AUTONOMOUS LANGUAGE PROCESSING FOR BUSINESS SOLUTIONS

KOR JIA LI

A project report submitted in partial fulfilment of the requirements for the award of Bachelor of Science (Honours) Actuarial Science

Lee Kong Chian Faculty of Engineering and Science
Universiti Tunku Abdul Rahman

September 2021

DECLARATION

I hereby declare that this project report is based on my original work except for citations and quotations which have been duly acknowledged. I also declare that it has not been previously and concurrently submitted for any other degree or award at UTAR or other institutions.

Signature : 

Name : Kor Jia Li

ID No. : 17UEB05353

Date : 23 August 2021

APPROVAL FOR SUBMISSION

I certify that this project report entitled **"AUTONOMOUS LANGUAGE PROCESSING FOR BUSINESS SOLUTIONS"** was prepared by **KOR JIA LI** has met the required standard for submission in partial fulfilment of the requirements for the award of Bachelor of Science (Honours) Actuarial Science at Universiti Tunku Abdul Rahman.

Approved by,

Signature	:	
Supervisor	:	Dr. Tan Wei Lun
Date	:	8 SEPTEMBER 2021

ACKNOWLEDGEMENTS

I would like to thank everyone who had contributed to the successful completion of this project. I would like to express my gratitude to my research supervisor, Dr. Tan Wei Lun for his invaluable advice, guidance and his enormous patience throughout the development of the research.

In addition, I would like to express my gratitude to my loving parent and friends who had helped me and given me encouragement when I faced problem in doing this project.

ABSTRACT

A speech analytics solution worked with the combination of speech recognition and Natural Language Processing (NLP). It converted spoken sentences into written words by using Python Programming and with the help of Google Cloud Speech To Text API. Speech recognition steps included receiving "speech" either through microphone or audio files firstly. Then, the “speech” converted from physical sound into an electrical signal. The electrical signal was then being converted into digital data using an analogue-to-digital converter. Lastly, a model was used to convert the audio into text once it has been digitized. NLP helped a computer to understand languages spoken by humans. It was explained as an automated way of analysing the written text by following some theories and technologies. In the business area, speech analytics were used to make predictions and developed an understanding of the clients’ metrics. In this study, we focused on the languages such as Malay languages and mixed languages which were commonly used in Malaysia. Most of the call recordings data that used were basically containing these two languages. As Malay and mixed languages were not the worldwide languages, it increased the difficulty of developing a speech analytics solution that converted these two languages accurately into written text. Therefore, we expected that the results of this research improved the accuracy of speech analytics solutions so that it increased the efficiency of the insurance company in dealing with their clients. The accuracy of the speech analytics solutions in converting the spoken word into written text was investigated with Word Recognition Rate and an accuracy scale table used as a reference. There were two factors such as “Time Cut Point” and audio’s speed, being investigated in order to determine whether it would bring any effect towards the accuracy of text transcription. Different “Time Cut Points” and audio’s speed used in manipulating the data. Both factors were analysed together in a combination form. The best combination was chosen for both evaluation methods (WRR and accuracy scale table).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS / ABBREVIATIONS

S	Number of substitutions
D	Number of deletions
I	Number of insertions
N	Number of words
H	Number of words recognized correctly

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Background

Speech analytics solution was built by the combination of speech recognition and natural language processing (NLP). Speech recognition steps included receiving "speech" either through microphone or audio files firstly. Then, the “speech” converted from physical sound into an electrical signal. The electrical signal was then being converted into digital data using an analogue-to-digital converter. Lastly, a model was used to convert the audio into text once it has been digitized. Speech recognition is also called Speech to Text as well. Speech to Text was a software that can transcribe the voice into text. It was a unique and autonomous application nowadays because it was very convenient and effective in converting audio files into written words. This software had become popular worldwide, especially for those who need to generate content with long descriptions. This software was definitely handy in transcription since it was used automatically, and thus, it eventually increased working efficiency and productivity.

NLP helped a computer to understand languages spoken by humans. It was explained as an automated way of analysing the written text by following some theories and technologies. NLP aimed to be human-like language understanding. NLP was able to paraphrase the text, changed the text into different languages, identified the answer for questions related to the text content, and thus, made some conclusions based on the text. NLP had two focuses such as language processing and language generation. Language processing was explained as “reader” or “listener” as it analysed the languages of the text and then outputted a meaningful and relevant representation. Meanwhile, the language generation acted as “writer” or “speaker” as it referred to language production from the relevant representation. It required to decide what needs to be generated at what time and what point that was suitable in the interaction. Speech recognition was a software system that can transcript audio into text based on a set of “dictionaries”. Hence, NLP acted as a dictionary containing grammar and vocabulary that used by the speech

analytics. The combination of NLP and speech analytics increased the accuracy by training a model using machine learning and the model functioned to understand the calling purpose rather than matching a group of words. However, it took a longer time and resources to develop the training model. To develop a high accuracy speech analytics solution, we needed the help of NLP. The function of the speech analytics solution was to recognize the "speech" and converted it into written texts. Hence, it was just like inputted a continuous speech and get an output of written text.

There were many speech analytics business use cases around the world. In the business area, speech analytics were used to make predictions and developed an understanding of the clients' metrics such as Client Experience metrics, Client Satisfaction, Net Promoter Score and Client Effort. By using speech analytics solution, every client's voice tone and spoken words were being analysed and measured in order to make predictions on the client satisfaction level and NPS scoring was received from the client based on the call. Every call's satisfaction level from the clients would be obtained with the application of speech analytics. Besides, speech analytics also used as a tool to improve the performance of the customer service department staff. Since the speech analytics used to listen to every single call, it ensured that the staff followed the rules and processes when dealing with the client in the call. Speech analytics defined the keywords and phrases, then analysed their languages and voice tone for sentiment analysis, and eventually get scores based on various criteria for every dial in or dial out calls. By having this, the staffs improved their communication and interaction with the clients after their defect was identified by the speech analytics. The application of speech analytics solution was also used to determine the sentiment and emotion of both parties in the call as mentioned. Both parties' sentiment and emotion were detected by analysing their interaction. The analysis showed the results of how high the emotion level such as level of anger, happiness, and sadness of the people in the conversation. Not surprising if the speech analytics solution was used to detect fraud. Most insurance companies used this to minimize the number of fraud cases. Speech analytics was used to analyse the call recordings as discussed. Hence, it gave the outputs of transcription of text

and some interaction analytics. By using all the outputs given, a supervised learning prediction model was built to predict either the call was a fraud call or a non-fraud call. By doing so, the unwarranted pay-out claims were rejected by the insurance company to those unqualified policyholders.

In this study, we focus on the languages such as Malay languages and mixed languages which were commonly used in Malaysia. Most of the call recordings data used in this study were basically containing these two languages. As Malay and mixed languages were not worldwide languages, it increased the difficulty of developing a speech analytics solution that converted these two languages accurately into written text. Therefore, we expected that the results of this research improved the accuracy of speech analytics solutions so that it increases the efficiency of the insurance company in dealing with their clients.

1.2 Problem Statement

In this study, the problem that we are addressing is the transcription's inaccuracy in detecting the Malay language and mixed language. As there were no many speech recognition APIs developed based on Malay or mixed languages, therefore the accuracy becomes not so preferable. This inaccuracy is important to be addressed in a country like Malaysia as it is a norm in this country to mix languages. As the user base and data collection are mostly Malaysians addressing these languages.

Everyone has different ways of speaking. They could speak in various kinds of pronunciations or accents could have slow or fast speed and low or high frequency when they were talking to others. All these led to a problem that the API cannot capture their spoken sentence properly and then converted it into text inaccurately, which lower the accuracy of the transcription.

1.3 Objectives

The objectives of this study are as below:

1. To determine the factors that affect the accuracy of the Google API on the text transcription in Malay and mixed (Rojak) languages, with the use of speech analytics solution.
2. To investigate the best combination of the various time cut point and audio's speed that gives the highest accuracy of the transcription.

1.4 Significance

The significance of this study was to increase the speech analytics solution accuracy of converting the Malay and mixed languages which were not commonly used in the world. Developing the speech analytics solution can provide some other business solutions to the insurance company. The business solution is it can be used to speed up the call recordings quality checking. Besides, it can identify the client calling purpose and then speed up the progress of the client request because it may reduce the time taken to deal with the client problems.

This study assisted insurance companies in Malaysia to increase their efficiency in dealing with clients' issues and requests. The developed speech analytics solution can be used to boost client engagement. Hence, it offered a better client service and enhanced business outcomes for insurance companies in Malaysia.

1.5 Scope

Sun Life Malaysia Assurance Berhad was a life insurance and family takaful provider. We only focused on the clients who requested to cancel their insurance policy from Sun Life Malaysia Assurance Berhad. The beneficial parties belonged to the customer service team and respective teams related to insurance policy cancellation.

The accuracy of the speech analytics solutions in converting the spoken word into written text was investigated. The developing tools of the speech analytics solution were Python Programming languages and some packages that converted the voice into text. The packages used under the

Python module was known as SpeechRecognition. There were the other two online tools such as Google Speech To Text API and Google Recognizer. All of the packages or online tools were being tried to determine each accuracy in converting voice to text. High accuracy and more dynamic packages or tools were used in developing the speech analytics solution. With a suitable speech analytics solution, the accuracy could be improved.

CHAPTER 2

LITERATURE REVIEW

2.1 Reference List

A total of 11 references were used for literature review part. All of them sourced from the journals, online websites and the articles. Table 2.1 shows the references used in this project.

Table 2.1: Reference List.

Reference list format	In-Text Citation
Amir, G. and Murtaza, H., 2014. Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management,* 35(2015), pp. 137-144.	(Amir and Murtaza, 2014)
Beth, W., 2019. *Word Error Rate Mechanism, ASR Transcription and Challenges in Accuracy Measurement*. [Online]. Available at: <https://www.gmrtranscription.com/blog/word-error-rate-mechanism-asr-transcription-and-challenges-in-accuracy-measurement>. [Accessed 24 July 2021].	(Beth, 2019)
Bogdan, I., 2019. Evaluating Google Speech-to-Text API's Performance for Romanian e-Learning Resources. *Information Economică*, 23(1), pp. 19-20.	(Bogdan, 2019)
Chowdhury, G.G., 2003. Natural language processing. *Annual review of information science and technology*, *37*(1), pp. 51-89.	(Chowdhury, 2003)
James, F.A., n.d. *Natural Language Processing*. [Online]. Available at: <https://dl.acm.org/doi/pdf/10.5555/1074100.1074630>. [Accessed 2 March 2021].	(James, n.d.)
Juang, B. H. and Lawrence, R. R., 2004. *Automatic Speech Recognition – A Brief History of the Technology Development.* [Online]. Available at: <https://folk.idi.ntnu.no/gamback/teaching/TDT4275/literature/juang_rabiner04.pdf>. [Accessed 1 March 2021].	(Juang and Lawrence, 2004)
Liddy, E.D., 2001. In Encyclopedia of Library and Information Science, Marcel Decker. Inc.- Natural Language Processing.	(Liddy, 2001)
Santosh, G., Bharti, G. and Pravin, Y., 2010. A Review on Speech Recognition Technique. *International Journal of Computer Application (0975-8887),* 10(16), pp. 16-24.	(Santosh, Bharti and Pravin, 2010)

Table 2.1 (Continued)

Shaghayegh, E., 2019. *Word Error Rate (WER) for Recognition of Natural Interactions.* [Online]. Available at: <https://init.cise.ufl.edu/2018/04/03/word-error-rate-wer-for-recognition-of-natural-interactions/>. [Accessed 24 July 2021].	(Shaghayegh, 2018)
Scott, S. and Chung, Q.B., 2018. Making a case for speech analytics to improve customer service quality: Vision, implementation, and evaluation. *International Journal of Information Management,* 45(2019), pp. 223-232.	(Scott and Chung, 2018)

2.2 Literature Review

Speech analytics did analyse and extraction of the information from the audio data which were unstructured. Originally, speech analytics was known as audio analytics, it transformed its name into speech analytics when it applied to the languages spoken by the human. However, these two terms we always used in vice versa as it was a technique that applied to human spoken audio. Speech analytics application was popular and mostly used in the customer call centre or customer service centre, and also the healthcare. In customer service centre, speech analytics helped to do an analysis of call recordings, which then improved the experience of the customer, did evaluation on the performance of the agents, increased turnover rates of sales, monitored and controlled the compliance with various policies such as privacy policies, determined the problems of product and services, and get more understanding towards customer behaviours. It was used for the real-time calls as well, by giving recommendations based on the customer's historical interaction, and eventually gave the agents some feedback or ideas on how to communicate with the customers efficiently. In healthcare, it supported some diagnosis and treatment of particular medical conditions that bring impacts to those patient's communication patterns such as a patient that undergoes depression or patient with schizophrenia (a mental disorder that interprets the reality in abnormal perspectives). Besides, the speech analytics also functioned to do analysis on the baby's cries, which interpreted the health and emotional condition of the baby from his or her cries. Speech analytics had two technological ways such as transcript-based and phonetic-based approaches. The transcript-based

approaches were also called as Large-Vocabulary continuous speech recognition (LVCSR) systems. LVCSR had two-phase process, the first phase was the system trying to convert the speech content of the audio using automatic speech recognition algorithms by trying to do matching between the sounds and the words, and then the words were identified according to a dictionary which was predefined. If there were no correct words found, a similar word was returned as output, which was an index file that can be searched and contained the info of sequence of the words spoken in the audio file. While, in the second phase, a method known as standard text-based were applied to identify the search label in the index file. The phonetic-based approaches were worked with the phonemes (different and unique sound units in a particular language that differentiate the words with other words, for example, the phonemes of "k" and "b" distinguish the "cat" and "bat" meaning) or sounds. This system had two phases, indexing and searching. In the first phase, the phonetic-based system was trying to translates the speech input into a phonemes sequence, which was different compare with the transcript-based system that converted into a word sequence. While in the second phase, output from the first phases was searched by the system for the representation of phonetic of the search terms (Amir and Murtaza, 2014).

The general terms of speech recognition were a human-computer interface, speech processing, pattern recognition, modelling technique and signal processing. It was defined as the conversion signal of the speech into a word sequences by implementing means Algorithms as a program of a computer. In the 1960s, computer scientists had done some research on how a computer recorded, analysed, and then understood the speech of human languages. In the 1980s, the first system that was used to analyse human speech developed, but it had its limitations. The languages of spoken had dominated human communication. Thus, it was normal that humans expected to have speech interfaces with the computer. There were many types of speech that had their own type of utterances and being recognized by the speech recognition system. The type of speeches were isolated words, connected words, continuous speech and spontaneous speech. Speech recognition had four working stages such as analysis, feature extraction, modelling and testing.

For analysis, there were three analysis techniques such as segmentation analysis, sub-segmental analysis and suprasegmental analysis. The difference between them was in terms of frame size and time taken to shift, and also the content extracted. For feature extraction, it had to deal with the issues in categorization which aimed to reduce the input vector dimensionality but remained the signal discriminating power. There were some feature extraction methods such as principal component analysis (PCA), linear discriminate analysis (LDA), independent component analysis (ICA) and etc. All of the methods had their own property and implementation procedure. For modelling, it aimed to generate models of speakers. There were some modelling used to recognize the speech, such as acoustic-phonetic, pattern-recognition, template-based, dynamic time warping, knowledge-based, statistical-based, learning-based, artificial intelligence and stochastics approaches. For testing or matching, there were different types of matching as well which included whole-word and sub-word matchings. (Santosh, Bharti and Pravin, 2010).

