rule to work with it. Before using this service, an Azure account is a must to activate this service. Subscribe the Azure Cognitive Service and get a subscription key to use these services. After activating the Custom Voice Service, a set of voice data and transcripts has to submit to the Azure portal for training purpose. The time to train the voice model is depending on the size and number of voice data. After the voice data has been trained to voice model, it can be used as a preferred language and voice for the speech synthesis part.u

Azure Text to Speech Service

Figure 4.9: Logo of Azure Cognitive Services

Figure 4.9 shows the logo of Azure Cognitive Services. Azure Cognitive Services is a combination of all voice and speech services. Azure Text to Speech Service is the core service that uses this project and it is a subservice from Azure Cognitive Services. To activate the Text to Speech Service, the developer has to create an Azure account and subscribe to Azure Cognitive Services. After activating the Text to Speech Service, is can be used as a library in Python programming language by import the library and use its method. There is a language and voice selection in this service, the default language will be selected by the service. But this project wants an output with speaker natural voice. Thus, the reason to use this service is that it can easily use the voice model train by Azure Custom Voice Service.

4.3. Functional Requirement

The requirement below is the basic functionality needed in the system which includes:

- The microphone is able to capture the speech from the speaker.
- The system is able to convert the speech to text.
- The system in the speaker device able to transmit the text data to the listener device.
- The system in the listener device able to receive the text data from the speaker device.
- The system is able to reconstruct the text back to speech.

- The system is able to call the operating system to play the speech.
- The data transmit is maintain in small size

4.4. Assumption

There are some assumptions have to made in order to miss generate a correct output and avoid some errors which include:

- The language speaks by the user must be United State English.
- The slang of the user must be avoided.
- User is expected to use the system in a quiet environment.
- User voice has to train before using the system.
- Only least than 2 user voice can store in the voice model database

4.5. System Performance Definition

There is a system performance definition in this project. This project is not focused on speech and accuracy but the voice is nature. The voice in reconstructed speech has to be similar compared with the original speech. The project objective is to deliver the message clearly in a bandwidth constraint situation and the original voice from the speaker. If the reconstruction process using the default system voice, it has no difference between some speeches to text system like iPhone Siri. So, ensuring the voice is nature is important and it is the project aim.

4.6. Evaluation Plan

After developing the system, some of the test data and technique will apply the system to ensure it works well. The process to evaluate the naturalness of the speech is to demonstrate the system used by the developer and get some feedback and rating about the naturalness of the speech. Next, the accuracy of the converted speech needs to evaluate too. A set of test text data that need to speak by the developer. Next, the developer uses the system to generate the speech and match the test text data and speech similarities.

4.7. Implementation Issues and Challenges

There are a few limitations founded in this project. Since there is limited time to complete this project, some of the results may not perfect and few assumptions have to support this system. Below are the limitations of the proposed natural speech reconstruction system:

Language and Slang

The user has to speak in English and the slang of the user may not recognize due to the recognition library is not support. The recognition library has various language to recognize but the correct system only can detect English as the language that needs to reconstruct.

Noise

The noise in the environment may affect the accuracy of the voice recognizing. This system can cope with some noise but most of the noises are surely affect the accuracy of the system then the system may perform very badly. Creating a voice recognition that can perfectly cope with noise is a high level of study and require more time to develop it.

Dynamic voice recognition

To train the machine to learn about a human voice, the machine needs a lot of voice data and keyword to work on it. In the current state, only less than 2 people data will be collected to show the system recognize these two people and reconstruct the speech base on who is the speaker.

4.8. Project Timeline

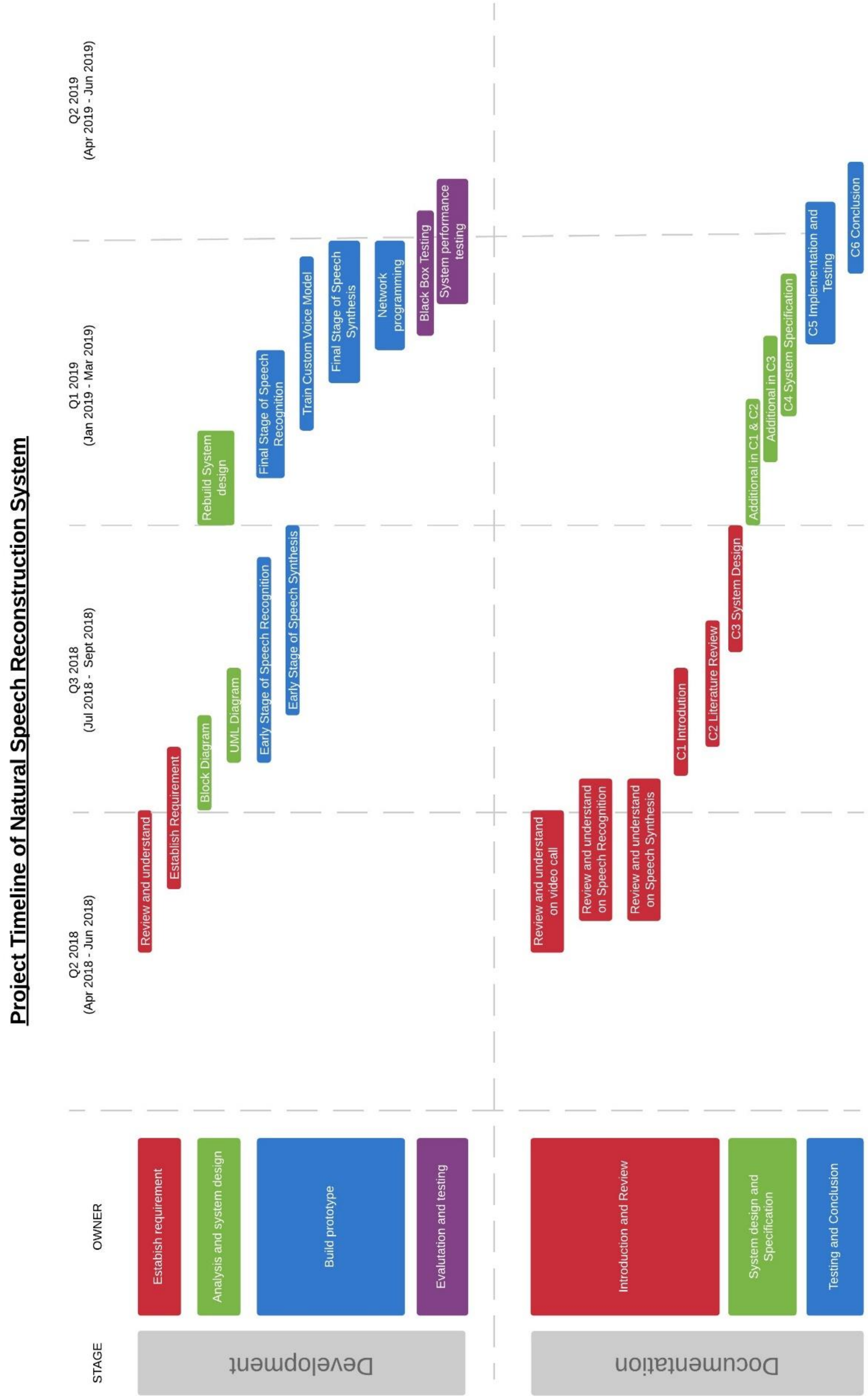


Figure 4.10: Timeline for this project

Chapter 5 Implementation and Testing

5.1. System Implementation

This section is discussing the system implementation that specifies how the system archive the output and how to implement the technique. This project is a voice and speech related project, there are some cognitive services used which is speech recognition and speech synthesis. For custom voice model implementation, it is a machine learning based technique to train the voice model and able to use in speech synthesis.