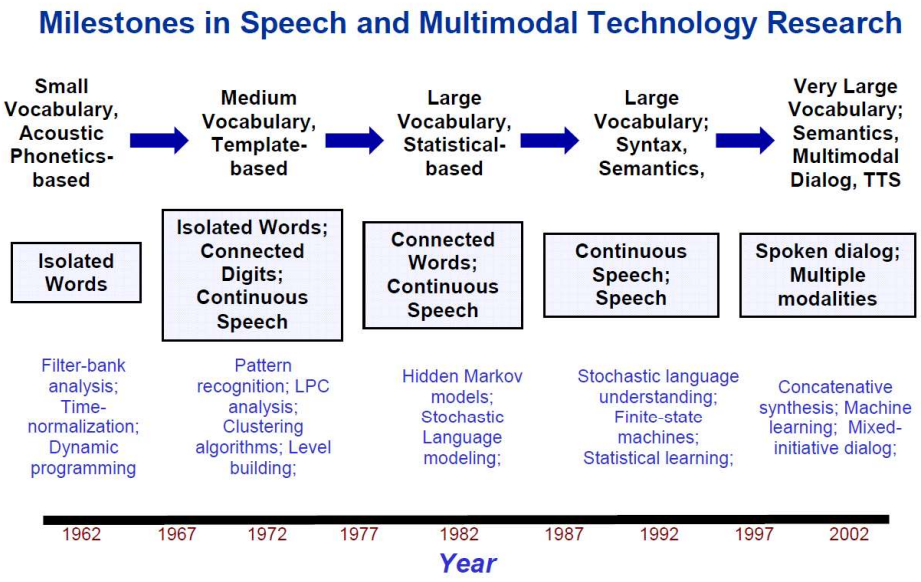


Figure 2.1: Speech Recognition Technology Development Timeline.

Figure 2.1 showed the timeline of the development of speech recognition technology from the 1960s to nowadays. In the 1960s, the technology only recognized 10 to 100 vocabularies of isolated words using

acoustic phonetics based approaches. This technology is developed by using filter-bank analysis, time normalization and dynamic programming. In the 1970s, the vocabularies recognized were around 100 to 1000 isolated words, connected digits and continuous speech using template-based approaches. Pattern recognition, LPC analysis, clustering algorithms and level building were used to increase the capability of this technology in recognizing the vocabularies. In the 1980s, large vocabularies such as connected words and continuous speech, around 1000 to unlimited were recognized using statistical-based approaches. Hidden Markov models and stochastic language modelling were being used to improve the performance. In the 1990s, a speech recognition system was able to detect large vocabularies of continuous speech or speech, with an unlimited language model. Stochastic language understanding, finite-state machines and statistical learning were applied to develop the technology. In the 2000s, the system recognized very huge vocabularies with full semantic models and text-to-speech integration synthesis systems and multi-modal inputs. Concatenative synthesis, machine learning and mixed-initiative dialog were used in improving the speech recognition systems. After so long duration of research, this technology was introduced in the marketplace and became very famous among the beneficial party (Juang and Lawrence, 2004).

Natural Language Processing (NLP) was an automated approach to do text analysis based on some theories and technologies. It was a range of computational techniques which are theoretically motivated, used for analysing and interpreting the texts that occurred naturally at only one or more than one level of analysis of linguistic, in order to accomplish human-like language processing for many tasks or applications. Its automated techniques must emphasize as there were many other techniques used to achieve a specific analysis of language. While the texts that occurred naturally could be explained as “Naturally occurring texts”, it could be any mode, any genre and any language, could be written or oral, and the only requirement was it must be the human languages used for communication. Human-like language processing interpreted that NLP was treated as Artificial Intelligence (AI) discipline. The goal of NLP was to achieve human-like language processing or

human-like language understanding. NLP originated from Natural Language Understanding (NLU), and NLU had some goals such as paraphrased a text input, translated it into other languages, answered some content related questions and made conclusions from the text content. NLP is able to achieve the goals mentioned but only except the last goals (make a conclusion). There were some NLP applications such as information retrieval, information extraction, question-answering, summarization, machine translation, and dialogue systems. Information retrieval retrieved the information from the text, information extraction put focus on tagging, recognition and extraction of key information, and question-answering produced a list of possible information to answer the user's questions (Liddy, 2001).

Natural Language Processing was research and application that study the way of how the computer understands natural languages in terms of text or speech. NLP aimed to make the computer understood the languages so that it could carry out some tasks by itself. NLP's foundations lie in some disciplines such as linguistic, computer and information sciences, mathematics, artificial intelligence, electrical and electronic engineering, robotics, and also psychology (Chowdhury, 2003).

Natural Language Processing was referred to as a computer system that can do analysis and try to understand human languages and eventually produced other human languages. The input of NLP could be spoken words, written text, or input of keyboards. NLP's task could be a translation of the input into other languages, understanding and representing the text content, database builder, summary generator and etc. All natural languages involved simple lexical ambiguity (a word can be used as a noun or verb), semantic ambiguity (a word that had more than ten meanings in every dictionary), referential ambiguity (a sentence that does not mention clearly about the target), structural or syntactic ambiguity (a word that does not mention precise action) and pragmatic ambiguity (a question that either need to answer yes or no, or could be a request). All these ambiguities are found at all levels of problems increased the difficulty in developing NLP (James, n.d.).

Google Speech Recognition was introduced in the year 2008. It was launched for Apple iPhone in the form of the Google Voice Search. Google has stored huge quantities of data in its own server, and it also had some machine learning algorithms. Thus, with the large data and ready algorithms, Google developed its first large scale Automatic Speech Recognition (ASR) system. This Google ASR system was considered the launch of the modern ASR algorithms. In the year 2017, Google Cloud Speech to Text API was introduced to the public. This API supported many languages, around 90 languages in the first launch, and more languages being add on until it could support 150 languages nowadays. Google Cloud STT API not only could perform ASR, but however, it also provided other extra features such as real-time streaming, automatic punctuation, and auto-detection of languages (Bogdan, 2019).

The traditional way to improve client service quality was done by observing a prefix number of agents' interactions with the client, then a form of evaluation would be filled. If there was any improvement needed for the agent, training to the agent by face to face would be carried out. Nowadays, the way to improve the quality had changed due to the launch of call recordings and speech analytics. The assessment team was now able to not only monitored the agents, but they could also get some client-related information from the interaction (Scott and Chung, 2018).

2.3 Word Error Rate and Word Recognition Rate

Speech recognition performance was evaluated by calculating its accuracy and speed. For accuracy measure, Word Error Rate (WER) or Word Recognition Rate (WRR) were used while for the speed measure, it evaluated with calculated the time taken for a transcription to be done. The formula of WER was as below:

$$WER = \frac{S+D+I}{N} \tag{2. 1}$$

where

S = Number of substitutions

D = Number of deletions

I = Number of insertions

N = Number of words

WRR was the opposite of WER, its formula was as below:

$$WRR = 1 - WER = \frac{N-S-D-I}{N} = \frac{H-I}{N} \quad (2.2)$$

where $H = N - (S + D)$ is the number of words recognized correctly. Both accuracy measures were applicable in determining the accuracy of speech recognition (Santosh, Bharti and Pravin, 2010).

Word Error Rate had some limitations which it neglected the words' importance, evaluated all the errors in a document with the same score, and this actually was not applicable in the real world as the words were important because some of them were the keywords in a transcription. If WER just ignored it, the accuracy would not be accurate enough as it would be led to a low understanding of the transcription. Besides, WER also ignored the speaker labels, for example, the first sentence was spoken by person A, the second sentence was spoken by person B, in way of calculating WER, we would not match the sentence with the person but only focused on counting the words recognized correctly (Beth, 2019).

Word Error Rate was derived from Levenshtein Distance, an algorithm that calculated the minimum edit distance between both strings, mainly two strings. Word Error Rate could be used in determining the recognizer accuracy, and it was applicable to text or speech. Word Recognition Rate could be computed by getting the inverse of Word Error Rate (Shaghayegh, 2018).

CHAPTER 3

METHODOLOGY AND WORK PLAN

3.1 Flowchart

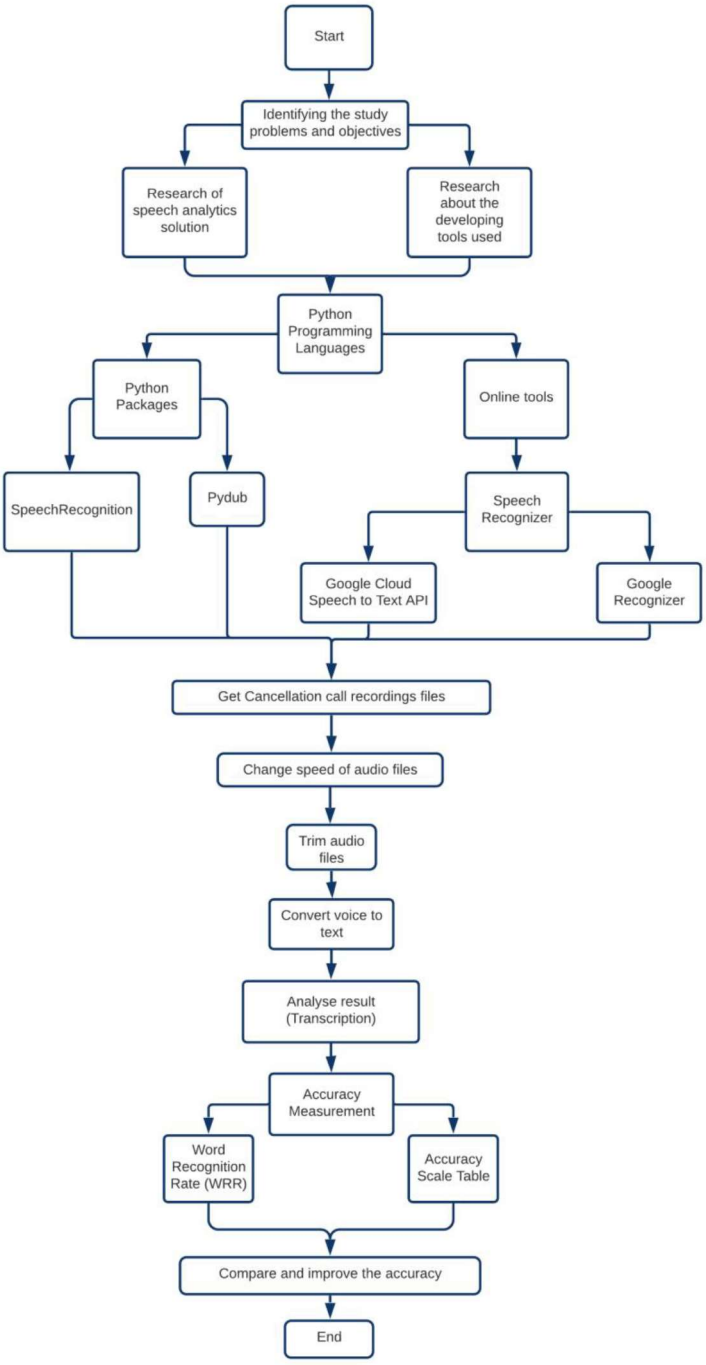


Figure 3.1: Project's Flow.

3.2 Research Methodology

This study was aimed to develop a high accuracy speech analytics solution that could convert spoken words into written texts. Qualitative data were used and sourced from Sun Life Malaysia Assurance Berhad. The data used were the call recordings related to policy cancellation. The data were considered secondary data as they were collected from the company mentioned. This study was to develop a high accuracy speech analytics solution that could provide business solutions to Sun Life Malaysia Assurance Berhad. The business solution could be an improvement of the efficiency in dealing with the clients' problems and issues. The main target of this study was the clients from the company itself especially those who decided to cancel their insurance policy. The call recordings related to policy cancellation were being analysed and used as the training data.

The analysis was done based on languages and words. Content analysis that could be done was categorized the text transcription output from the speech analytics solution and interpreted the meaning of the text itself. An experiment was not used as this study aimed to produce outputs related to languages and words, but not testing the relationship between the data variables. Validity was determined in this study as it referred to the accuracy of a measure, for example, the accuracy of the speech analytics solution in converting the spoken words into written text.

Generally, Speech recognition or Speech to Text (STT) was one of the computer science subfields that was able to recognize and transcript the spoken words into written text using computers. In STT, some models were important in speech recognition algorithms which were including acoustic modelling and language modelling. The acoustic model could turn the audio signals into phonetic representation, which was a representation that does not require to know and understand the spoken languages priorly. Language modelling is grouped in terms of domain knowledge of the words, sentence structures and grammar for the languages. The models were developed by using machine learning algorithms with the model based on probability.

Language models were functioned to introduce the knowledge of languages into the words to become text in speech recognition, by provided correct spelling words to the text context. The acoustic model not really could differentiate the spelling of the words as it is only based on sound. The words output from the acoustic model was less likely accurate as it depended on the probability distribution over various and distinct words, calculated by depending on the likelihood of the sequence of words output from the sound signal. In this case, the language model helped to increase the accuracy by outputting the correct spelling words. The common language model used was known as N-gram. N-gram was a continuous sequence of n items such as letters and words, from the given speech. It could be collected from the speech or text collection.

Hidden Markov models (HMMs) were being used in developing speech recognition systems. It had become a traditional automatic speech recognition solution as it was used with advances for speech recognition a long time ago. HMMs were convenient and effective for identifying the patterns across time. HMMs were applied due to a signal of speech could be looked as a signal of piecewise or short-time stationary, meaning if the signal divided into the time frames that were short enough, it could be interpreted accurately in each time frame as a summation of sinusoids in the existence of white noises. Speech could be interpreted as a stationary process on a scale that was short in time (ten milliseconds).

Based on the flowchart in Figure 3.1, this study began with doing some title related researches in order to increase the understanding of the title. The researches that being done were related to speech analytics solutions such as how to develop a speech analytics solution, and what kind of developing tools would be used? In the process of doing researches, it has been found that Python Programming Languages would be suitable and easy for a beginner to design and develop a speech analytics solution.

The speech analytics solution used some tools such as packages and modules in Python Programming and some online recognizers. For example, SpeechRecognition in Python worked with the recognizer in order to

recognize the spoken words and convert them into written text. There were a few recognizers available such as recognize_google(), recognize_google_cloud(), recognize_ibm() and recognize_wit(). All these recognizers required the connection of the internet. In this study, we would try to approach recognize_google_cloud(), which was known as Google Speech to Text API (Google STT). Google STT was an API powered by AI technologies from Google that could accurately transcript the speech into text. It was selected as the most suitable recognizer because of its various features. The features were including global vocabulary that supports 125 languages, could be applied in a real-time situation, consisting of speech adaptation which could configure speech recognition by providing hints or keys to improve the accuracy, noise robustness that could minimize the surrounding noises, could prefix languages code such as Malay and mixed languages in Malaysia and so on.

As all the recognizers had a limitation that the audio's duration could not be more than one minute, thus Pydub module in Python was used to trim the audio files. For example, if there was an audio file in which duration was two minutes, Pydub would be used to trim the audio into two short audios, each one minute in duration to ensure that it could pass successfully to the recognizer for conversion purposes. After the conversion, we validated the accuracy of the text transcription by listened and compared the text transcription with the call recordings. A way to improve the accuracy was figured out as the accuracy, in the beginning, was too low.

Speech analytics solution could use to analyse the call recordings of the clients in order to increase the understanding of the clients' needs. It could be explained as a process of interpreting the conversation in the call recordings to obtain more information related to the clients. By doing so, the company could further improve their way of communication with the clients and also for future interaction purposes. The speech analytics solution was designed in such a way that it contained an element of automatic recognition of speech, which means it could be used to detect the spoken words and then converted them to written text.

There are some other future use cases of speech analytics solutions developed such as it could use to identify the calling purpose of the client. Besides, speech analytics solution also can be used to speed up quality checking in Sun Life Malaysia Assurance Berhad. This is because only a few call recordings are being used for quality checking, however, with the use of the solution developed, the size of quality checking can be increased due to the automation.

3.3 Work Schedule (Gantt Chart)

Table 3.1: Project I Work Schedule.

Task	Weeks													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Project Title Registration	■													
Biweekly Report Submission			■		■		■		■		■			
Literature Review Study		■	■	■	■	■	■	■	■	■	■			
Proposal Writing		■	■	■	■	■	■							
Proposal Submission							■							
Interim Report Writing							■	■	■	■	■	■		
Interim Report Submission												■		
Project Presentation													■	

Table 3.2: Project II Work Schedule.

Task	Weeks													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Research Materials Collection	■	■	■	■	■	■	■	■	■					
Research Materials Analysis	■	■	■	■	■	■	■	■	■	■	■			
Literature Review Study		■	■	■	■	■	■	■	■					
Final Report Writing		■	■	■	■	■	■	■	■	■	■	■		
Final Report Submission												■		
Project Presentation													■	

3.4 Accuracy Measurement

Word Recognition Rate would be used in evaluating the performance or results of the transcription. In addition, the accuracy scale table designed

would be used as a reference as well in assessing the accuracy of the text transcription.

Word Recognition Rate could be calculated using a formula which was as below:

$$WRR = 1 - WER = \frac{N-S-D-I}{N} = \frac{H-I}{N} \quad (3.\ 1)$$

where

N = Number of words

$H = N - (S + D)$ = Number of words recognized correctly

The higher the WRR calculated, the better the performance and the higher the accuracy of the text transcription.

The accuracy scale table would evaluate the accuracy by assessing "Text Understanding", "Wrong Spelling of Word" and "Completeness of Transcription" of the transcription, which its measure and criteria were shown as below:

Table 3.3: Accuracy Scale Table.

Scale parameter	**Accuracy Description**	**Text Understanding**	**Wrong Spelling of Word**	**Completeness of Transcription**
1	Very low accuracy	<20%	> 20 words	<20%
2	Low accuracy	20% - < 40%	15 – 20 words	20% - < 40%
3	Medium accuracy	40% - < 60%	10 – 15 words	40% - < 60%
4	High accuracy	60% - < 80%	5 – 10 words	60% - < 80%
5	Very high accuracy	> 80%	1 – 5 words	> 80%

Table 3.3 show 5 different scales for sentence accuracy in 3 categories. The scale of "1" indicated that the accuracy was very low, the text only provided less than 20% understanding, more than 20 words' spelling were

wrong, and the completeness of transcription was less than 20%. The next scale is “2”, indicating that the accuracy was low, which the text provided 20% to less than 40% understanding, 15 to 20 words’ spelling were wrong, and the completeness of transcription was 20% to 40%. The third scale is “3”, indicating that the accuracy was medium, which the text provided 40% to 60% understanding, 10 to 15 words’ spelling were wrong, and the completeness of transcription was 40% to 60%. Furthermore, scale “4”, indicated that the accuracy was high, which the text provided 60% to less than 80% understanding, 5 to 10 words’ spelling were wrong, and the completeness of transcription was 60% to 80%. The last scale, “5” indicated that the accuracy was very high, the text provided more than 80% understanding, only less than 5 words’ spelling were wrong, and the completeness of transcription was more than 80%.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Introduction

In this chapter, the results' analyse were done by manipulating the audio files. There were two different manipulations, and the first manipulation was trimming different duration or Time Cut Points. Another manipulation was related to trim different duration together with changing the speed of audio files at the same time to obtain different analyse and results.

4.2 Time Cut Point(s)

There were some factors that would be investigated in order to determine whether it would bring any effect towards the accuracy of text transcription. The first factor would be related to time, a term named "Time Cut Point" would be used, which was the duration of every single short audio file being cut. Different durations would be used in manipulating the data such as 59 seconds, 20 seconds, 10 seconds, and 5 seconds in this study.

Table 4.1: Original Text of "C9733930".

(...Bunyi phone kau tu tu tu tu tu...)	
Staff:	Selamat petang
Client:	Hello
Staff:	Selamat petang
Client:	Hello
Staff:	Boleh bercakap dengan cik Rohani Harista
Client:	Em saya
Staff:	Cik Harista, minta maaf berganggu ya. Saya buat panggilan daripada Sunlife Malaysia Assurance
Client:	Em hmm
Staff:	Okay. Ini berkaitan dengan satu notis permohonan penyerahan
Client:	En. Pembatalan tu kan
Staff:	Ya betul untuk pembatalan. Kalau saya lihat dekat sini cik ada lagi sekali pada 11 hari bulan hari tu hantarkan dokumen ataupun hantarkan emel tanpa dokumen cik. Cik ada terima belum borang nilai serahan yang kita hantar itu
Client:	Borang tu saya dah serahkan tapi saya tak ada saksi pun. Masa open tu tak da saksi pun buat
Staff:	Oh macam ini cik harista, untuk maklumat cik, borang tu maksudnya cik dah terima lah borang yang kita hantar dua kali ini
Client:	Ah I dah terima dan saya dah print out dah cuma belum isi lagi

Table 4.1 shows the first 59 seconds of the original text of the call recording file named “C9733930”. The call recording was manually listened to in order to type out the original text conversation between the staff and the client. The complete transcription could be referred to Appendix A. The first 59 seconds’ original text would be used as a reference in order to make comparison among Time Cut Point(s). The unit of time being used in Python code was in milliseconds, meaning 1 second would be inserted as 1000 milliseconds. The output of text transcription would be shown in a table form and saved as a Microsoft Excel File. The table would include each start time of the short audio file (cut using different Time Cut Point), followed by each end time in the second column, and lastly the text transcription in the third column.

Table 4.2: Time Cut Point (59 seconds) of "C9733930".

Start_Time	End_Time	Transcription
0	59000	bunyi phone kau tu tu tu tu tu tu tu tak maafkan aku ya Saya dah ada Buat panggilan daripada solat Perlis insuran CIDB berkata dengan satu Doctors permohonan penyerahan pembatalan tu kan dia betul-betul pembalasan lihat dekat sini cik cik ada lagi lagi sekali pada 11 haribulan hari tu hantarkan dokumen ataupun hantarkan email tanpa dokumen cik cik dah terima belum borang Biasalah orang dah hantar kat awak Tapi saya tak suka memaksa cik dah terima laporan kita dah hantar gaduh kali ni saya dah terima saya dah print out

From Table 4.2, it could be observed that the transcription was not so accurate as the original text in Table 4.1. The text understanding is only around 45% in this state. There were 11 words that were wrong, and there were listed in Table 4.3. The completeness of transcription only approximately 45% as there were some sentences that were not converted successfully such as “Selamat Petang. Hello. Selamat Petang. Hello. Boleh bercakap dengan cik Rohani Harista. Em saya. Cik Harista, minta” and “Borang tu saya dah serahkan tapi saya tak ada saksi pun. Masa open tu tak da saksi pun buat. Oh macam ini cik harista, untuk maklumat cik, borang tu maksudnya.” Overall, the accuracy scale for Time Cut Point (59 seconds) was “3”, which indicated that the accuracy was at a medium level. The complete transcription could be referred to Appendix B.

Table 4.3: Spelling Errors of Time Cut Point (59 Seconds) of “C9733930”.

Wrong Spelling	Correct Spelling
Solat	Sunlife
Malaysia	Perlis
CIDB	Ini
Berkata	Berkaitan
Doctors	Notis
Pembalasan	Pembatalan
Email	Emel
Boring	Borang
Biasalah	Nilai
Orang	Serahan
Gaduh	Dua

Table 4.4 shows that the transcription was not so accurate as of the original text in Table 4.1. The text understanding is only around 50% in this state. There were 14 words that were wrong, and there were listed in Table 4.5. The completeness of transcription only approximately 50% as there were some sentences that were not converted successfully such as “Selamat Petang. Hello. Selamat Petang. Hello. Boleh bercakap dengan cik Rohani Harista. Em saya. Cik Harista, minta” and “Oh macam ini cik harista, untuk.” Overall, the accuracy scale for Time Cut Point 20 seconds was “3”, which indicated that the accuracy was at a medium level. The complete transcription could be referred to Appendix C.

Table 4.4: Time Cut Point (20 seconds) of "C9733930".

Start_Time	End_Time	Transcription
0	20000	bunyi pun kata tu tu tu tu tu tu tak maafkan aku ya Saya dah ada buat
20000	40000	Pergilah daripada solat di Sinsuran pagi ini berkata dengan satu Doctors permohonan penyerahan pembatalan Bukan dia betul-betul pembatalan Kalau saya lihat dekat sini cakap lagi sekali pada 11 haribulan hari tu hantarkan dokumen ataupun hantarkan email tanpa dokumen cik
40000	60000	cik dah terima ke belum borang Biasalah orang dah hantar kan terus terang kan Tapi saya tak Tipulah tidak kasi tahu Untuk makluman isi borang tu maksud dia cik dah terima laporan kita dah hantar gaduh kali ni saya dah terima saya dah print out mudah

Table 4.5: Spelling Errors of Time Cut Point (20 Seconds) of “C9733930”.

Wrong Spelling	Correct Spelling
Pergilah	Panggilan
Solat	Sunlife
Malaysia	Di
Sinsuran	Assurance
Berkata	Berkaitan
Doctors	Notis
Cakap	Cik ada
Email	Emel
Biasalah	Nilai
Orang	Serahan
Makluman isi	Maklumat cik
Dia	Nya
Laporan	Lah borang
Gaduh	Dua

Table 4.6: Time Cut Point (10 seconds) of "C9733930".

Start_Time	End_Time	Transcription
0	10000	bunyi phone kata tut tut tut tut
10000	20000	datang sini boleh bercakap dengan Cik Rohani hasrita Saya minta maaf ganggu Ya saya dah dapat
20000	30000	lagilah daripada solat di Sinsuran pagi ini berkata dengan satu notis permohonan penyerahan pembatalan Bukan dia betul-betul
30000	40000	Kalau saya lihat dekat sini chat-chat ada lagi lagi sekali pada 11 haribulan hari tu hantarkan dokumen ataupun hantarkan email tanpa dokumen cik
40000	50000	cik dah terima ke belum borang Biasalah orang dah hantar kan terus terang kan Tapi saya tak tipu sama saya
50000	60000	oh baca dikasi tahu tu maklumat sikit borang tu maksud dia cik dah terima laporan kita dah hantar gaduh kali ni saya dah terima saya dah pening awak pun dah

In Table 4.6, it could be observed that the transcription was not so accurate as the original text in Table 4.1. The text understanding is only around 55% in this state. There were 15 words that were wrong, and there were listed in Table 4.7. The completeness of transcription only approximately 60% as there were some sentences that are not converted successfully such as “Selamat Petang. Hello. Selamat Petang. Hello.” Overall, the accuracy scale for Time Cut Point 10 seconds) is “3”, which indicated that the accuracy was at a medium level. The complete transcription could be referred to Appendix D.

Table 4.7: Spelling Errors of Time Cut Point (10 Seconds) of “C9733930”.

Wrong Spelling	Correct Spelling
Dah dapat lagilah	Buat panggilan
Solat	Sunlife
Di	Malaysia
Sinsuran	Assurance
Berkata	Berkaitan
Chat chat	Cik cik
Cakap	Cik ada
Email	Emel
Biasalah	Nilai
Orang	Serahan
Makluman sikit	Maklumat cik
Dia	Nya
Laporan	Lah borang
Gaduh	Dua
Pening awak	Print out

Table 4.8: Time Cut Point (5 seconds) of "C9733930".

Start_Time	End_Time	Transcription
0	5000	
5000	10000	bunyi pun kata tut tut tut tut
10000	15000	datang sini boleh bercakap lagi
15000	20000	Cik Rohani hasrita Saya minta maaf ganggu Ya saya dah ada buat
20000	25000	Pergilah daripada solat di Sinsuran pagi ini berkata dengan satu
25000	30000	Pesta Bunga dan penyerahan pembatalan tu kan dia betul-betul
30000	35000	Kalau saya lihat dekat sini chat-chat ada lagi lagi sekali pada 11 haribulan hari tu
35000	40000	hantarkan dokumen ataupun hantarkan email tanpa dokumen cik
40000	45000	cik dah terima ke belum borang Biasalah orang dah hantar kan terus
45000	50000	sudah hantar Tapi saya tak tipu siapa-siapa saya tak ada kat situ
50000	55000	oh baca dikasi tahu untuk maklumat cik borang tu masa dia cik dah tu
55000	60000	mana borang kita dah hantar gaduh kali ni saya dah terima saya dah print out mudah

Table 4.8 shows that the transcription was not so accurate as of the original text in Table 4.1. The text understanding is only around 60% in this state. There were 11 words that were wrong, and there were listed in Table 4.9. The completeness of transcription only approximately 60% as there were some sentences that are not converted successfully such as “Selamat Petang. Hello. Selamat Petang. Hello.” Overall, the accuracy scale for Time Cut Point 5 seconds was “4”, which the accuracy could be considered as high level. The complete transcription could be referred to Appendix E.

Table 4.9: Spelling Errors of Time Cut Point (5 Seconds) of “C9733930”.

Wrong Spelling	Correct Spelling
Bercakap lagi	Bercakap dengan
Solat	Sunlife
Di	Malaysia
Sinsuran	Assurance
Pesta bunga dan	Notis permohonan
Chat chat	Cik cik
Email	Emel
Biasalah	Nilai
Orang	Serahan
Masa dia	Maksudnya
Tu mana	Terima
Gaduh	Dua

4.3 Time Cut Point(s) and Speed(s)

Other than the first factor that related to time, there was another factor was undergo investigation which is related to the speed of audio, and there were 5 different speeds being tested, including 0.5X (0.5 times slower than normal speed), 0.75X, 1X (normal speed), 1.25X and 1.5X. Both factors (Time Cut Point and speed) would be analysed together in a combination form, for example, 5s with 0.5X, 5s with 0.75X and so on. For one call recording, it would have 20 combinations in total. In this study, a total of 30 call recording audio files were analysed.

As usual, by using Python code for speech recognition (convert spoken word into written text), the transcription would be output and being analysed. As mentioned, different combinations of time cut point and speed would be analysed. By taking one call recording as an example, the speed of the audio file would be modified in the beginning. Then, different time cut points would be manipulated for each audio file which was at a different speed. When the text transcriptions were obtained, the analysis would be started. There were two methods in evaluating the results as mentioned before, Word Recognition Rate (WRR) and accuracy scale table used as a reference. In this study, only the first minutes of call recording would be analysed.

The first step of doing the analysis was that the original text was manually listened to and typed in a conversation form, as it would be used as a reference in comparing and analysing those different combinations. Next, the number of correct words recognized in the data resulted from programming (text transcription) and the number of words in the original text would be calculated in order to compute WRR. Then, for the accuracy scale table, text understanding would be evaluated, words that have wrong spelling would be figured out, and the percentage of text completeness would be determined. The last step was to identify the best results of combination for WRR and accuracy scale table.

Each of the call recording audio files had its own naming as the data were collected from Sun Life Malaysia Assurance Berhad. The call recording named “C9735524” would be explained in detail on how the analysis was being done. Table 4.10 was the original text of “C9735524” that has been manually listened to and written in conversation format.

Table 4.10: Original Text of "C9735524".

Staff:	Helo Sun life Malaysia. Boleh saya bantu? Helo?
Client:	helo Assalamualaikum
Staff:	Mualaikum salam
Client:	Helo saya Nur Syahirah Binti Mahadzir. Pegawai CIMB
Staff:	Okay kenapa ya
Client:	Saya nak tanya pasal sun life
Staff:	Okey tentang apa tu
Client:	Okey saya nak cancel lah sun life ni
Staff:	Okey yang itu saya kena semak dulu boleh baca saya tak IC ke nombor polisi ke
Client:	Okey IC sembilan tujuh kosong lima eh sembilan tujuh kosong enam
Staff:	Okey
Client:	25
Staff:	Okey
Client:	385164
Staff:	5164 Okey sebentar ya Saya cuba semak dulu rekod yang dimasukkan dulu
Client:	ya macam mana plan ya
Staff:	macam mana
Client:	plan sun mozi shield
Staff:	bukan. saya semak dulu sebentar ya
Client:	Okey
Staff:	saya tengok kat sini di bawah nama Nur Syahirah Binti Mahadzir ya Nanti saya nak buat pengesahan sedikit boleh tak

Table 4.10 shows the first minute of the original text of “C9735524”, while the complete original text could refer to Appendix F. It would be used as a reference in order to make comparisons among the combination of time cut point(s) and speed(s). After that, by using Python, speech recognition would take place by converting the new combination from spoken word into written text. The output of text transcription would be then presented in a table that included the start time of the short audio file (cut using different Time Cut Point), followed by each end time and the text transcription in the last column. The time unit used is in milliseconds, where 1 second is represented as 1000 milliseconds.