5.1.1. Custom Voice Model

The part will explain the implementation of custom voice model. A machine learning based service, Azure Custom Voice Services, provide a great environment to train the custom voice model. To train a voice model that use on the speech synthesis part, few speech data are needed as training data. A voice training dataset consists of a set of audio files, along with a text file that contains the transcripts of the audio files. To produce a good voice model, make the recordings in a quiet room with a high-quality microphone. Consistent volume, speaking rate, speaking pitch, and expressive mannerisms of speech is essential for building a great digital voice. The dataset has to archive in a .zip file before uploading to the database. The format of the speech data is stated below:

• File format	: RIFF (.wav)
• Sampling rate	: at least 16,000 Hz
• Sample format	: PCM, 16-bit
• File name	: Numeric, with .wav extension
• Archive format	:.zip
• Maximum archive size	: 200 MB

If the format of the voice dataset is not meeting the requirement, the dataset will be rejected by the service. For the sampling rate, those voice data that has the sampling rate lower than 16,000 Hz in a .zip file, the service will reject them and only accept those voice file with equal or higher than 16,000 Hz.

Next, the transcription file is important for the training process. The service gives some guideline about the format of the transcription file too. The transcription file is a plain text file (.txt file). Each line of the transcription file must have the name of an audio file, followed by a tab character, and finally, its transcript and no blank lines are allowed. It is important to provide a 100% accurate transcriptions of the corresponding audio recordings to the service to train a voice model. The reason behind this is because the service is using Supervised Machine Learning. The text act as a sample and the voice act as a result of the prediction. The training process is to match them one by one using the given voice dataset. The following table shows the example of the format of the transcription file:

Name of audio file	Transcript
001	This is the waistline, and it's falling.
002	We have trouble scoring.
003	It was Janet Maslin.

Table 5.1 Example of transcription file

After all the data are approved by the service, the process will proceed to train the voice model. Azure provides an environment to train the voice data to a voice model. The developer only chooses which data need to train, the language of the voice model and the gender of the speaker. After that, name the voice model as a determinant of the voice model to use on the speech synthesis of Azure. It takes time base of the size of the uploaded voice dataset. The time range of training is from 30 minutes for hundreds of utterances to 40 hours for 20,000 utterances.

Finally, Azure able to develop the speech synthesis endpoint and suit the system. The endpoint is in a URL form and will be used in the code on the project.

5.1.2. Speech Recognition

For the speech recognition part in this project, Google Speech Recognition API is used. Speech recognition has to speak some word and sentences to the microphone, and the captured speech is converting to text. Google Speech Recognition API is powerful and able to recognise the speech correctly, but the limitation is Internet connection is needed to use its service. By using Python programing language, it needed to install PyAudio library and PortAudio library to capture the speech using a microphone. Besides, Google Speech Recognition API has to be installed early too. After setting up the microphone and API, use "r" as the recognizer and “m” as the microphone in the code. If “m’ captured something, the system put the speech to “r” for recognition using “r.recognize_google(audio)”. It will return a text in string data type from Google Speech Recognition API. The text is the important data that need to send it to listener device.

5.1.3. Speech Synthesis

This section is explaining the speech synthesis implementation on natural speech reconstruction system. Speech synthesis plays an important role in natural speech reconstruction system because this technique generates the output. The tools to achieve this is using Azure Speech Synthesis Service. This service and Azure Custom Voice Model comes from Azure Cognitive Service, so they can use together. For default of the service, the language and gender of the voice are defined by the service. But the voice model can be changed to some predefine voice model from the service itself. Since both of the services come from the same category, they can integrate together and use on this system. The voice model created by Azure Custom Voice Model is a choice that can use on Azure Speech Synthesis Service. Each voice model has a name to identify the model. Thus, change the name of the voice and language option on the Azure Speech Synthesis Service to the name of the Azure Custom Voice Model created before, the output of the system will be the speaker's voice.

5.2. System Testing

System testing is the process of evaluating the system based on the system requirement and the non-system requirement. The purpose of testing is to ensure the system is well developed and it is able to deliver to the user. A testing result will be generated after a few appropriate testing methods. Black Box Testing is used to test on the system based on the functional requirement. Performance and accuracy of the system is the consideration of this non-system requirement testing. System performance testing will be tested on the bandwidth usage and the accuracy of the output match with the original input. In this testing, there are some methods to test and they are different from the system requirement testing.

5.2.1. Black Box Testing

Test	**Test Case**	**Expected Outcome**	**Actual Outcome**	**Result**
1	Initialize Microphone	Able to capture speech from user	Able to capture voice and word from speaker	Pass
2	Speech recognition	Able to convert speech to text.	Text file is generated from the system	Pass
3	Transmit data	Able to transmit the text data	The data is transmitted after the conversion	Pass
4	Receive data	Able to receive data from speaker device	The data is received and speaker IP address is displayed	Pass
5	Speech synthesis	Able to convert text to speech	Generated a speech based on the text	Pass
6	Play the audio file	Able to call system to play the audio	Audio file is playing after the conversion	Pass
7	Small size of data transmission	Transmit a small size of data	The size of data transmit is small then voice data	Pass

Table 5.2 Black box testing result

Table 5.2 show the result of Black Box Testing on natural reconstruction system. Black Box Testing is also called functional testing. All the test case is based on the functional requirement. Black box testing not concern with the internal mechanisms of a system; these are focus solely on the outputs generated in response to selected inputs and execution conditions (Nidhra & Dondeti, 2012). The table has clearly shown all the test case with an expected outcome and actual outcome. After all the test case is being tested on the system, the result of all test cases are passed, which mean all the functional requirement is being developed and designed in this system very well.

5.2.2. System Performance Testing

Bandwidth Testing

The objective of this system is to develop a natural speech reconstruction system that can cope with bandwidth constraint. Thus, there are some testing between the bandwidth and the system. The participation system in this testing is Skype, Wechat and Natural Speech Reconstruction system. The first testing is tested on the bandwidth usage for the data transmission of these three systems. Both existing video calling system transmits video data listener device while Natural Speech Reconstruction system transmits text data to listener device. The testing environment is provided by NetLimiter 4 which can view the bandwidth usage graph per second and the total size of transmitted data. This testing is tested on a normal and stable connection. The table below is the result of bandwidth usage for data transmission of three systems.

Name of system	Bandwidth usage per second (KB)
Skype	50 KB
Wechat	60 KB
Natural Speech Reconstruction	10 KB

Table 5.3 Bandwidth usage for three systems

The next testing is tested on the adaptability for data transmission in an emergency network of these three systems. NetLimiter 4 able to congest the network into a bad network with a certain speed rate. The figures below are the bandwidth graph example of the networks.