Table 4.11: Combination of Time Cut Point of 5s and Speed of 1X of "C9735524".

Start_Time	End_Time	Transcription
0	5000	
5000	10000	Hello Hello assalamualaikum assalamualaikum
10000	15000	Syahirah ada Okey okey saya
15000	20000	kacau sangat lain Okey tentang Lepas tu Okey saya nak tulis
20000	25000	Okey yang itu saya kena semak dulu boleh baca saya tak IC ke nombor polisi ke
25000	30000	okey 705
30000	35000	Wassalam okey
35000	40000	385 1645 14 Okey sebentar ya
40000	45000	cuba semak dua ekor termasuk data nya
45000	50000	Lan bukan
50000	55000	saya semak dulu sebentar ya Okey saya tengok kat sini dia bawa nama Nur Syahirah
55000	60000	Terima kasih ya Nanti saya nak buat pengesan sedikit boleh tak

Table 4.11 shows the transcription of "C9735524" in a combination of time cut point 5 seconds and normal speed, while for the complete transcription could refer to Appendix G. The transcription in Table 4.11 was not as accurate as of the original text in Table 4.10. For WRR computation, there were a total of 56 words that were recognized correctly, and the number of words in the original text is 114 words. Hence, WRR was computed to be 49%. Next, by doing some analysis, the text understanding only understood around 30%, 8 words which were wrong in spelling, and all of them were tabulated in Table 4.12. The text completeness was approximately 30% because there were some sentences that were not converted such as "Sun life Malaysia. Boleh saya bantu? Helo?" and "Mualaikum salam. Helo saya Nur." Overall, the accuracy scale for the combination of 5s and 1X was "2", which indicated that the accuracy was at a low level.

Table 4.12: Spelling Errors of Time Cut Point 5s and Speed 1X of "C9735524".

Wrong Spelling	Correct Spelling
Dulu rekod yang dimasukkan dulu	dua ekor termasuk data nya
Di bawah	Dia bawa
Pengesahan	Pengesan

Table 4.13: Summary Result of "C9735524" for Time Cut Point of 5s.

C9735524 (First minute of recording) **Total words in original text = 114 words**					
Time cut point: 5s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	5	25	30	10	1
Wrong Spelling Word	10	18	8	17	0
Text completion (%)	10	30	30	15	2
Scale	1	2	2	1	1
Word Recognition Rate (WRR) (%)	33	52	49	25	5

By referring to Table 4.13, the results could be explained in this way. For example, 0.5 times of speed for “C9735524” with time cut point of 5 seconds, the understanding towards the text transcription only 5%, 10 words had wrongly spelt and 10% of the text were transcript completely if compared to its original text. With reference to the accuracy scale table, it was under Scale 1, which had very low accuracy. In calculating WRR, the final result achieved 33%. The complete summary result of “C9735524” could refer to Appendix R.

With the steps explained above, the analysis continued for the rest. Each call recording audio file had 20 combinations. There were a total of 30 call recording audio files with different combinations that would be analysed. The results of each combination for each call recording audio file were tabulated separately in a table form and they could be referred to Appendix H until Appendix AK.

Table 4.14: Summary of Best Combination for WRR & Accuracy Scale Table.

Call Recording Audio File	Scale	WRR
C9733930	1X & 5s	1X & 5s
C9814494	1X & 5s	1X & 5s
C10347935	1X & 5s	0.75X & 5s
C10206965	0.75X & 5s	0.75X & 5s
C10208770	0.75X & 5s	0.75X & 5s
C10237881	1X & 10s	0.75X & 5s
C10395406	1X & 5s	1X & 5s
C10415521	1X & 10s	0.75X & 5s
C10551221	1X & 10s	1X & 10s
C10804449	1X & 5s	1X & 5s
C9735524	1X & 10s	0.75X & 10s
C10209992	1X & 10s	1X & 10s
C10371808	1X & 5s	1X & 5s
C10411974	1X & 5s	1X & 5s
C10503847	1X & 5s	0.75X & 10s
C9760826	1X & 10s	1X & 10s
C9798109	0.75X & 20s	0.75X & 20s
C9802561	0.75X & 10s	0.75X & 10s
C9822062	0.75X & 5s	0.75X & 5s
C9838810	0.75X & 10s	0.75X & 5s
C9846248	1X & 5s	1X & 5s
C9855059	0.75X & 59s	0.75X & 59s
C9907261	0.75X & 10s	0.75X & 10s
C9935118	1X & 5s	1X & 5s
C9937436	1X & 5s	1X & 5s
C10052504	1X & 5s	1X & 5s
C10063278	1X & 5s	1X & 5s
C10077643	1X & 5s	1X & 5s
C10078155	1X & 5s	1X & 5s
C10081916	1X & 5s	1X & 5s

Table 4.14 is the summary of the best combination in WRR and accuracy scale table for all 30 call recording audio files.

Table 4.15: Best Combination of Each Audio Files for Accuracy Scale Table Measurement.

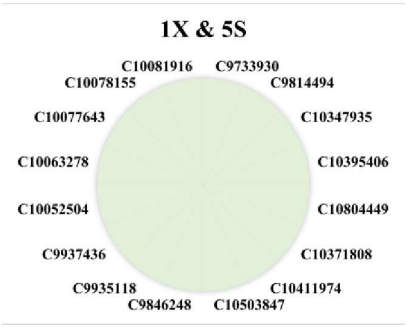
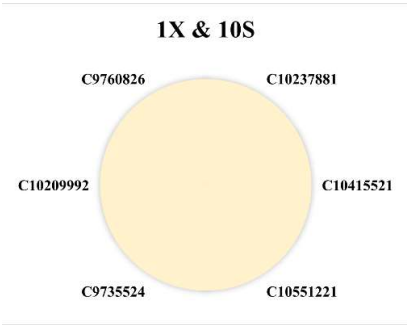
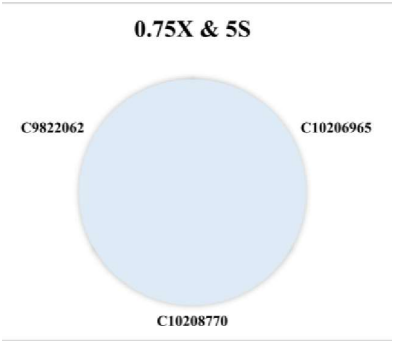
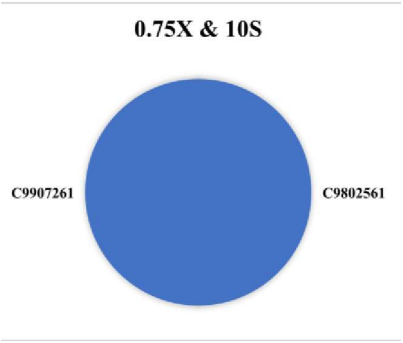
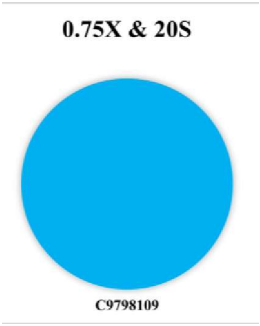
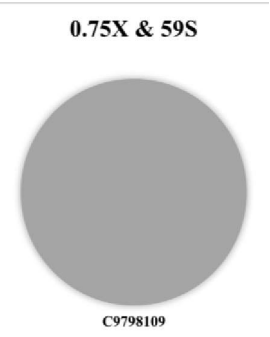
(a) Combination of 1X & 5s	(b) Combination of 1X & 10s
	
(c) Combination of 0.75X & 5s	(d) Combination of 0.75X & 10s
	
(e) Combination of 0.75X & 20s	(f) Combination of 0.75X & 59s
	

Table 4.15 gives the best combination for each audio file by the accuracy measurement with reference to the accuracy scale table. From Table 4.15(a), there were a total of 16 audio files that obtained the best accuracy in the combination of 1X and 5s, which included “C9733930”, “C9814494”, “C10347935”, “C10395406”, “C10804449”, “C10371808”, “C10411974”,

“C10503847”, “C9846248”, “C9935118”, “C9937436”, “C10052504”, “C10063278”, “C10077643”, “C10078155”, and “C10081916”. From Table 4.15(b), there were a total of 6 audio files that obtained the best accuracy in the combination of 1X and 10s, which included “C10237881”, “C10415521”, “C10551221”, “C9735524”, “C10209992”, and “C9760826”. From Table 4.15(c), there were a total of 3 audio files that obtained the best accuracy in the combination of 0.75X and 5s, which included “C10206965”, “C10208770”, and “C9822062”. From Table 4.15(d), there were a total of 2 audio files that obtained the best accuracy in the combination of 0.75X and 10s, which included “C9802561” and “C9907261”. From Table 4.15(e), there was only 1 audio file that obtained the best accuracy in the combination of 0.75X and 20s, which was “C9798109”. From Table 4.15(f), there was also only 1 audio file that obtained the best accuracy in the combination of 0.75X and 59s, which was “C9855059”.

Table 4.16: Best Combination of Each Audio Files for WRR Measurement.

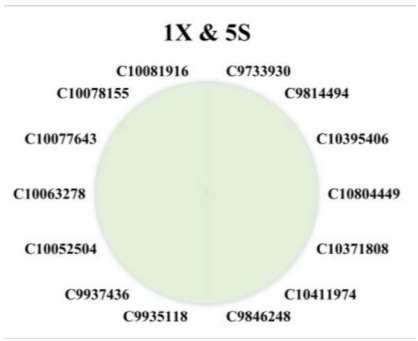
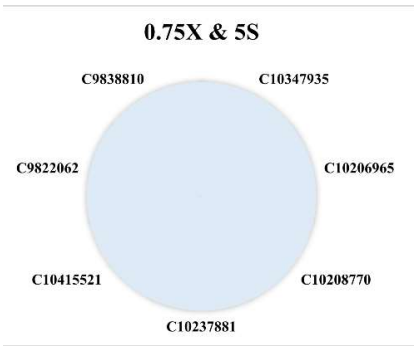
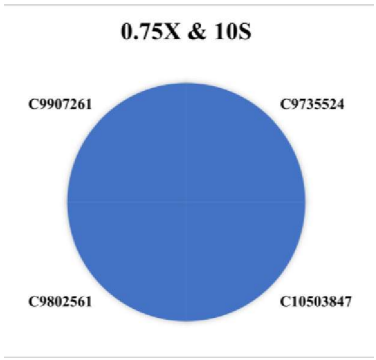
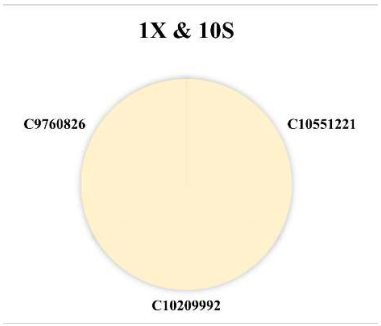
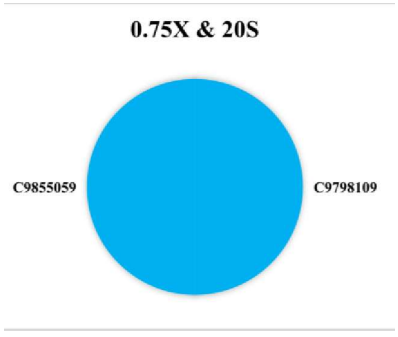
(a) Combination of 1X & 5s 	(b) Combination of 0.75X & 5s 
(c) Combination of 0.75X & 10s 	(d) Combination of 1X & 10s 
(e) Combination of 0.75X & 20s 	

Table 4.16 shows the best combination for each audio file by the accuracy measurement with reference to Word Recognition Rate (WRR).

From Table 4.16(a), there were a total of 14 audio files that obtained the best accuracy in the combination of 1X and 5s, which included “C9733930”, “C9814494”, “C10395406”, “C10804449”, “C10371808”, “C10411974”, “C9846248”, “C9935118”, “C9937436”, “C10052504”, “C10063278”, “C10077643”, “C10078155”, and “C10081916”. From Table 4.16(b), there were a total of 7 audio files that obtained the best accuracy in the combination of 0.75X and 5s, which included “C10347935”, “C10206965”, “C10208770”, “C10237881”, “C10415521”, “C9822062”, and “C9838810”. From Table 4.16(c), there were a total of 4 audio files that obtained the best accuracy in the combination of 0.75X and 10s, which included “C9735524”, “C10503847”, “C9802561”, and “C9907261”. From Table 4.16(d), there were a total of 3 audio files that obtained the best accuracy in the combination of 1X and 10s, which included “C10551221”, “C10209992”, and “C9760826”. From Table 4.16(e), there were only 2 audio files that obtained the best accuracy in the combination of 0.75X and 20s, which included “C9798109” and “C9855059”.

Table 4.17: Number of Audio Files in Each Best Combination.

Combination	Scale	WRR
1X & 5s	16/30	14/30
0.75X & 5s	3/30	7/30
1X & 10s	6/30	3/30
0.75X & 10s	3/30	4/30
0.75X & 20s	1/30	2/30
0.75X & 59s	1/30	0/30

Figure 4.1 shows the total number of audio files for each combination when each of them performed their accuracy evaluation with reference to the accuracy scale table and Word Recognition Rate (WRR). By referring to the graph above, we observed that there were different combinations of time cut point and speed chosen to be the best combination for a particular audio file.

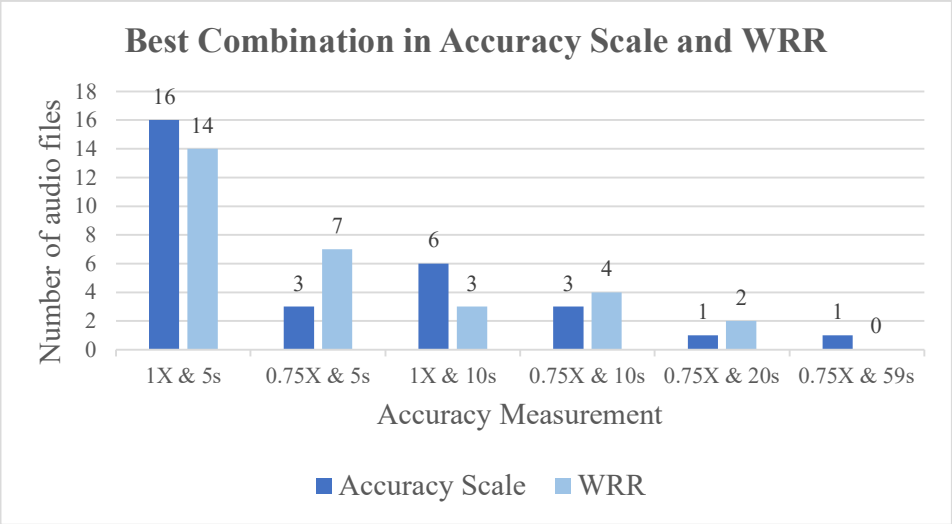


Figure 4.1: Number of Audio Files that Obtained Best Accuracy in Different Combinations in Each Accuracy Measurement.

For the accuracy scale table, a total of 16 out of 30 audio files had achieved their best result (accuracy) when the combination was with a time cut point of 5 seconds and normal speed. For WRR, a total of 14 out of 30 audio files had the best result when the combination was with a time cut point of 5 seconds and normal speed. From these results, approximately half of the samples were achieved higher accuracy in the combination of 1X and 5s no matter whether the text transcriptions were evaluated with reference to the accuracy scale table or calculated using the WRR formula. Thus, it could be led to a conclusion that the audio file performed better when it was being cut for every 5 seconds at normal speed. With this, we identified the best combination used to manipulate the audio file in the future to get a better result, which the combination of 1X and 5s was preferable.