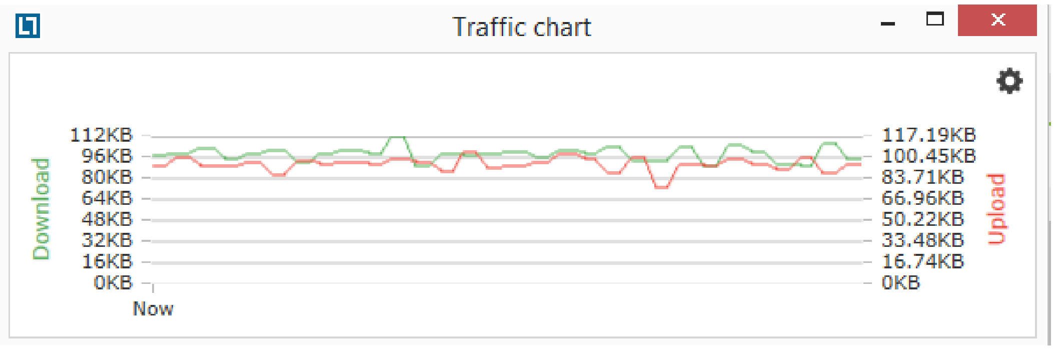


Figure 5.1: Network limited in 100 KB per second

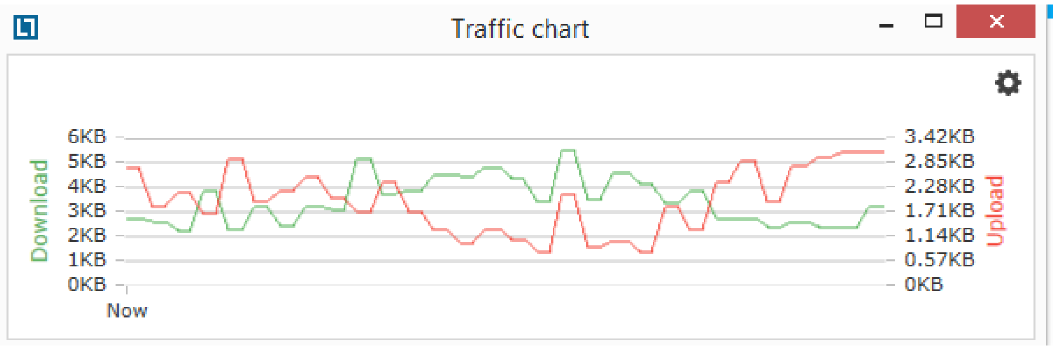


Figure 5.2: Network limited in 5 KB per second

These networks will be applied on Skype, Wechat and Natural Speech Reconstruction system. Assume the last network, network limited in 5KB per second, as an emergency network. Besides, assume a basic video calling system is applying Natural Speech Reconstruction system, the figure show the graph of the result on this emergency network testing.

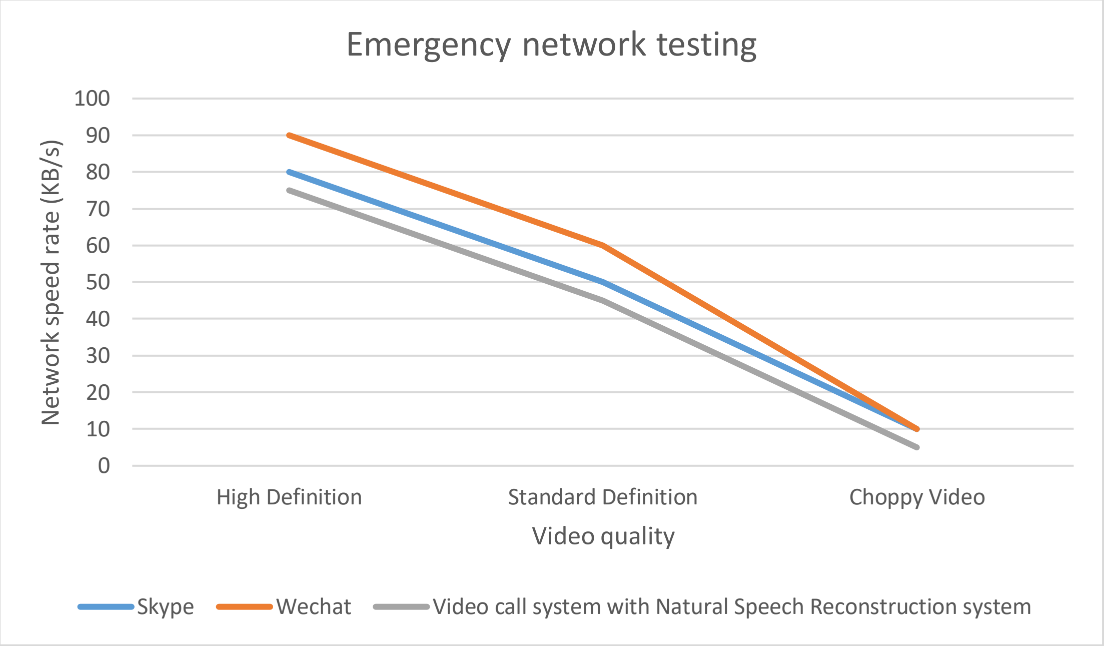


Figure 5.3: Result of the emergency network testing

Name of system	Network speed rate that shut down the system (KB/s)	Action after detecting emergency network
Skype	8	Choppy video quality
Wechat	10	Pausing video and terminate the system.
Natural Speech Reconstruction	5	Text data able to transmit to listener device to reconstruct

Table 5.4: Result of the emergency network testing in table form

This testing will focus more on what is the bandwidth speed threshold of each system to reach a pausing or choppy video. The video call system with Natural Speech Reconstruction system has the lowest bandwidth speed threshold because the text data is still able to transmit under the emergency network. When there is an emergency network, Skype and Wechat will have a pausing video with no voice at all during the video call because the file of transmission is too large to transmit in the emergency network. These systems will terminate if the file is unable to transmit anymore.

Naturalness

The aim of this project is to develop a system that can construct the speech back from the text but with a natural voice. Hence, naturalness is a must to test in this system. In this testing, two existing systems are used to compare with this system. The process of the evaluation separate to two part, the first part is given the observation and listening test the speech and voice come from Skype, Wechat and Natural Speech Reconstruction system. After the observation and listening test, a short survey is given to get the feedback and score on the naturalness based on each system. The respondents are Chan Ling Hui, Chan Ling Hao and Mah Hong Wai. The table below shows the result of each respondent based on the systems.