An audio file with the short time cut point (5 seconds) was better in captured the words when converted the spoken words into written text. As a combination that achieved the best accuracy found, therefore the performance of speech analytics solution designed had improved.

Chapter 5

CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

Different accuracy measurements were used in evaluating the accuracy of the text transcription converted with the help of Python programming and Google Speech To Text API. One of the accuracy measurements applied was the calculation of the Word Recognition Rate by counting the number of words correctly recognized and the total number of words in the original text and finding the ratio between both of them. The second method used to assess the accuracy was to make reference to the accuracy scale table. In the accuracy scale table, three elements were being examined, such as text understanding, words wrongly spelt, and text completion.

In addition, there were two factors that were determined in order to identify the method that can apply to the audio sample so that it could achieve the best accuracy. The factors investigated were related to the time and speed of the audio samples. Different combinations of time cut point and speed were applied on each sample (call recording).

As a result, the best combination that gives the best accuracy was investigated. With reference to the accuracy scale table, the best combination belonged to the combination of 1X and 5s, as there were a total of 16 out of 30 audio files that achieve the best accuracy when they were tested in this combination. On the other hand, for WRR, a total of 14 out of 30 audio files achieved the highest accuracy when they were tested in the combination of 1X and 5s as well. Hence, these results led to a conclusion that the best combination could be applied to the audio samples to get the highest accuracy was with normal speed and time cut point of 5 seconds.

5.2 Recommendations for Future Work

5.2.1 Recommendation 1

In the future, further research on how to increase the reliability of speech to text can be approached. Generally, one of the important factors that affect the accuracy of speech recognition was related to user characteristics, such as the

way of speaking and its languages, the way the speaker pronounces a word, the speaker's accent, etc. In this field, AI engines could be trained by feeding it with pre-recorded audio files of different speakers. With this, it could increase the accuracy as the database had different kinds of data that could be estimated and used for text transcription. When speech recognition is used in the specific terminology in the industry, for example, in this project was insurance industry, a special training on the AI engine can be applied which could be fed by common sentences or words used by the customer service speakers in the insurance field. It would help the machine to transcript the audio files more accurately.

5.2.2 Recommendation 2

Under this project, Malay and mixed languages used in Malaysia had become one of the factors that bring effects on the accuracy as there were not having many databases of these kinds of languages used in the speech recognition APIs developed in the world. More data were needed to be feed into the engine such as the natural language toolkits library for Malay languages.

5.2.3 Recommendation 3

Another approach that can be further research was the time taken to convert the spoken words into written text. As we know, there were different lengths of duration needed for different speech analytics solutions. In this case, further research could be done by comparing the time taken for Google Speech To Text API and other APIs available.

REFERENCE

Amir, G. and Murtaza, H., 2014. Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management,* 35(2015), pp. 137-144.

Beth, W., 2019. *Word Error Rate Mechanism, ASR Transcription and Challenges in Accuracy Measurement*. [Online]. Available at: https://www.gmrtranscription.com/blog/word-error-rate-mechanism-asr-transcription-and-challenges-in-accuracy-measurement [Accessed 24 July 2021].

Bogdan, I., 2019. Evaluating Google Speech-to-Text API's Performance for Romanian e-Learning Resources. *Information Economic*ă, 23(1), pp. 19-20.

Chowdhury, G.G., 2003. Natural language processing. *Annual review of information science and technology*, 37(1), pp. 51-89.

James, F.A., n.d. *Natural Language Processing*. [Online]. Available at: <https://dl.acm.org/doi/pdf/10.5555/1074100.1074630> [Accessed 2 March 2021].

Juang, B. H. and Lawrence, R. R., 2004. *Automatic Speech Recognition – A Brief History of the Technology Development.* [Online]. Available at: <https://folk.idi.ntnu.no/gamback/teaching/TDT4275/literature/juang_rabiner04.pdf>. [Accessed 1 March 2021].

Liddy, E.D., 2001. In Encyclopedia of Library and Information Science, Marcel Decker. Inc.- Natural Language Processing.

Santosh, G., Bharti, G. and Pravin, Y., 2010. A Review on Speech Recognition Technique. *International Journal of Computer Application (0975-8887),* 10(16), pp. 16-24.

Shaghayegh, E., 2019. *Word Error Rate (WER) for Recognition of Natural Interactions.* [Online]. Available at: <https://init.cise.ufl.edu/2018/04/03/word-error-rate-wer-for-recognition-of-natural-interactions/>. [Accessed 24 July 2021].

Scott, S. and Chung, Q.B., 2018. Making a case for speech analytics to improve customer service quality: Vision, implementation, and evaluation. *International Journal of Information Management,* 45(2019), pp. 223-232.

APPENDICES

Appendix A: Original Text Transcription of "C9733930"

(...Bunyi phone kau tu tu tu tu tu...)	
Staff:	Selamat petang
Client:	Hello
Staff:	Selamat petang
Client:	Hello
Staff:	Boleh bercakap dengan cik Rohani Harista
Client:	Em saya
Staff:	Cik Harista, minta maaf berganggu ya. Saya buat panggilan daripada Sunlife Malaysia Assurance
Client:	Em hmm
Staff:	Okay. Ini berkaitan dengan satu notis permohonan penyerahan
Client:	En. Pembatalan tu kan
Staff:	Ya betul untuk pembatalan. Kalau saya lihat dekat sini cik ada lagi sekali pada 11 hari bulan hari tu hantarkan dokumen ataupun hantarkan emel tanpa dokumen cik. Cik ada terima belum borang nilai serahan yang kita hantar itu
Client:	Borang tu saya dah serahkan tapi saya tak ada saksi pun. Masa open tu tak da saksi pun buat
Staff:	Oh macam ini cik harista, untuk maklumat cik, borang tu maksudnya cik dah terima lah borang yang kita hantar dua kali ini
Client:	Ah I dah terima dan saya dah print out dah cuma belum isi lagi
Staff:	Ok macam ni kalau saya lihatkan di sini, cik harista untuk bahagian saksi tu, saksi tu sesiapa sahaja lebih 18 tahun, okay, selain yang berdaftar di atas polisi lah, maksudnya selain daripada cik harista dan juga cik hashima yang ini ibu ya? Sebab ... ibu sebagai penama cumanya macam sekarang ini, saksi tu cik harista perlu nyatakan orang selain yang

	didaftarkan di dalam polisi
Client:	Oh
Staff:	Ha. Sebab masuk hari tu takda saksi sebab cik berdaftar dengan pegawai bank, tak perlu ada saksi, sebab dia pendaftaran. Tapi untuk pembatalan ini, borang penyerahan tu, cik kena lengkapkan bahagian pemegang polisi, cik boleh semak di bahagian pengakuan, bahagian pemegang polisi cik kena lengkapkan, bahagian saksi tu sesiapa saja lebih daripada 18 tahun selain daripada cik harista and juga cik hashima.
Client:	Oh maksudnya siapa siapa pun boleh lah
Staff:	Ya betul
Client:	Cuma nak saksi yang saya ni buat isi borang tu je lah kan
Staff:	Ya betul Cik harista
Client:	Saya pun fikir masa tu saya buat pun tak ada pun saksi pun nak buat itu, tulis depan pegawai bank tu, macam tu je, dah printout dah semuanya
Staff:	Ah boleh. Cik dah jelaskan?
Client:	Ah dah jelas
Staff:	Nanti cik harista hantarkan kat kita. Tapi cik perlu pastikan ya tandatangan cik ada buat perubahan tak?
Client:	Rasanya macam tu lah
Staff:	Ah maksudnya kalau kata takda, cik perlu pastikan tandatangan cik sama dengan pendaftaran yang sebelum ini ya. Sebab kalau kata tak sama, nanti kita akan hantar satu lagi borang untuk tandatangan. Tapi macam sekarang ni cik harista lengkapkan borang tu dulu, bahagian butiran cik kena lengkap, bahagian bank pun kena lengkap, bahagian saksi, bahagian pengakuan tu perlu lengkap dengan tandatangan cik. Kemudian cik harista nyatakan ataupun compilekan bersama dengan salinan kad pengenalan
Client:	Ah IC saya lah?
Staff:	Ah ya

Client:	Ah okay okay kay. Ah pembatalan ini masa berapa lama?
Staff:	Proses penyerahan dalam tempoh 7 ke 14 hari bekerja.
Client:	Oh dua minggu lah
Staff:	Ya betul
Client:	Em okok nanti saya buat. Sebab saya dah print out dah belum isi je lagi
Staff:	Okay jelas. Tapi saya buat panggilan ni tujuan untuk buat penerangan lah sebab kita ada terima notis. Saya takut cik tak terima borang. Tapi kalau dah terima borang, bahagian saksi tu cik kena nyatakan selain yang berdaftar dengan polisi
Client:	Masa yang mula mula tu saya tak nampak borang tu kat bawah so saya call balik semula
Staff:	Oh okay okay
Client:	Borang tu kat bahagian bawah tu
Staff:	Ya dekat bahagian bawah. Maksudnya kalau cik dah terima, cik boleh isikan dan hantar balik kepada kita ya
Client:	Ah okay okay
Staff:	Okay Minta maaf kerana ganggu itu saja. Terima kasih.
Client:	Sama

Appendix B: Time Cut Point (59 Seconds) of "C9733930"

Start_Time	End_Time	Transcription
0	59000	bunyi phone kau tu tu tu tu tu tu tu tak maafkan aku ya Saya dah ada Buat panggilan daripada solat Perlis insuran CIDB berkata dengan satu Doctors permohonan penyerahan pembatalan tu kan dia betul-betul pembalasan lihat dekat sini cik cik ada lagi lagi sekali pada 11 haribulan hari tu hantarkan dokumen ataupun hantarkan email tanpa dokumen cik cik dah terima belum borang Biasalah orang dah hantar kat awak Tapi saya tak suka memaksa cik dah terima laporan kita dah hantar gaduh kali ni saya dah terima saya dah print out
59000	118000	anak sendiri lagi kalau saya lihat dekat sini

		kasi tahu untuk bahagian sat situ situ sesiapa sahaja lebih pada 18 tahun lagi selain yang berdaftar di atas kepala sila bersedia selain daripada cita-cita dan juga Si hasi mahir ibu ya semuanya akan ibu sebagai penama cumanya macam sekarang ni Sat situ yang senang betul dia tekan orang sana yang didapatkan di dalam polisi Oh sebab masuk ayat terakhir tu tak ada saksi sebab berdaftar berdaftar dengan pegawai bank tak perlu ada kasih sebab dia berani untuk pembatalan borang penyerahan tu cikgu dilengkapkan bahagian pemegang polisi siap boleh semak di bahagian pengakuan sebagai pegangan pemegang polisi cakap dengan Tokan bahagian saat itu sesiapa saja Lebih daripada itu lebih dapat 18 tahun selain daripada Encik hasrita dan juga Perang Suriah 50 maksudnya siapa-siapa pun bolehlah betul
118000	177000	semenax axian saya ni buat isi borang tu jelaskan Ya betul saya tak fikir sebab itu macam Sebab tu Tak ada pun Tak ada pun tak boleh depan pengawasan tua macam tu abang dah tak ada kerja Nanti kasi tahu Hantar kat dia kat kita tapi kita perlu pastikan tak ada tangan cik ada buat perubahan tak Rasanya macam tulah abang saja kalau kata tak ada saya perlu pastikan tandatangan cek sama dengan pendaftaran Yang sebelumnya Sebab kalau kata tak sama tapi kita kena tekan satu lagi barang tersebut tandatangan daripada sekarang ni letakkan borang tu dulu bahagian butir Acik kena nangkap bahagian pengenalan sedap bahagian sakit sendi bahagian pengakuan tu perlu dekat dengan tandatangan cek kepada jantan Nyatakan ataupun kepentingan bersama dengan salinan kad pengenalan IC
177000	236000	tujuh kan 14 hari bekerja Tunggulah Ya betul okey okey okey okey nanti saya buat saya dah balik dah Belum isi lagi okey jahlah tapi saya buat panggilan itu juga untuk buat penerangan Esok kita ada terima notis saya Saya takut saya tak terima borang Tapi kalau dah tiba borang bahagian sat situ cari kedai yang tekan selain yang berkaitan dengan politik semasa Yang mula-mula tu saya tak nampak orang tu kat bawah kopek dia ni dia ada Express kat bawah tu daripada bagi ubat batuk je sekarang ni dah tiba nanti saya isikan hantar kembali

		kepada kita ya okey okey okey okey sama-sama

Appendix C: Time Cut Point (20 Seconds) of "C9733930"

Start_Time	End_Time	Transcription
0	20000	bunyi pun kata tu tu tu tu tu tu tak maafkan aku ya Saya dah ada buat
20000	40000	Pergilah daripada solat di Sinsuran pagi ini berkata dengan satu Doctors permohonan penyerahan pembatalan Bukan dia betul-betul pembatalan Kalau saya lihat dekat sini cakap lagi sekali pada 11 haribulan hari tu hantarkan dokumen ataupun hantarkan email tanpa dokumen cik
40000	60000	cik dah terima ke belum borang Biasalah orang dah hantar kan terus terang kan Tapi saya tak Tipulah tidak kasi tahu Untuk makluman isi borang tu maksud dia cik dah terima laporan kita dah hantar gaduh kali ni saya dah terima saya dah print out mudah
60000	80000	lagi kalau saya lihat dekat sini kasi tahu untuk bahagian sat situ sat sat situ sesiapa sahaja lebih pada 18 tahun lagi selain yang berdaftar di atas polisi la maksudnya selain daripada cita-cita dan juga cik Hasimah yang ibu ya semuanya akan ibu sebagai penama cumanya macam
80000	100000	cara bisa situ ya Setan betul dia tekan orang salah yang didapatkan di dalam polisi Oh sebab masuk ayat terakhir tu tak ada saksi sebab berdaftar berdaftar dengan pegawai bank tak perlu ada kasih sebab dia buat pendaftaran untuk pembatalan borang penyerahan tu cikgu dilengkapkan bahan
100000	120000	koleksi kad-kad cik boleh saya nak dia bahagia pengakuan sebagai pegang pemegang polisi cikgu dekatkan bahagian saat situ sesiapa sahaja lebih daripada itu lebih daripada 18 tahun selain daripada Cik Hafiz ada juga parasit Hazimah masuknya siapa-siapa bolehlah betul nak kat saya ni
120000	140000	isi borang tu jelah kan Ya betul saya tak fikir sebab itu macam Sebab tu Tak ada pun saksi pun nak tunggu depan pegawai mentua macam tu Jah dia tak dapat jalan Sudah jelas dalam Nanti kasi tahu Hantar kat dia kat kita tapi kita perlu pastikan ya tandatangan cek

		ada buat perubahan tak
140000	160000	Rasanya macam tulah abang sakit kalau kata tak ada chat kau lepaskan tandatangan cek sama dengan pendaftaran Yang sebelumnya Sebab kalau kata tak sama tapi kita kena tekan satu lagi barang tersebut tak ada tak jadi macam sekarang ni kita letakkan borang tu dulu bahagian butiran chicken lengkap bahagian pengenalan
160000	180000	bagai saksi bahagian pengakuan tu perlu dekat dengan tandatangan cek kepada Jah tekan ataupun kepilkan bersama dengan salinan kad pengenalan IC okey okey okey proses bayaran dalam tempoh 7 ke 14 hari bekerja
180000	200000	okey okey okey nanti saya buat tapi awak dah belum isteri lagi okey jahlah tapi saya buat panggilan itu untuk buat perangai lah sebab kita ada terima notis saya Saya takut yang cakap Cuba borang Tapi kalau dah cuba kau orang bahagia kat situ cari kedai yang tekan selain yang berkaitan dengan politik semasa Yang mula-mula tu saya tak nampak
200000	220000	dekat bawah Kompleks ibu dia ni dia ada Express kat bawah tu daripada pakai spek mata tu besok je sekarang ni dah tiba nanti saya isikan hantar kembali kepada kita ya Okey Okey Okey Okey itu saja Suria cinema