	Name of system		
Name of respondent	Skype	Wechat	Natural Speech Reconstruction System
Chan Ling Hui	9	8	7
Chan Ling Hao	8	8	5
Mah Hong Wai	7	7	6
Total Score	24	23	18
Average Score	8	7.67	6

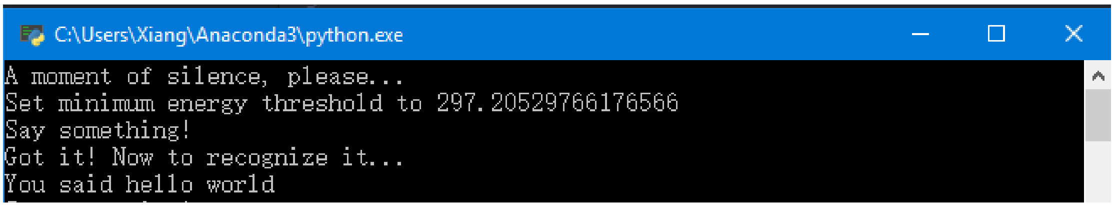
Table 5.5: Naturalness Score for three systems

The naturalness score for Skype and Wechat have the higher score compare with Natural Speech Reconstruction System in the testing result. A comment from Chan Ling Hao state the speech come from Natural Speech Reconstruction System has the voice of the speaker, but the naturalness is still lower than Skype and Wechat. For Chan Ling Hui, three of them are good enough and Skype has the best naturalness. Based on the average score for each system, Skype has the highest score compared with other system and Natural Speech Reconstruction System has the lowest score. The reason behind this score is because Skype and Wechat are transmitting the video and audio data but Natural Speech Reconstruction System is transmitting text file and reconstruct the speech based on the text file.

Accuracy

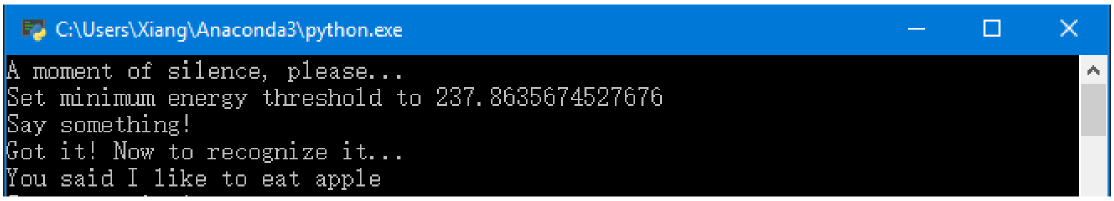
For testing part on this system, there are some sample sentences will test on the system. For a more visual result, this testing method only tests on the speech recognition part. Besides, Azure Speech Synthesis provides a very accurate service that can convert all word with correct pronunciation as long as the text file with no error such as spelling and extra spaces. The text generated by Google Speech Recognition API will not have any spelling error and extra spaces, only will recognize the wrong word. Hence if there is no spelling error, the testing is only applicable in the speech recognition part. This testing is applying on three people, which are Chan Ling Hui, Chan Ling Hao and Mah Hong Wai. The sample sentences are shown below:

- Hello World
- I like to eat apple.
- Hello I'm fine thank you.
- What fruit do you like to eat?
- What course do you study in your university?
- Today is Sunday let go outside to have some fun.



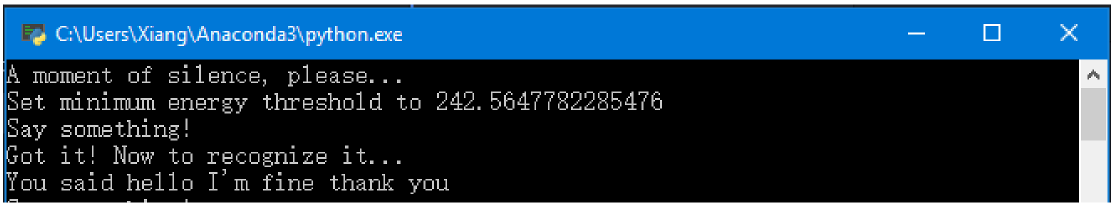
```
A moment of silence, please...
Set minimum energy threshold to 297.20529766176566
Say something!
Got it! Now to recognize it...
You said hello world
```

Figure 5.4: Testing result for sample sentence 1



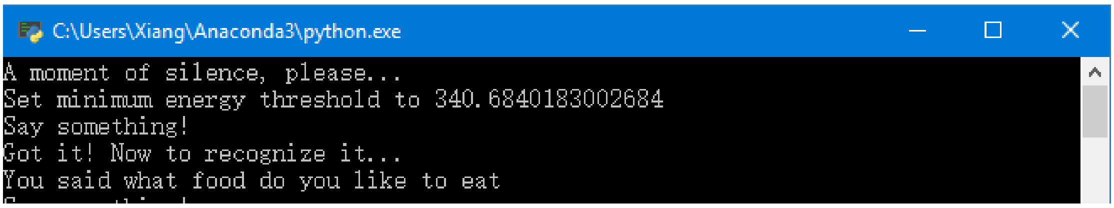
```
A moment of silence, please...
Set minimum energy threshold to 237.8635674527676
Say something!
Got it! Now to recognize it...
You said I like to eat apple
```

Figure 5.5: Testing result for sample sentence 2



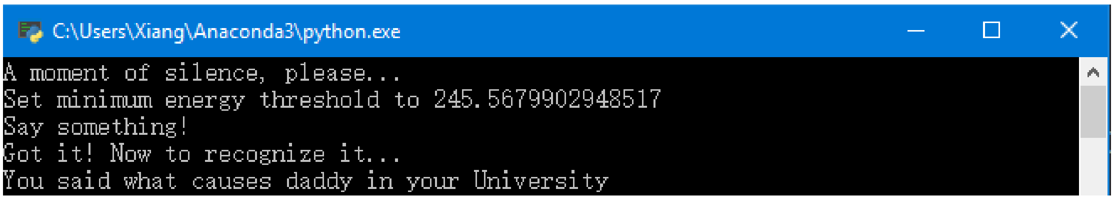
```
A moment of silence, please...
Set minimum energy threshold to 242.5647782285476
Say something!
Got it! Now to recognize it...
You said hello I'm fine thank you
```

Figure 5.6: Testing result for sample sentence 3



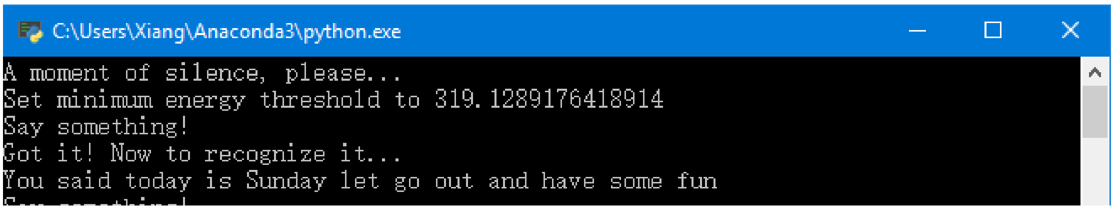
```
A moment of silence, please...
Set minimum energy threshold to 340.6840183002684
Say something!
Got it! Now to recognize it...
You said what food do you like to eat
```

Figure 5.7: Testing result for sample sentence 4



```
A moment of silence, please...
Set minimum energy threshold to 245.5679902948517
Say something!
Got it! Now to recognize it...
You said what causes daddy in your University
```

Figure 5.8: Testing result for sample sentence 5



```
A moment of silence, please...
Set minimum energy threshold to 319.1289176418914
Say something!
Got it! Now to recognize it...
You said today is Sunday let go out and have some fun
```

Figure 5.9: Testing result for sample sentence 6

The figures above are the result of Chan Ling Hui. There are 37 words in all sample sentence and there are only 5 words are wrongly recognised. The formula to calculate the accuracy of the system and the calculation is shown below:

$$accuracy = \frac{total\ of\ correct\ recognised\ word}{total\ of\ word\ that\ need\ to\ recognise} \times 100\%$$

$$accuracy = \frac{32}{37} \times 100\%$$

$$= 0.8649 \times 100\%$$

$$= 86.49\%$$

Figure 5.10: Accuracy formula and accuracy calculation

For speech synthesis, the result is in MP3 file so here only can be written down the testing result. There is no error on converting text to speech process. All the 37 words are clearly converted to speech and write into a MP3 file.

In short, the Google Text to Speech API is very accurate with a 86.49% of accuracy score, as long as the speech to text process is correct, the text to speech process will not cause any accuracy problem on the system. But the limitation on both API is Internet connection is needed to use the services provided by the API.

Chapter 6 Conclusion

Video calling is a popular way to contact family, friend and employee. It able people to see each other face at the same time hearing each other voice. Video calling can perform at an office meeting, cooking tutorial and friendly chitchat. But video calling relies on a higher bandwidth to support video streaming. When bandwidth reduces, the video quality will automatically reduce but worse is the video will pause at a very bad connection.

In short, this natural reconstruction system is developed for video calling system to help the message is delivered clear and complete in a condition with bad bandwidth. Natural reconstruction system will work when the network is bad condition and the message is unable to deliver to the listener. This will cause some conflict due to misunderstanding the unclear message.

In order to ensure the message is delivering well in bandwidth constraint, the only thing that at least needs to be transmitted is the speech but not video. But the size of an audio file is big too compare with a text file. So this system will generate a text file base on the speech and send it to the listener device to reconstruct the speech back from the text file.

There are some issues and challenges in this project due to the limitation of time. The slang and language are pre-set and only available for English. Other than English the system may not recognise and cannot generate the text for the reconstruction process. After that, noise is hard to avoid because creating a speech recognition system that can cope with noise is a high level of study and need more time to complete it. Lastly, dynamic voice recognition is a challenge in this project. This project will only get less than or equal to two people's voice to perform and implement in the system. This project mainly is to deliver the concept and the implementation of the system.

This project needs a few techniques to complete it, which is speech recognition, speech synthesis and machine learning. This three techniques are the main technique and have a bigger percentage on this project. By using speech recognition, it helps the system to convert the recorded speech into a text and write into a text file to ensure a small file will be delivered. After the file sent, using machine learning and train the machine to pronoun with the speaker's voice. Lastly, using synthesis to generate back the speech base on the given text. The device will play the speech after all work is done.

This concept and system will be tested before the delivery date. The evaluation method divided into two part, system requirement testing and performance testing. The result of system requirement testing is pass and all the requirement is perfectly delivered by the system. Next testing is tested on the performance which is bandwidth utility, naturalness and accuracy. A short observation and listening test is conducted first and follows by a survey to test the naturalness of the reconstructed speech. Although the test result is lower than other existing systems, it is still natural and the speech comes from the speaker's voice. Next, accuracy testing is using a normal microphone to record a spoken word to test the system. Written down the speech and compare the accuracy of the output with the speaker's speech. The project not only focuses on the implementation and involvement in the project, but it also focuses on the accuracy of the output. There is no point to develop this system if the output is not accurate and the main objective may not achieve at the end of the project.

Bibliography

Akhil, G. S., Manjunath, C. & Ashuthosh, V., 2016. Video Calling System Using BiomectricI Remote Authentaication. *International Journal of Electronics and Communication Engineering and Technology (IJECET),* 7(5).

Anon., n.d. *Sniffer – what it is and how to defend against sniffing | Avast.* [Online] Available at: https://www.avast.com/c-sniffer
[Accessed 21 March 2018].

Baset, S. A. & Schulzrinne, H. G., n.d. An Analysis of the Skype Peer-to-Peer Internet.

Das, P., Acharjee, K., Das, P. & Prasad, V., 2015. Voice Recognition System: Speech-To-Text. *Journal of Applied and Fundamental Sciences,* pp. 191-195.

Ghag, S. N., Kakade, K. D., Goyal, R. N. & Kamthane, A. A., 2014. Video Calling Over Wi-Fi Network using Android Phones. *Computer Engineering and Intelligent Systems.*

Hande, S. S., 2014. A Review on Speech Synthesis an Artificial Voice Production. *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering,* 3(3), pp. 8056-8063.

Hasrouty, C. A. et al., 2017. SVC Videoconferencing Call Adaptation and Bandwidth Usage in SDN Networks.

Kumar.Arikepudi, A., 2003. Software Prototyping.

Lukic, M. A., 2015. Benefits and Security Threats in Electronic Banking. 3(6), pp. 44-47.

Mitra, V., Franco, H., Graciarena, M. & Vergyri, D., 2014. Medium-Duration Modulation Cepstral Feature for Robust Speech Recognition. *2014 IEEE International Conference on Acoustic, Speech and Signal Processing,* pp. 1749-1753.

Nidhra, S. & Dondeti, J., 2012. Black Box and White Box Testing Techniques – a Literature Review. *International Journal of Embedded Systems and Applications,* 2(2), pp. 29-50.

Simon, A., Deo, M. S., Venkatesan, S. & Babu, D. R., 2015. An Overview of Machine Learning and its Applications. *International Journal of Electrical Sciences & Engineering,* 1(1), pp. 22-24.

Tamar, B. et al., 2010. TREND: A DYNAMIC BANDWIDTH ESTIMATION AND ADAPTATION. pp. 126-133.

T, S., V, U. & M, C., 2014. Duration Modelling Using Neural Networks for Hindi TTS System Considering Position of Syllable in a Word. *Procedia Computer Science 46 (2015),* pp. 60-67.

Wu, J., 2014. How WeChat, the Most Popular Social Network in. *Master of Applied Positive Psychology (MAPP).*

Xiao, M., Liu, Y., Guo, L. & Chen, S., 2015. Reducing Display Power Consumption for Real-time Video Calls on Mobile Devices. *Symposium on Low Power Electronics and Design*, pp. 285-230.

Appendices

Appendix A Final Year Project 2 Biweekly Report

FINAL YEAR PROJECT WEEKLY REPORT

*(Project I / __**Project II**__)*

Trimester, Year: Jan, 2019	**Study week no.: 2**
Student Name & ID: CHONG YEE XIANG 15ACB02458	
Supervisor: DR AUN YICHIET	
Project Title: NATURAL SPEECH RECONSTRUCTION SYSTEM WITH BANDWIDTH CONSTRAINT	

1. WORK DONE Previous work done by FYP1.
2. WORK TO BE DONE Do more review on Custom Voice Model, understand more on Custom Voice Model.
3. PROBLEMS ENCOUNTERED Lack of information and example from online resources.
4. SELF EVALUATION OF THE PROGRESS Slow progress, all progress on learning and understanding.