Appendix D: Time Cut Point (10 Seconds) of "C9733930"

Start_Time	**End_Time**	**Transcription**
0	10000	bunyi phone kata tut tut tut tut
10000	20000	datang sini boleh bercakap dengan Cik Rohani hasrita Saya minta maaf ganggu Ya saya dah dapat
20000	30000	lagilah daripada solat di Sinsuran pagi ini berkata dengan satu notis permohonan penyerahan pembatalan Bukan dia betul-betul
30000	40000	Kalau saya lihat dekat sini chat-chat ada lagi lagi sekali pada 11 haribulan hari tu hantarkan dokumen ataupun hantarkan email tanpa dokumen cik
40000	50000	cik dah terima ke belum borang Biasalah orang dah hantar kan terus terang kan Tapi saya tak tipu sama saya
50000	60000	oh baca dikasi tahu tu maklumat sikit borang

		tu maksud dia cik dah terima laporan kita dah hantar gaduh kali ni saya dah terima saya dah pening awak pun dah
60000	70000	lagi kalau saya lihat dekat sini kasi tahu untuk bahagian saksi tu saksi tu sesiapa sahaja lebih pada 18 tahun lagi selain yang bernama
70000	80000	datang sekarang sila bersedia selain daripada Cik hasrita Dan juga Hasimah yang ibu ya semuanya Nyatakan ibu sebagai penama cumanya macam
80000	90000	cara bisa situ ya Setan betul dia tekan orang salah yang didapatkan di dalam polisi Oh sebab masuk air tanah itu tak ada saksi sebab
90000	100000	tidak berdaftar berdaftar dengan pegawai bank tak perlu ada saksi sebab dia pendaftaran untuk pembatalan borang penyerahan tu cari kenalan Kapten bahan
100000	110000	koleksi kad-kad cik boleh saya nak dia bahagia pengakuan sebagai pegang pemegang polisi cikgu dekatkan bahagian saat situ sesiapa sahaja lebih daripada cik lebih dari
110000	120000	11 tahun selain daripada cita-cita dan juga perang Syria 50 maksudnya siapa-siapa pun bolehlah betul cuma nak sayang saya ni
120000	130000	isi borang tu jelaskan Ya betul saya tak fikir sebab itu macam Sebab tu Tak ada pun saksi pun nak tunggu depan pegawai mentua macam tu
130000	140000	apa dekat sana saya dah jelas Nanti kat situ hantar kat dia kat kita tapi cik perlu pastikan ya tandatangan cek ada buat perubahan tak
140000	150000	Rasanya macam tu lah abang sakit kalau kata tak ada saya perlu pastikan tandatangan cek sama dengan pendaftaran Yang sebelumnya Sebab kalau kata tak sama tapi kita kena tekan satu
150000	160000	orang tersebut tak ada tangan ada macam sekarang ni tak letakkan borang tu dulu bahagian butiran Cik kena lengkap bahagian pengenalan kau
160000	170000	bahagian saksi bahagian pengakuan tu perlu dekat dengan tandatangan cek kepada Jah tekan ataupun kepilkan bersama dengan salinan kad pengenalan IC
170000	180000	Saya nak tanya Okey okey-okey proses bayaran dalam tempoh 7 ke 14 hari bekerja
180000	190000	okey okey okey nanti saya buat tapi awak dah belum isteri lagi okey jahlah tapi saya buat panggilan itu untuk buat perangai lah sebab kita ada

190000	200000	Nanti saya Saya takut saya tak terima borang Tapi kalau dah tiba borang bahagian kat situ cari kedai yang tekan selain yang berada pada sisi manusia yang mula-mula tu saya tak nampak
200000	210000	dekat bawah episod kopek dia punya ni dia ada kat bawah tu daripada pakai spek mata tu besok je sekarang ni dah tiba nanti saya isikan hantar kembali kepada kita ya
210000	220000	Okey Okey Okey Okey kasi tahu lagu itu saja sudah si sama

Appendix E: Time Cut Point (5 Seconds) of "C9733930"

Start_Time	**End_Time**	**Transcription**
0	5000	
5000	10000	bunyi pun kata tut tut tut tut
10000	15000	datang sini boleh bercakap lagi
15000	20000	Cik Rohani hasrita Saya minta maaf ganggu Ya saya dah ada buat
20000	25000	Pergilah daripada solat di Sinsuran pagi ini berkata dengan satu
25000	30000	Pesta Bunga dan penyerahan pembatalan tu kan dia betul-betul
30000	35000	Kalau saya lihat dekat sini chat-chat ada lagi lagi sekali pada 11 haribulan hari tu
35000	40000	hantarkan dokumen ataupun hantarkan email tanpa dokumen cik
40000	45000	cik dah terima ke belum borang Biasalah orang dah hantar kan terus
45000	50000	sudah hantar Tapi saya tak tipu siapa-siapa saya tak ada kat situ
50000	55000	oh baca dikasi tahu untuk maklumat cik borang tu masa dia cik dah tu
55000	60000	mana borang kita dah hantar gaduh kali ni saya dah terima saya dah print out mudah
60000	65000	lagi kalau saya lihat dekat sini kasi tahu untuk bahagian saksi tu
65000	70000	saksi tu sesiapa sahaja lebih pada 18 tahun lagi selain yang bernama
70000	75000	datang sekarang sila bersedia selain daripada Cik hasrita Dan juga cik Hasimah
75000	80000	ibu makanlah ibu sebagai penama cumanya macam
80000	85000	cara bisa situ yang senang betul dia tekan orang salah yang didapatkan di dalam polis
85000	90000	oh oh sebab masuklah itu tak ada saksi sebab

90000	95000	tidak berdaftar berdaftar dengan pegawai bank tak perlu ada saksi sebab daripada
95000	100000	untuk pembatalan borang penyerahan tu cikgu dilengkapkan bahan
100000	105000	koleksi Kak Cik boleh saya nak dia bahagia pengakuan sebagai pegang-pegang
105000	110000	siapkan bahagian saksi tu sesiapa saja Lebih daripada 9
110000	115000	11 tahun selain daripada cita-cita dan juga pengasih azimat
115000	120000	masuknya siapa-siapa pun bolehlah saya betul-betul nak kat dengan saya ni
120000	125000	isi borang tu jahlah kan Ya betul saya tak fikir sebab itu macam Sebab tu
125000	130000	saksi pun tak boleh depan pegang orang tua macam tu
130000	135000	Abang dah tak sayang saya dah jelas Nanti chat
135000	140000	hantarkan dekat kita tapi cik perlu pastikan yang tandatangan cik ada buat perubahan tak
140000	145000	Rasanya macam tulah rasanya kalau kata tak ada telefon lepas ikan tak datang
145000	150000	cik sama dengan pendaftaran sebelum ni ya Sebab kalau kata tak sama nanti kita akan hantarkan satu
150000	155000	orang tersebut tak ada tangan tak ada macam sekarang ni tak letak ke
155000	160000	orang tu dulu bahagian butiran chicken lengkap bahagian bank pengenalan
160000	165000	bagai saksi bahagian pengakuan tu perlu dekat dengan tandatangan cek kepada jantan-jantan
165000	170000	dah sampai ke pekan bersama dengan salinan kad pengenalan
170000	175000	saya okey okey okey
175000	180000	proses penyerahan dalam tempoh 7 ke 14 hari bekerja
180000	185000	okey okey okey nanti saya buat babi
185000	190000	isi lagi okey jahlah tapi saya buat panggilan itu untuk buat padalangan Esok kita ada
190000	195000	Nanti saya Saya takut saya tak terima borang Tapi kalau dah terima borang bahagian 1
195000	200000	asyik kena jatuhkan selain yang berkaitan dengan politik semasa Yang mula-mula tu saya tak nampak
200000	205000	dekat bawah Kompleks ibu dia ni dia ada Express kat bawah tu ada
205000	210000	besok je sekarang ni dah tiba nanti saya isikan dah hantar kembali kepada kita ya

210000	215000	Okey Okey Okey Okey kasi tahu lagu itu saja sudah si sama

Appendix F: Original Text Transcription of "C9735524"

Staff:	Helo Sun life Malaysia. Boleh saya bantu? Helo?
Client:	helo Assalamualaikum
Staff:	Mualaikum salam
Client:	Helo saya Nur Syahirah Binti Mahadzir. Pegawai CIMB
Staff:	Okay kenapa ya
Client:	Saya nak tanya pasal sun life
Staff:	Okey tentang apa tu
Client:	Okey saya nak cancel lah sun life ni
Staff:	Okey yang itu saya kena semak dulu boleh baca saya tak IC ke nombor polisi ke
Client:	Okey IC sembilan tujuh kosong lima eh sembilan tujuh kosong enam
Staff:	Okey
Client:	25
Staff:	Okey
Client:	385164
Staff:	5164 Okey sebentar ya Saya cuba semak dulu rekod yang dimasukkan dulu
Client:	ya macam mana plan ya
Staff:	macam mana
Client:	plan sun mozi shield
Staff:	bukan. saya semak dulu sebentar ya
Client:	Okey
Staff:	saya tengok kat sini di bawah nama Nur Syahirah Binti Mahadzir ya Nanti saya nak buat pengesahan sedikit boleh tak
Client:	Ya boleh

Staff:	polisi yang dimaksudkan ini cik syahirah daftar melalui telefon ke pergi bank ke macam mana ya.
Client:	saya melalui telefon
Staff:	telefon boleh saya tahu bayar bulan-bulan tu berapa
Client:	bulan bulan dalam 78
Staff:	okey soalan terakhir Boleh saya tahu nombor telefon
Client:	Okey 011
Staff:	Okey
Client:	264
Staff:	okey
Client:	83211
Staff:	Terima kasih untuk pengesahan boleh saya tahu kenapa nak batalkan polisi tu cik
Client:	sebab saya nak ambil polisi lain saya nak ambil insurans lain
Staff:	begitu. yang polisi cik daftar ni yang 78 ni pun tengah tengok kat sini kita dah kurangkan bayaran dia sehingga 62.52 sen kemudian cik punya polis ini kalau tengok kat dalam sistem ada perlindungan sehingga tiga ratus ribu dan Agak tinggi juga cik betul pasti nak batalkan juga polis ini
Client:	Ya
Staff:	begitu. Macam saya cakap tadi kalau cik nak ambil insurans dekat syarikat lain, cik boleh ambil polisi ini sebagai polisi tambahan sebab kita dah kurangkan bayaran dia
Client:	Tak apa saya nak cancel
Staff:	begitu Lagipun polisi promosi Kalau cik batalkan sekarang mungkin lepas ni Tak adalah cik pasti nak batalkan ke tidak
Client:	Ya saya pasti
Staff:	Begitu. Tak apa. Cik syahirah dah cakap macam tu saya tak boleh nak tak boleh nak halang cik jadi untuk pembatalan ni saya boleh bantu melalui telefon dan untuk maklumat dia punya pembatalan ni akan ambil masa dalam 7 hari bekerja daripada hari ini Okey di mana bila kita nak batalkan kita akan hantar satu SMS dekat cik syahirah nanti ya

Client:	Ok baik
Staff:	selain itu
Client:	maksudnya Lepas batal tu dia tak potong dah kan
Staff:	kalau yang dia punya rumah tengah order Pergi Takkan padam sekali. Andaikan kata kalau cik dapat apa-apa SMS lepas 7 hari ketiga cik terus hantar balik dekat nombor yang sama untuk semak status
Client:	Okey baik terima kasih
Staff:	ada apa lagi kah
Client:	tak ada
Staff:	Okey kalau tak ada terima kasih sun life Malaysia Assalamualaikum

Appendix G: Time Cut Point of 5s and Speed of 1X of "9735524"

Start_Time	End_Time	Transcription
0	5000	
5000	10000	Hello Hello assalamualaikum assalamualaikum
10000	15000	Syahirah ada Okey okey saya
15000	20000	kacau sangat lain Okey tentang Lepas tu Okey saya nak tulis
20000	25000	Okey yang itu saya kena semak dulu boleh baca saya tak IC ke nombor polisi ke
25000	30000	okey 705
30000	35000	Wassalam okey
35000	40000	385 1645 14 Okey sebentar ya
40000	45000	cuba semak dua ekor termasuk data nya
45000	50000	Lan bukan
50000	55000	saya semak dulu sebentar ya Okey saya tengok kat sini dia bawa nama Nur Syahirah
55000	60000	Terima kasih ya Nanti saya nak buat pengesan sedikit boleh tak
60000	65000	telefon saya dimaksudkan itu saya dah daftar dekat telefon ke pergi bank ke macam mana ya saya
65000	70000	daripada kita bayar bulan-bulan tu berapa
70000	75000	ini adalah 78 Okey soalan terakhir boleh
75000	80000	nombor telefon HQ 011 Okey 2
80000	85000	Okey 83211 sakit masuk ke pengesahan
85000	90000	Kesian tau kenapa saya nak batalkan polisi

		cek Sebab saya
90000	95000	habis periksa nak ambil surat beranak
95000	100000	cetakan 7 8 kali pun tak tengok kat sini kita dah kurangkan bayaran dia sehingga 62
100000	105000	50 sen kemudian cik punya post ini kalau tinggal kat dalam
105000	110000	dating dengan single 300000 dan Agak tinggi juga cikgu tekan semula nak batalkan
110000	115000	tak boleh sini Itu macam saya cakap tadi cik Habis sudah lain dengan syarikat lain
115000	120000	ini dia boleh ambil polis yang sedap ni sebagai polisi tambahan Sebab kita pun akan marah dia
120000	125000	Tak apa saya nak sponsor lagi pun
125000	130000	kau masih di Kalabakan sekarang mungkin lepas ni Tak adalah Encik pasti ke nak makan ke tidak
130000	135000	ayah saya itu kenapa Hello Cik saya dah cakap macam tu saya tak boleh nak
135000	140000	tak boleh nak halang sikit Jadi untuk sementara ni saya boleh bantu bertelefon dan
140000	145000	untuk maklumat dia punya permata ni akan ambil masa dalam 7 hari bekerja daripada hari ini
145000	150000	dimana bila kita dapat rasa kita akan hantar satu SMS dekat cik Hari Raya
150000	155000	yang akibat kes lari daripada tu ya
155000	160000	maksud dia kalau dapat bantal tu suruh dia tak potong dah kan kalau muka ni yang dia punya rumah tu nak tutup
160000	165000	cakap Adam sekali kita kena beli
165000	170000	dapat apa-apa SMS selepas 3 hari bekerja Seterusnya kita balik dekat nombor yang sama untuk
170000	175000	masih sakit lagi ke
175000	180000	Okey kalau kita terima kasih kan bagi salam Malaysia Assalamualaikum

Appendix H: Result of "C9733930"

C9733930 (First minute of recording) **Total words in original text = 134 words**					
Time cut point: 59s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	25	50	45	35	5
Wrong Spelling Word	5	18	11	28	9
Text completion (%)	25	50	45	35	15
Scale	2	3	3	2	1
Word Recognition Rate (WRR) (%)	13	47	32	20	3
Time cut point: 20s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	30	45	50	30	5
Wrong Spelling Word	23	10	14	10	2
Text completion (%)	40	50	50	40	10
Scale	2	3	3	2	1
Word Recognition Rate (WRR) (%)	28	46	45.5	28	3
Time cut point: 10s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	35	50	55	25	5
Wrong Spelling Word	9	30	15	35	14
Text completion (%)	35	50	60	30	5
Scale	2	3	3	2	1
Word Recognition Rate (WRR) (%)	31	46	38	21	6
Time cut point: 5s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	35	40	60	20	5
Wrong Spelling Word	75	72	11	97	∞
Text completion (%)	35	40	60	20	5
Scale	2	2	4	2	1
Word Recognition Rate (WRR) (%)	32	39	49	19	6

Appendix I: Result of "C9814494"

C9814494 (First minute of recording) **Total words in original text = 104 words**					
Time cut point: 59s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	20	20	25	10	5
Wrong Spelling Word	14	7	1	7	3
Text completion (%)	20	20	25	10	5
Scale	1	1	2	1	1
Word Recognition Rate (WRR) (%)	23	20	17	10	3
Time cut point: 20s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	25	25	45	15	5
Wrong Spelling Word	9	11	7	14	8
Text completion (%)	30	25	45	15	5
Scale	2	2	3	1	1
Word Recognition Rate (WRR) (%)	30	36	51	24	14
Time cut point: 10s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	50	30	55	35	5
Wrong Spelling Word	28	14	9	14	6
Text completion (%)	40	30	50	35	5
Scale	3	2	3	2	1
Word Recognition Rate (WRR) (%)	51	34	59	39	11
Time cut point: 5s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	20	25	65	20	5
Wrong Spelling Word	42	32	9	38	18
Text completion (%)	20	25	50	20	5
Scale	1	2	4	1	1
Word Recognition Rate (WRR) (%)	41	46	60	42	20

Appendix J: Result of "C10347935"

C10347935 (First minute of recording) **Total words in original text = 115 words**					
Time cut point: 59s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	5	15	20	5	1
Wrong Spelling Word	16	5	8	5	0
Text completion (%)	10	20	10	10	5
Scale	1	1	1	1	1
Word Recognition Rate (WRR) (%)	11	13	12	8	3
Time cut point: 20s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	10	25	40	5	3
Wrong Spelling Word	6	11	15	5	0
Text completion (%)	10	25	35	10	5
Scale	1	2	2	1	1
Word Recognition Rate (WRR) (%)	28	29	36	14	5
Time cut point: 10s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	10	50	50	25	5
Wrong Spelling Word	24	29	9	16	16
Text completion (%)	20	50	40	30	5
Scale	1	3	3	2	1
Word Recognition Rate (WRR) (%)	20	50	41	33	8
Time cut point: 5s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	5	45	55	15	5
Wrong Spelling Word	84	21	12	26	13
Text completion (%)	10	50	55	15	5
Scale	1	2	3	1	1
Word Recognition Rate (WRR) (%)	19	51	41	30	16

Appendix K: Result of "C10206965"

C10206965 (First minute of recording) **Total words in original text = 142 words**					
Time cut point: 59s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	0	10	10	1	5
Wrong Spelling Word	8	25	19	2	9
Text completion (%)	1	10	15	1	5
Scale	1	1	1	1	1
Word Recognition Rate (WRR) (%)	3	32	26	4	6
Time cut point: 20s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	15	25	25	5	2
Wrong Spelling Word	42	26	32	38	1
Text completion (%)	15	25	25	5	2
Scale	1	2	2	1	1
Word Recognition Rate (WRR) (%)	25	37	39	13	4
Time cut point: 10s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	5	15	20	5	0
Wrong Spelling Word	44	25	38	12	0
Text completion (%)	5	15	20	5	0
Scale	1	1	1	1	1
Word Recognition Rate (WRR) (%)	23	38	30	13	0
Time cut point: 5s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	5	30	30	10	5
Wrong Spelling Word	5	37	53	35	18
Text completion (%)	10	30	30	15	5
Scale	1	2	2	1	1
Word Recognition Rate (WRR) (%)	35	41	39	25	6

Appendix L: Result of "C10208770"

C10208770 (First minute of recording) **Total words in original text = 105 words**					
Time cut point: 59s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	3	15	5	0	0
Wrong Spelling Word	3	6	5	0	0
Text completion (%)	3	20	5	0	0
Scale	1	1	1	1	1
Word Recognition Rate (WRR) (%)	11	23	5	3	1
Time cut point: 20s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	10	15	10	5	2
Wrong Spelling Word	6	10	8	1	0
Text completion (%)	10	20	15	5	2
Scale	1	1	1	1	1
Word Recognition Rate (WRR) (%)	17	32	15	11	5
Time cut point: 10s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	20	10	20	5	2
Wrong Spelling Word	5	9	21	0	1
Text completion (%)	25	10	30	5	2
Scale	2	1	2	1	1
Word Recognition Rate (WRR) (%)	29	33	30	13	7
Time cut point: 5s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	15	25	25	10	2
Wrong Spelling Word	16	14	19	1	2
Text completion (%)	15	30	30	10	2
Scale	1	2	2	1	1
Word Recognition Rate (WRR) (%)	27	38	31	18	8

Appendix M: Result of "C10237881"

C10237881 (First minute of recording) **Total words in original text = 104 words**					
Time cut point: 59s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	1	5	10	15	10
Wrong Spelling Word	0	4	7	10	5
Text completion (%)	1	10	10	20	10
Scale	1	1	1	1	1
Word Recognition Rate (WRR) (%)	3	26	24	31	29
Time cut point: 20s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	15	15	20	25	5
Wrong Spelling Word	12	9	11	15	10
Text completion (%)	20	20	30	30	10
Scale	1	1	2	2	1
Word Recognition Rate (WRR) (%)	35	36	40	46	19
Time cut point: 10s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	15	25	35	5	5
Wrong Spelling Word	13	11	11	5	5
Text completion (%)	20	30	40	7	7
Scale	1	2	2	1	1
Word Recognition Rate (WRR) (%)	37	47	50	14	14
Time cut point: 5s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	25	30	35	15	10
Wrong Spelling Word	14	7	13	7	8
Text completion (%)	25	30	35	25	10
Scale	2	2	2	2	1
Word Recognition Rate (WRR) (%)	39	54	48	39	24

Appendix N: Result of "C10395406"

C10395406 (First minute of recording) **Total words in original text = 123 words**					
Time cut point: 59s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	10	25	25	5	0
Wrong Spelling Word	23	16	25	8	4
Text completion (%)	15	30	35	5	0
Scale	1	2	2	1	1
Word Recognition Rate (WRR) (%)	34	42	40	16	2
Time cut point: 20s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	15	12	25	5	1
Wrong Spelling Word	30	20	23	12	6
Text completion (%)	20	20	25	15	1
Scale	1	1	2	1	1
Word Recognition Rate (WRR) (%)	36	41	36	28	6
Time cut point: 10s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	20	25	25	5	1
Wrong Spelling Word	33	24	23	13	8
Text completion (%)	20	30	30	15	1
Scale	1	2	2	1	1
Word Recognition Rate (WRR) (%)	36	45	46	23	6
Time cut point: 5s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	15	25	30	5	0
Wrong Spelling Word	26	20	23	9	7
Text completion (%)	20	30	35	15	2
Scale	1	2	2	1	1
Word Recognition Rate (WRR) (%)	30	41	51	20	7

Appendix O: Result of "C10415521"

C10415521 (First minute of recording) **Total words in original text = 130 words**					
Time cut point: 59s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	5	5	10	2	0
Wrong Spelling Word	1	4	12	3	0
Text completion (%)	5	5	10	5	2
Scale	1	1	1	1	1
Word Recognition Rate (WRR) (%)	15	27	22	12	5
Time cut point: 20s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	15	12	20	10	0
Wrong Spelling Word	9	2	13	8	6
Text completion (%)	20	20	25	10	1
Scale	1	1	2	1	1
Word Recognition Rate (WRR) (%)	25	28	25	14	5
Time cut point: 10s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	10	5	25	10	0
Wrong Spelling Word	4	8	11	8	0
Text completion (%)	10	10	30	10	0
Scale	1	1	2	1	1
Word Recognition Rate (WRR) (%)	19	24	27	16	2
Time cut point: 5s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	5	25	25	5	0
Wrong Spelling Word	12	14	10	6	0
Text completion (%)	5	25	25	5	0
Scale	1	2	2	1	1
Word Recognition Rate (WRR) (%)	21	32	23	12	6

Appendix P: Result of "C10551221"

C10551221 (First minute of recording) **Total words in original text = 99 words**					
Time cut point: 59s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	0	5	35	15	5
Wrong Spelling Word	0	8	11	9	1
Text completion (%)	1	10	35	15	5
Scale	1	1	2	1	1
Word Recognition Rate (WRR) (%)	3	25	47	28	16
Time cut point: 20s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	5	25	35	5	5
Wrong Spelling Word	4	12	14	10	1
Text completion (%)	5	30	35	15	5
Scale	1	2	2	1	1
Word Recognition Rate (WRR) (%)	11	45	48	27	12
Time cut point: 10s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	5	25	35	10	5
Wrong Spelling Word	4	16	19	10	6
Text completion (%)	5	30	40	15	5
Scale	1	2	2	1	1
Word Recognition Rate (WRR) (%)	20	40	51	25	21
Time cut point: 5s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	5	15	35	5	5
Wrong Spelling Word	20	25	21	9	4
Text completion (%)	10	25	40	20	15
Scale	1	1	2	1	1
Word Recognition Rate (WRR) (%)	16	39	49	35	12

Appendix Q: Result of "C10804449"

C10804449 (First minute of recording) **Total words in original text = 67 words**					
Time cut point: 59s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	5	25	0	5	0
Wrong Spelling Word	2	1	0	2	1
Text completion (%)	5	30	0	10	2
Scale	1	2	1	1	1
Word Recognition Rate (WRR) (%)	24	46	0	22	10
Time cut point: 20s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	25	20	35	15	1
Wrong Spelling Word	3	1	5	2	1
Text completion (%)	25	25	35	20	1
Scale	2	2	2	1	1
Word Recognition Rate (WRR) (%)	42	40	43	37	10
Time cut point: 10s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	25	25	35	20	0
Wrong Spelling Word	1	2	5	1	1
Text completion (%)	25	25	35	20	2
Scale	2	2	2	1	1
Word Recognition Rate (WRR) (%)	43	45	45	39	7
Time cut point: 5s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	20	25	40	10	1
Wrong Spelling Word	9	1	2	4	1
Text completion (%)	25	30	40	15	5
Scale	2	2	2	1	1
Word Recognition Rate (WRR) (%)	37	46	51	27	19

Appendix R: Result of "C9735524"

C9735524 (First minute of recording) **Total words in original text = 114 words**					
Time cut point: 59s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	0	5	10	1	0
Wrong Spelling Word	5	3	5	3	0
Text completion (%)	0	5	10	2	0
Scale	1	1	1	1	1
Word Recognition Rate (WRR) (%)	9	31	25	15	4
Time cut point: 20s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	10	20	30	5	0
Wrong Spelling Word	7	11	8	1	0
Text completion (%)	15	25	30	10	0
Scale	1	2	2	1	1
Word Recognition Rate (WRR) (%)	38	45	47	25	5
Time cut point: 10s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	10	20	35	5	0
Wrong Spelling Word	9	14	10	6	0
Text completion (%)	15	25	35	10	0
Scale	1	2	2	1	1
Word Recognition Rate (WRR) (%)	38	54	50	27	0
Time cut point: 5s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	5	25	30	10	1
Wrong Spelling Word	10	18	8	17	0
Text completion (%)	10	30	30	15	2
Scale	1	2	2	1	1
Word Recognition Rate (WRR) (%)	33	52	49	25	5

Appendix S: Result of "C10209992"

C10209992 (First minute of recording) **Total words in original text = 84 words**					
Time cut point: 59s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	5	5	35	15	5
Wrong Spelling Word	0	0	3	2	1
Text completion (%)	5	5	35	15	10
Scale	1	1	2	1	1
Word Recognition Rate (WRR) (%)	17	23	60	35	26
Time cut point: 20s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	5	5	30	25	5
Wrong Spelling Word	0	0	2	9	2
Text completion (%)	10	10	30	25	5
Scale	1	1	2	2	1
Word Recognition Rate (WRR) (%)	35	31	57	52	24
Time cut point: 10s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	10	20	40	15	10
Wrong Spelling Word	3	2	5	7	2
Text completion (%)	15	40	45	20	15
Scale	1	2	3	1	1
Word Recognition Rate (WRR) (%)	38	52	74	40	33
Time cut point: 5s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	30	35	35	25	10
Wrong Spelling Word	4	2	5	5	1
Text completion (%)	30	40	40	30	15
Scale	2	2	2	2	1
Word Recognition Rate (WRR) (%)	51	62	69	49	36

Appendix T: Result of "C10371808"

C10371808 (First minute of recording) **Total words in original text = 80 words**					
Time cut point: 59s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	5	0	0	1	0
Wrong Spelling Word	5	0	0	1	0
Text completion (%)	5	0	0	2	1
Scale	1	1	1	1	1
Word Recognition Rate (WRR) (%)	26	3	1	18	8
Time cut point: 20s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	5	5	10	1	0
Wrong Spelling Word	4	5	5	1	2
Text completion (%)	5	5	10	2	2
Scale	1	1	1	1	1
Word Recognition Rate (WRR) (%)	29	30	34	19	19
Time cut point: 10s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	10	10	15	2	1
Wrong Spelling Word	5	5	2	0	3
Text completion (%)	10	10	20	5	5
Scale	1	1	1	1	1
Word Recognition Rate (WRR) (%)	36	36	43	25	19
Time cut point: 5s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	5	10	30	5	5
Wrong Spelling Word	4	5	6	5	3
Text completion (%)	10	15	30	10	10
Scale	1	1	2	1	1
Word Recognition Rate (WRR) (%)	38	44	48	34	28

Appendix U: Result of "C10411974"

C10411974 (First minute of recording) **Total words in original text = 95 words**					
Time cut point: 59s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	1	25	15	15	5
Wrong Spelling Word	0	4	0	0	3
Text completion (%)	1	25	20	15	5
Scale	1	2	1	1	1
Word Recognition Rate (WRR) (%)	14	41	31	22	17
Time cut point: 20s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	15	15	20	5	1
Wrong Spelling Word	3	4	5	1	0
Text completion (%)	20	20	25	10	2
Scale	1	1	2	1	1
Word Recognition Rate (WRR) (%)	39	34	40	28	15
Time cut point: 10s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	25	25	30	20	5
Wrong Spelling Word	7	5	5	3	1
Text completion (%)	30	30	30	20	5
Scale	2	2	2	1	1
Word Recognition Rate (WRR) (%)	45	43	48	36	14
Time cut point: 5s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	20	25	40	20	2
Wrong Spelling Word	5	7	5	4	0
Text completion (%)	20	25	40	25	5
Scale	1	2	2	2	1
Word Recognition Rate (WRR) (%)	40	48	62	47	15

Appendix V: Result of "C10503847"

C10503847 (First minute of recording) **Total words in original text = 99 words**					
Time cut point: 59s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	1	20	30	20	2
Wrong Spelling Word	0	6	5	5	3
Text completion (%)	1	20	30	20	5
Scale	1	1	2	1	1
Word Recognition Rate (WRR) (%)	5	47	49	46	20
Time cut point: 20s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	5	25	30	15	2
Wrong Spelling Word	7	7	5	6	3
Text completion (%)	10	30	30	15	5
Scale	1	2	2	1	1
Word Recognition Rate (WRR) (%)	21	52	47	34	17
Time cut point: 10s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	15	30	20	15	2
Wrong Spelling Word	9	8	5	4	2
Text completion (%)	15	30	25	20	10
Scale	1	2	2	1	1
Word Recognition Rate (WRR) (%)	32	54	37	31	22
Time cut point: 5s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	20	30	40	25	5
Wrong Spelling Word	10	11	10	9	3
Text completion (%)	25	35	40	30	5
Scale	2	2	2	2	1
Word Recognition Rate (WRR) (%)	40	54	51	47	17

Appendix W: Result of "C9760826"

C9760826 (First minute of recording) **Total words in original text = 91 words**					
Time cut point: 59s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	10	5	35	10	1
Wrong Spelling Word	3	1	7	5	4
Text completion (%)	10	5	35	10	2
Scale	1	1	2	1	1
Word Recognition Rate (WRR) (%)	32	22	54	31	15
Time cut point: 20s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	20	30	35	20	1
Wrong Spelling Word	6	8	3	8	9
Text completion (%)	20	30	40	25	5
Scale	1	2	2	1	1
Word Recognition Rate (WRR) (%)	49	56	59	41	21
Time cut point: 10s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	25	30	40	10	5
Wrong Spelling Word	9	8	3	7	8
Text completion (%)	30	35	40	20	5
Scale	2	2	2	1	1
Word Recognition Rate (WRR) (%)	50	53	62	34	25
Time cut point: 5s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	25	20	40	20	0
Wrong Spelling Word	8	10	5	9	6
Text completion (%)	25	30	40	20	1
Scale	2	2	2	1	1
Word Recognition Rate (WRR) (%)	45	52	59	44	13

Appendix X: Result of "C9798109"

C9798109 (First minute of recording) **Total words in original text = 136 words**					
Time cut point: 59s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	2	5	10	1	0
Wrong Spelling Word	4	5	8	4	0
Text completion (%)	5	10	10	2	1
Scale	1	1	1	1	1
Word Recognition Rate (WRR) (%)	26	37	32	21	4
Time cut point: 20s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	2	20	10	1	1
Wrong Spelling Word	7	8	7	3	2
Text completion (%)	3	20	10	1	2
Scale	1	1	1	1	1
Word Recognition Rate (WRR) (%)	21	39	31	15	20
Time cut point: 10s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	2	15	10	1	2
Wrong Spelling Word	7	7	8	2	2
Text completion (%)	2	15	10	1	2
Scale	1	1	1	1	1
Word Recognition Rate (WRR) (%)	21	35	32	14	17
Time cut point: 5s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	1	5	10	1	1
Wrong Spelling Word	7	5	5	4	2
Text completion (%)	5	5	10	5	1
Scale	1	1	1	1	1
Word Recognition Rate (WRR) (%)	19	29	34	24	18

Appendix Y: Result of "C9802561"

C9802561 (First minute of recording) **Total words in original text = 99 words**					
Time cut point: 59s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	25	50	50	50	10
Wrong Spelling Word	5	8	6	11	7
Text completion (%)	30	50	50	50	20
Scale	2	3	3	3	1
Word Recognition Rate (WRR) (%)	47	78	75	74	37
Time cut point: 20s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	40	60	50	30	1-
Wrong Spelling Word	11	6	6	10	4
Text completion (%)	40	70	50	40	10
Scale	2	3	3	2	1
Word Recognition Rate (WRR) (%)	61	81	75	62	32
Time cut point: 10s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	40	60	50	40	10
Wrong Spelling Word	10	6	7	11	5
Text completion (%)	45	70	50	40	10
Scale	3	4	3	2	1
Word Recognition Rate (WRR) (%)	64	82	73	65	38
Time cut point: 5s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	20	45	50	35	5
Wrong Spelling Word	8	8	9	6	5
Text completion (%)	30	50	50	40	15
Scale	2	3	3	2	1
Word Recognition Rate (WRR) (%)	45	70	71	58	38

Appendix Z: Result of "C9822062"

C9822062 (First minute of recording) **Total words in original text = 105 words**					
Time cut point: 59s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	10	15	5	1	0
Wrong Spelling Word	5	8	3	3	2
Text completion (%)	15	20	5	1	1
Scale	1	1	1	1	1
Word Recognition Rate (WRR) (%)	27	44	19	4	6
Time cut point: 20s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	15	10	15	1	0
Wrong Spelling Word	8	7	6	2	2
Text completion (%)	20	20	20	1	1
Scale	1	1	1	1	1
Word Recognition Rate (WRR) (%)	40	34	42	7	7
Time cut point: 10s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	10	10	15	1	1
Wrong Spelling Word	8	8	5	6	3
Text completion (%)	15	10	20	5	2
Scale	1	1	1	1	1
Word Recognition Rate (WRR) (%)	37	33	43	16	9
Time cut point: 5s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	5	15	10	5	1
Wrong Spelling Word	5	8	7	5	3
Text completion (%)	15	25	20	10	2
Scale	1	1	1	1	1
Word Recognition Rate (WRR) (%)	31	44	35	26	9

Appendix AA: Result of "C9838810"

C9838810 (First minute of recording) **Total words in original text = 121 words**					
Time cut point: 59s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	1	10	25	5	1
Wrong Spelling Word	0	0	5	2	0
Text completion (%)	2	10	30	5	1
Scale	1	1	2	1	1
Word Recognition Rate (WRR) (%)	10	21	43	18	5
Time cut point: 20s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	10	20	25	10	0
Wrong Spelling Word	2	3	3	4	0
Text completion (%)	15	25	25	10	0
Scale	1	1	2	1	1
Word Recognition Rate (WRR) (%)	36	40	40	24	0
Time cut point: 10s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	15	35	25	10	0
Wrong Spelling Word	4	3	5	6	0
Text completion (%)	20	40	30	15	0
Scale	1	2	2	1	1
Word Recognition Rate (WRR) (%)	40	51	42	31	0
Time cut point: 5s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	10	35	30	10	1
Wrong Spelling Word	3	6	8	3	0
Text completion (%)	10	40	30	15	1
Scale	1	2	2	1	1
Word Recognition Rate (WRR) (%)	36	57	46	33	7

Appendix AB: Result of "C9846248"

C9846248 (First minute of recording) **Total words in original text = 91 words**					
Time cut point: 59s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	0	0	15	5	0
Wrong Spelling Word	0	0	1	2	0
Text completion (%)	0	0	15	5	0
Scale	1	1	1	1	1
Word Recognition Rate (WRR) (%)	1	2	33	16	2
Time cut point: 20s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	2	15	10	10	10
Wrong Spelling Word	4	5	4	4	4
Text completion (%)	2	15	10	10	10
Scale	1	1	1	1	1
Word Recognition Rate (WRR) (%)	12	36	24	29	27
Time cut point: 10s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	5	10	15	5	2
Wrong Spelling Word	6	4	6	2	6
Text completion (%)	5	10	20	10	2
Scale	1	1	1	1	1
Word Recognition Rate (WRR) (%)	24	31	40	27	16
Time cut point: 5s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	2	15	20	10	2
Wrong Spelling Word	6	5	7	3	3
Text completion (%)	2	20	20	10	2
Scale	1	1	1	1	1
Word Recognition Rate (WRR) (%)	20	35	44	26	14

Appendix AC: Result of "C9855059"

C9855059 (First minute of recording) **Total words in original text = 75 words**					
Time cut point: 59s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	5	10	2	0	0
Wrong Spelling Word	0	2	3	0	0
Text completion (%)	5	10	2	0	0
Scale	1	1	1	1	1
Word Recognition Rate (WRR) (%)	12	25	16	3	3
Time cut point: 20s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	2	10	2	2	0
Wrong Spelling Word	2	3	1	0	0
Text completion (%)	2	10	2	2	0
Scale	1	1	1	1	1
Word Recognition Rate (WRR) (%)	13	28	11	9	4
Time cut point: 10s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	2	5	5	2	0
Wrong Spelling Word	2	3	2	1	0
Text completion (%)	2	5	5	2	0
Scale	1	1	1	1	1
Word Recognition Rate (WRR) (%)	12	20	20	11	4
Time cut point: 5s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	1	5	2	0	0
Wrong Spelling Word	0	3	3	0	0
Text completion (%)	0	5	2	1	0
Scale	1	1	1	1	1
Word Recognition Rate (WRR) (%)	1	16	16	3	3

Appendix AD: Result of "C9907261"

C9907261 (First minute of recording) **Total words in original text = 111 words**					
Time cut point: 59s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	2	1	0	2	2
Wrong Spelling Word	0	1	1	1	1
Text completion (%)	2	2	0	2	2
Scale	1	1	1	1	1
Word Recognition Rate (WRR) (%)	9	4	3	14	12
Time cut point: 20s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	20	5	25	0	5
Wrong Spelling Word	2	4	4	0	3
Text completion (%)	20	10	25	0	5
Scale	1	1	2	1	1
Word Recognition Rate (WRR) (%)	34	25	42	3	16
Time cut point: 10s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	20	30	25	0	5
Wrong Spelling Word	3	3	3	0	0
Text completion (%)	20	30	25	0	5
Scale	1	2	2	2	1
Word Recognition Rate (WRR) (%)	41	48	42	0	11
Time cut point: 5s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	10	30	30	10	5
Wrong Spelling Word	3	7	7	4	4
Text completion (%)	15	30	30	10	10
Scale	1	2	2	1	2
Word Recognition Rate (WRR) (%)	36	51	48	21	20

Appendix AE: Result of "C9935118"

C9935118 (First minute of recording) **Total words in original text = 79 words**					
Time cut point: 59s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	1	1	5	10	1
Wrong Spelling Word	0	0	0	1	0
Text completion (%)	1	2	5	20	1
Scale	1	1	1	1	1
Word Recognition Rate (WRR) (%)	5	13	5	38	9
Time cut point: 20s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	10	25	10	10	1
Wrong Spelling Word	4	5	1	1	0
Text completion (%)	15	30	20	10	1
Scale	1	2	1	1	1
Word Recognition Rate (WRR) (%)	37	48	27	32	11
Time cut point: 10s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	25	30	10	5	5
Wrong Spelling Word	4	5	2	1	1
Text completion (%)	25	30	20	10	5
Scale	2	2	1	1	1
Word Recognition Rate (WRR) (%)	42	49	35	22	19
Time cut point: 5s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	5	30	40	10	1
Wrong Spelling Word	6	2	5	3	0
Text completion (%)	10	35	40	15	2
Scale	1	2	2	1	1
Word Recognition Rate (WRR) (%)	30	53	56	34	18

Appendix AF: Result of "C9937436"

C9937436 (First minute of recording) **Total words in original text = 79 words**					
Time cut point: 59s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	10	15	40	5	2
Wrong Spelling Word	2	2	7	2	3
Text completion (%)	10	15	40	10	2
Scale	1	1	2	1	1
Word Recognition Rate (WRR) (%)	25	35	54	22	16
Time cut point: 20s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	30	40	35	10	2
Wrong Spelling Word	2	3	1	3	5
Text completion (%)	30	40	40	15	2
Scale	2	2	2	1	1
Word Recognition Rate (WRR) (%)	47	61	49	24	11
Time cut point: 10s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	35	45	40	10	1
Wrong Spelling Word	7	2	1	2	1
Text completion (%)	35	50	40	10	1
Scale	2	3	2	1	1
Word Recognition Rate (WRR) (%)	57	63	54	27	11
Time cut point: 5s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	25	45	50	25	5
Wrong Spelling Word	8	4	3	3	4
Text completion (%)	30	50	50	25	5
Scale	2	3	3	2	1
Word Recognition Rate (WRR) (%)	53	61	65	41	11

Appendix AG: Result of "C10052504"

C10052504 (First minute of recording) **Total words in original text = 69 words**					
Time cut point: 59s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	0	0	1	0	0
Wrong Spelling Word	0	0	0	0	0
Text completion (%)	0	0	2	1	0
Scale	1	1	1	1	1
Word Recognition Rate (WRR) (%)	1	1	14	3	0
Time cut point: 20s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	5	1	1	0	0
Wrong Spelling Word	2	2	0	0	0
Text completion (%)	5	2	1	0	0
Scale	1	1	1	1	1
Word Recognition Rate (WRR) (%)	14	9	12	3	1
Time cut point: 10s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	5	7	5	2	2
Wrong Spelling Word	2	2	4	1	1
Text completion (%)	10	10	10	5	2
Scale	1	1	1	1	1
Word Recognition Rate (WRR) (%)	19	23	20	12	6
Time cut point: 5s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	5	10	10	5	2
Wrong Spelling Word	1	3	2	2	1
Text completion (%)	5	10	10	5	2
Scale	1	1	1	1	1
Word Recognition Rate (WRR) (%)	16	29	29	17	13

Appendix AH: Result of "C10063278"

C10063278 (First minute of recording) **Total words in original text = 125 words**					
Time cut point: 59s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	5	5	15	5	1
Wrong Spelling Word	4	6	7	6	3
Text completion (%)	10	10	20	10	1
Scale	1	1	2	1	1
Word Recognition Rate (WRR) (%)	23	22	28	25	9
Time cut point: 20s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	5	20	20	3	2
Wrong Spelling Word	5	7	8	4	1
Text completion (%)	10	20	20	3	2
Scale	1	1	1	1	1
Word Recognition Rate (WRR) (%)	21	35	30	15	9
Time cut point: 10s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	25	25	25	2	1
Wrong Spelling Word	8	8	6	4	2
Text completion (%)	30	25	25	3	1
Scale	2	2	2	1	1
Word Recognition Rate (WRR) (%)	39	38	36	13	6
Time cut point: 5s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	20	25	25	10	2
Wrong Spelling Word	6	9	8	3	4
Text completion (%)	20	30	30	10	5
Scale	1	2	2	1	2
Word Recognition Rate (WRR) (%)	30	42	42	19	13

Appendix AI: Result of "C10077643"

C10077643 (First minute of recording) **Total words in original text = 94 words**					
Time cut point: 59s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	1	5	5	1	0
Wrong Spelling Word	2	1	1	0	0
Text completion (%)	1	5	5	1	0
Scale	1	1	1	1	1
Word Recognition Rate (WRR) (%)	6	13	13	4	2
Time cut point: 20s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	5	5	5	0	0
Wrong Spelling Word	2	2	1	0	0
Text completion (%)	5	5	5	0	0
Scale	1	1	1	1	1
Word Recognition Rate (WRR) (%)	12	12	17	2	0
Time cut point: 10s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	5	7	5	0	0
Wrong Spelling Word	3	2	3	0	0
Text completion (%)	5	7	5	1	0
Scale	1	1	1	1	1
Word Recognition Rate (WRR) (%)	16	20	19	9	1
Time cut point: 5s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	5	10	10	7	0
Wrong Spelling Word	5	3	2	2	0
Text completion (%)	5	10	10	10	0
Scale	1	1	1	1	1
Word Recognition Rate (WRR) (%)	23	26	28	22	1

Appendix AJ: Result of "C10078155"

C10078155 (First minute of recording) **Total words in original text = 88 words**					
Time cut point: 59s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	0	0	0	10	5
Wrong Spelling Word	3	3	4	5	9
Text completion (%)	0	1	0	10	5
Scale	1	1	1	1	1
Word Recognition Rate (WRR) (%)	0	3	0	33	26
Time cut point: 20s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	10	20	20	25	2
Wrong Spelling Word	4	2	5	7	6
Text completion (%)	10	20	20	25	2
Scale	1	1	1	2	1
Word Recognition Rate (WRR) (%)	26	38	31	44	16
Time cut point: 10s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	20	35	45	30	10
Wrong Spelling Word	3	5	5	8	6
Text completion (%)	20	40	50	30	10
Scale	1	2	3	2	1
Word Recognition Rate (WRR) (%)	36	56	58	48	22
Time cut point: 5s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	30	45	50	45	20
Wrong Spelling Word	3	2	7	7	4
Text completion (%)	30	50	50	45	25
Scale	2	3	3	3	2
Word Recognition Rate (WRR) (%)	44	59	61	53	31

Appendix AK: Result of "C10081916"

C10081916 (First minute of recording) **Total words in original text = 110 words**					
Time cut point: 59s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	1	30	20	10	5
Wrong Spelling Word	0	7	2	0	0
Text completion (%)	1	30	20	10	5
Scale	1	2	1	1	1
Word Recognition Rate (WRR) (%)	5	47	35	25	14
Time cut point: 20s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	10	30	30	10	0
Wrong Spelling Word	5	8	4	0	0
Text completion (%)	10	40	30	10	0
Scale	1	2	2	1	1
Word Recognition Rate (WRR) (%)	25	50	41	19	8
Time cut point: 10s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	25	30	30	10	1
Wrong Spelling Word	4	10	6	2	0
Text completion (%)	30	30	35	10	1
Scale	2	2	2	1	1
Word Recognition Rate (WRR) (%)	39	45	45	26	5
Time cut point: 5s					
Speed	**0.5X**	**0.75X**	**1X**	**1.25X**	**1.5X**
Text understanding (%)	20	30	40	10	5
Wrong Spelling Word	9	11	11	4	2
Text completion (%)	20	30	40	10	5
Scale	1	2	2	2	1
Word Recognition Rate (WRR) (%)	36	46	55	30	14