_________________________ _________________________

Supervisor’s signature Student’s signature

FINAL YEAR PROJECT WEEKLY REPORT

*(Project I / **Project II**)*

Trimester, Year: Jan, 2019	**Study week no.: 5**
Student Name & ID: CHONG YEE XIANG 15ACB02458	
Supervisor: DR AUN YICHIET	
Project Title: NATURAL SPEECH RECONSTRUCTION SYSTEM WITH BANDWIDTH CONSTRAINT	

1. WORK DONE

Finish the review and understanding on Custom Voice Model.

2. WORK TO BE DONE

- Check back and validate previous work.
- Find way and solution on Custom Voice Model.

3. PROBLEMS ENCOUNTERED

Lack of idea on training the Custom Voice Model.

4. SELF EVALUATION OF THE PROGRESS

Fast progress, need a fast catch up progress after Chinese New Year.

Supervisor’s signature

Student’s signature

FINAL YEAR PROJECT WEEKLY REPORT

*(Project I / **Project II**)*

Trimester, Year: Jan, 2019	**Study week no.: 7**
Student Name & ID: CHONG YEE XIANG 15ACB02458	
Supervisor: DR AUN YICHIET	
Project Title: NATURAL SPEECH RECONSTRUCTION SYSTEM WITH BANDWIDTH CONSTRAINT	

1. WORK DONE

A voice data model is trained and prepare to be used. The system able to reconstruct the speech with speaker voice.

2. WORK TO BE DONE

- Separate the system, find solution for data transmission.
- System Requirement Testing.

3. PROBLEMS ENCOUNTERED

Less understanding on network programming.

4. SELF EVALUATION OF THE PROGRESS

Normal progress, everything back on schedule.

____________________________ ____________________________

Supervisor’s signature Student’s signature

FINAL YEAR PROJECT WEEKLY REPORT

*(Project I / **Project II**)*

Trimester, Year: Jan, 2019	**Study week no.: 9**
Student Name & ID: CHONG YEE XIANG 15ACB02458	
Supervisor: DR AUN YICHIET	
Project Title: NATURAL SPEECH RECONSTRUCTION SYSTEM WITH BANDWIDTH CONSTRAINT	

1. WORK DONE

The job scope on data transmission and system requirement testing is done.

2. WORK TO BE DONE

- System Performance Testing.
- Documentation.

3. PROBLEMS ENCOUNTERED

Less reliable bandwidth monitor from online resources.

4. SELF EVALUATION OF THE PROGRESS

Normal progress, everything back on schedule.

_________________________ _________________________

Supervisor's signature Student's signature

FINAL YEAR PROJECT WEEKLY REPORT

*(Project I / **Project II**)*

Trimester, Year: Jan, 2019	**Study week no.: 11**
Student Name & ID: CHONG YEE XIANG 15ACB02458	
Supervisor: DR AUN YICHIET	
Project Title: NATURAL SPEECH RECONSTRUCTION SYSTEM WITH BANDWIDTH CONSTRAINT	

1. WORK DONE

A result is generated from the system performance testing and the documentation is done.

2. WORK TO BE DONE

- Finalize documentation.
- Prepare for presentation.

3. PROBLEMS ENCOUNTERED

Documentation formatting issues.

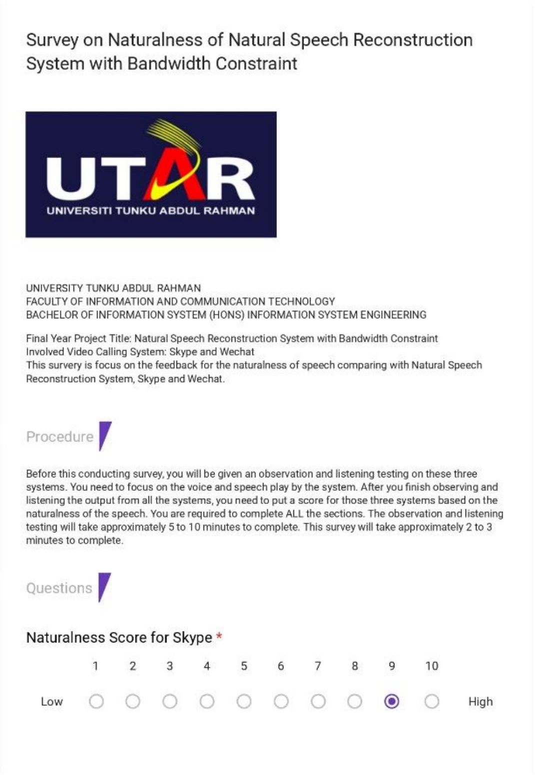
4. SELF EVALUATION OF THE PROGRESS

Normal progress, everything back on schedule.

_________________________ _________________________

Supervisor's signature Student's signature

Appendix B Survey result for each respondent



Survey on Naturalness of Natural Speech Reconstruction System with Bandwidth Constraint

UTAR
UNIVERSITI TUNKU ABDUL RAHMAN

UNIVERSITY TUNKU ABDUL RAHMAN
FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY
BACHELOR OF INFORMATION SYSTEM (HONS) INFORMATION SYSTEM ENGINEERING

Final Year Project Title: Natural Speech Reconstruction System with Bandwidth Constraint
Involved Video Calling System: Skype and Wechat
This survery is focus on the feedback for the naturalness of speech comparing with Natural Speech Reconstruction System, Skype and Wechat.

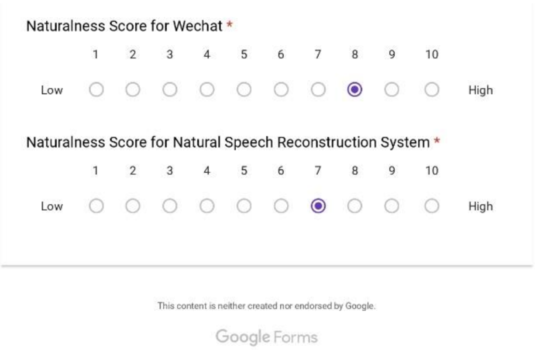
Procedure

Before this conducting survey, you will be given an observation and listening testing on these three systems. You need to focus on the voice and speech play by the system. After you finish observing and listening the output from all the systems, you need to put a score for those three systems based on the naturalness of the speech. You are required to complete ALL the sections. The observation and listening testing will take approximately 5 to 10 minutes to complete. This survey will take approximately 2 to 3 minutes to complete.

Questions

Naturalness Score for Skype *

	1	2	3	4	5	6	7	8	9	10	
Low	○	○	○	○	○	○	○	○	●	○	High



Naturalness Score for Wechat *

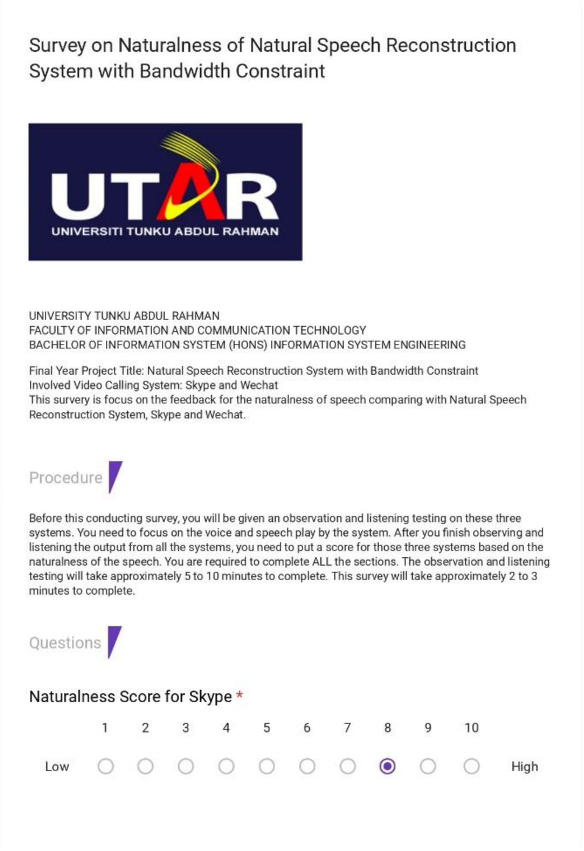
	1	2	3	4	5	6	7	8	9	10	
Low	○	○	○	○	○	○	○	●	○	○	High

Naturalness Score for Natural Speech Reconstruction System *

	1	2	3	4	5	6	7	8	9	10	
Low	○	○	○	○	○	○	●	○	○	○	High

This content is neither created nor endorsed by Google.

Google Forms



Survey on Naturalness of Natural Speech Reconstruction System with Bandwidth Constraint

UNIVERSITY TUNKU ABDUL RAHMAN
FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY
BACHELOR OF INFORMATION SYSTEM (HONS) INFORMATION SYSTEM ENGINEERING

Final Year Project Title: Natural Speech Reconstruction System with Bandwidth Constraint
Involved Video Calling System: Skype and Wechat
This survery is focus on the feedback for the naturalness of speech comparing with Natural Speech Reconstruction System, Skype and Wechat.

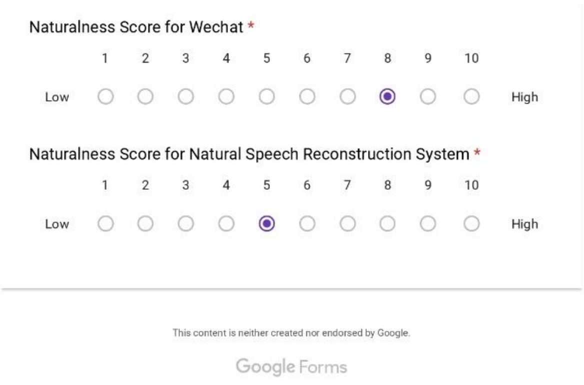
Procedure

Before this conducting survey, you will be given an observation and listening testing on these three systems. You need to focus on the voice and speech play by the system. After you finish observing and listening the output from all the systems, you need to put a score for those three systems based on the naturalness of the speech. You are required to complete ALL the sections. The observation and listening testing will take approximately 5 to 10 minutes to complete. This survey will take approximately 2 to 3 minutes to complete.

Questions

Naturalness Score for Skype *

	1	2	3	4	5	6	7	8	9	10	
Low	○	○	○	○	○	○	○	●	○	○	High



Naturalness Score for Wechat *

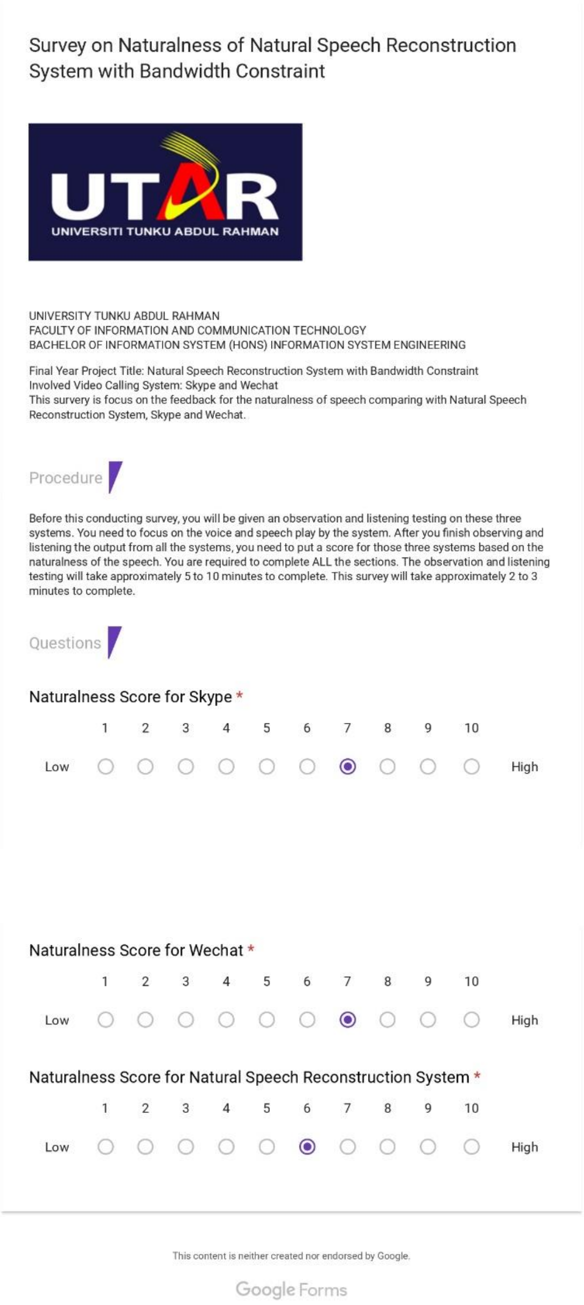
	1	2	3	4	5	6	7	8	9	10	
Low	○	○	○	○	○	○	○	●	○	○	High

Naturalness Score for Natural Speech Reconstruction System *

	1	2	3	4	5	6	7	8	9	10	
Low	○	○	○	○	●	○	○	○	○	○	High

This content is neither created nor endorsed by Google.

Google Forms



Survey on Naturalness of Natural Speech Reconstruction System with Bandwidth Constraint

UNIVERSITY TUNKU ABDUL RAHMAN
FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY
BACHELOR OF INFORMATION SYSTEM (HONS) INFORMATION SYSTEM ENGINEERING

Final Year Project Title: Natural Speech Reconstruction System with Bandwidth Constraint
Involved Video Calling System: Skype and Wechat
This survery is focus on the feedback for the naturalness of speech comparing with Natural Speech Reconstruction System, Skype and Wechat.

Procedure

Before this conducting survey, you will be given an observation and listening testing on these three systems. You need to focus on the voice and speech play by the system. After you finish observing and listening the output from all the systems, you need to put a score for those three systems based on the naturalness of the speech. You are required to complete ALL the sections. The observation and listening testing will take approximately 5 to 10 minutes to complete. This survey will take approximately 2 to 3 minutes to complete.

Questions

Naturalness Score for Skype *

	1	2	3	4	5	6	7	8	9	10	
Low	○	○	○	○	○	○	●	○	○	○	High

Naturalness Score for Wechat *

	1	2	3	4	5	6	7	8	9	10	
Low	○	○	○	○	○	○	●	○	○	○	High

Naturalness Score for Natural Speech Reconstruction System *

	1	2	3	4	5	6	7	8	9	10	
Low	○	○	○	○	○	●	○	○	○	○	High

This content is neither created nor endorsed by Google.

Google Forms

Appendix C Transcript of Voice model dataset

01 I am a movie fanatic.

02 When friends want to know what picture won the Oscar in 2001 or who voiced Optimus Prime in Transformers, they ask me.

03 However, my buddies have stopped asking me if I want to go out to the movies.

04 While I love movies as much as ever, I find it more enjoyable to wait for a movie's release on Netflix because of the inconvenience of going out, the temptations of the concession stand, and the behavior of some patrons.

05 First of all, just getting to the theater presents difficulties.

06 Leaving a home equipped with an HDTV and surround sound isn't attractive on a cold or rainy night.

07 Even if the weather cooperates, there is the hassle of looking for a parking space and the lines.

08 There is also the worry of whether you and your friends will get all your seats together.

09 Although none of these hindrances are insurmountable, it's much easier to stay seated on your sofa.

10 Second, the theater offers tempting snacks that I don't really need.

11 At home I can control myself because there is no ice cream in the freezer, we don't have sodas in the fridge, and my snacks tend to be healthy, like fruits, nuts, and juices.

12 At the movies, even if I only buy a Diet Coke, the smell of fresh popcorn dripping with butter soon overcomes me.

13 And what about the nachos with cheese and the Snickers and M&M's?

14 I'm better off without all those temptations.

15 Finally, some of the other patrons are even more of a problem than the concession stand.

16 Little kids race up and down the aisles, making noise.

17 Teenagers try to impress their friends by talking back to the actors on the screen or otherwise making fools of themselves.

18 Some adults aren’t any better, commenting loud enough to reveal plot twists that are supposed to be a secret until the movie’s end.

19 What am I doing here, I ask myself.

20 After arriving home from the movies one night, I decided I had had enough.

21 I was not going to be a moviegoer anymore.

22 I was tired of the problems involved in getting to the theater, resisting unhealthy snacks, and dealing with the patrons.

23 The next day, I arranged to have premium movie channels added to my cable TV service, and I got a Netflix membership.

24 I may now see movies a bit later than other people, but I’ll be more relaxed watching box office hits in the comfort of my own living room.

25 Man's best friend has historically been considered a dog.

26 But dogs are not the only animal friend whose camaraderie people enjoy.

27 For many people, a cat is their best friend.

28 Despite what dog lovers may believe, cats make excellent house pets because they are good companions, they are civilized members of the household, and they are easy to care for.

29 Cats are good companions.

30 Many cats are affectionate.

31 They will snuggle up and ask to be petted or scratched under the chin, and who can resist a purring cat?

32 If they're not feeling affectionate, cats are generally quite playful.

33 They love to chase balls and feathers or just about anything dangling from a string.

34 And when they’re tired from chasing laser pointers, cats will curl up in your lap to nap.

35 Cats are loyal housemates.

36 Cats are also civilized housemates.

37 Unlike dogs, cats don’t bark or make other loud noises.

38 Most cats don't even meow that often.

39 Cats don't usually have accidents.

40 Mother cats train their kittens to use the litter box, and most cats will use it without fail from that time on.

41 Cats do have claws, but a tall scratching post in a favorite cat area of the house will often keep the cat content to leave the furniture alone.

42 Compared with other pets, cats are actually quite polite.

43 Cats are easy to care for.

44 They don’t have to be walked because they get plenty of exercise in the house as they play.

45 Even cleaning their litter box can be a quick, painless procedure.

46 Cats also take care of their own grooming.

47 Bathing a cat is almost never necessary because under ordinary circumstances cats clean themselves.

48 Cats are so easy to care for they can be left home alone for a few hours without fear.

49 Cats are low maintenance, civilized companions.

50 People who have small living quarters or less time for pet care should appreciate these characteristics of cats.

FYP2

ORIGINALITY REPORT

5%	3%	3%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.cse.dmu.ac.uk Internet Source	1%
2	Christelle Al Hasrouty, Cristian Olariu, Vincent Autefage, Damien Magoni, John Murphy. "SVC Videoconferencing Call Adaptation and Bandwidth Usage in SDN Networks", GLOBECOM 2017 - 2017 IEEE Global Communications Conference, 2017 Publication	1%
3	eprints.utar.edu.my Internet Source	<1%
4	www.cs.binghamton.edu Internet Source	<1%
5	cdn-lnx1.nwu.ac.za Internet Source	<1%
6	www.rroij.com Internet Source	<1%
7	airccse.org Internet Source	<1%

Universiti Tunku Abdul Rahman			
Form Title : Supervisor's Comments on Originality Report Generated by Turnitin for Submission of Final Year Project Report (for Undergraduate Programmes)			
Form Number: FM-IAD-005	Rev No.: 0	Effective Date: 01/10/2013	Page No.: 1of 1

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY

Full Name(s) of Candidate(s)	
ID Number(s)	
Programme / Course	
Title of Final Year Project	

Similarity	**Supervisor's Comments** **(Compulsory if parameters of originality exceeds the limits approved by UTAR)**
Overall similarity index:__________% **Similarity by source** Internet Sources: _________________% Publications:____________________ % Student Papers:__________________ %	
Number of individual sources listed of more than 3% similarity: __________	
Parameters of originality required and limits approved by UTAR are as Follows: **(i) Overall similarity index is 20% and below, and** **(ii) Matching of individual sources listed must be less than 3% each, and** **(iii) Matching texts in continuous block must not exceed 8 words** ***Note: Parameters (i) – (ii) shall exclude quotes, bibliography and text matches which are less than 8 words.***	

Note: Supervisor/Candidate(s) is/are required to provide softcopy of full set of the originality report to Faculty/Institute

Based on the above results, I hereby declare that I am satisfied with the originality of the Final Year Project Report submitted by my student(s) as named above.

____

Signature of Supervisor

Name:__________________________

Date: ___________________________

____

Signature of Co-Supervisor

Name:__________________________

Date: ___________________________

UNIVERSITI TUNKU ABDUL RAHMAN

FACULTY OF INFORMATION & COMMUNICATION TECHNOLOGY (KAMPAR CAMPUS)

CHECKLIST FOR FYP2 THESIS SUBMISSION

Student Id	
Student Name	
Supervisor Name	

TICK (√)	**DOCUMENT ITEMS** Your report must include all the items below. Put a tick on the left column after you have checked your report with respect to the corresponding item.
	Front Cover
	Signed Report Status Declaration Form
	Title Page
	Signed form of the Declaration of Originality
	Acknowledgement
	Abstract
	Table of Contents
	List of Figures (if applicable)
	List of Tables (if applicable)
	List of Symbols (if applicable)
	List of Abbreviations (if applicable)
	Chapters / Content
	Bibliography (or References)
	All references in bibliography are cited in the thesis, especially in the chapter of literature review
	Appendices (if applicable)
	Poster
	Signed Turnitin Report (Plagiarism Check Result - Form Number: FM-IAD-005)

*Include this form (checklist) in the thesis (Bind together as the last page)

I, the author, have checked and confirmed all the items listed in the table are included in my report. ________________________ (Signature of Student) Date:	Supervisor verification. Report with incorrect format can get 5 mark (1 grade) reduction. ________________________ (Signature of Supervisor) Date: