DEVELOPMENT OF INTERVIEW TRANSCRIPTS APPLICATION USING SPEECH RECOGNITION METHOD

Muhammad Elra Ghifary R[1], Richi Dwi Agustia[2]

[1,2] Informatics Engineering Program – Computer University of Indonesia
St. Dipatiukur No. 112-114, District of Coblong, Bandung City, West Java
E-mail: elraghifary@gmail.com[1], richi@email.unikom.ac.id[2]

ABSTRACT

PT Kadence International conducts interviews to obtain research data, in the form of transcripts, made by the transcriber. After the interview is over, there is still a conversation that has yet to be written the transcriber and cannot give the transcript to the moderator directly when the interview is finished. This means that the transcript can exceed the specified time period to transcribe the interview. The system used in storing transcripts uses local storage which may result in the loss of transcripts required or even misplacement of transcripts. A system was proposed to fix this issue using speech recognition to convert voice speech into written words as the interview progresses. This would utilize Google's Web Speech API technology combined with noise suppression technology to reduce noise from the voice of conversation between moderator and respondent. Transcript management system built on web technology. The result of the research is the development of interview transcripts application using speech recognition method. Testing for word accuracy found to be around 95% accurate, the technology using speech recognition was found useful. In conclusion, the interview transcripts application using speech recognition method can facilitate the transcriber in the making and managing of interview transcripts.

Keywords: Transcripts, Interview, Web Application, Speech Recognition Method, Noise Suppression

1. PRELIMINARY

1.1 Background

PT Kadence International is a consumer survey company engaged in marketing research where interviews are conducted to obtain research data on a research study. Interviews were conducted in the form of In-depth Interview and Focus Group Discussion where moderators were interviewers, respondents as interviewees, and transcriber as persons who transcribe the content of the interview. Moderator duty to lead the interview and provide questions to respondents to obtain research data in the form of transcripts. As evidence of interviews, there are voice recorder and camera tools used to record interview activities. After the interview finished, the transcriber gets the result of the interview in the form of sound recordings. The transcript which is a copy of the content of the interview conversation is done by the transcriber during the interview and the transcript is used as the data to be analyzed by the moderator which is then reported to the client who has the role of the person providing the research study. Transcriber works on transcripts using Microsoft Word software and then stored on hard drive storage media and is only given two to three days to collect transcripts to moderators.

Based on the interview with Mrs. Ely as a transcriber from PT Kadence International, it is said that the transcriber cannot give the transcript at the end of the interview because of a missed conversation. Transcriber listens to the recorded audio of the interview and then proceeds to transcript the results of the previous interview to fill in the missed conversation. Thus, the process of listening to the content of an interview conversation in accordance with the duration of the sound recording can result in the transcript over time. The management of transcripts that still use local storage may cause the loss of transcripts required where the transcriber says this happens when the computer is damaged.

The Notula Rapat application which is an automatic transcription system to convert a speech into writing directly (real-time) in Bahasa Indonesia with a microphone as input. Notula Rapat has the features to manage the necessary transcription process meetings such as leadership meetings, focus group discussions, interviews, and others using speech recognition technology. [1] Speech recognition is a technology that can convert the input of speech into text that can be read. Speech recognition can identify the spoken voice by digitizing signals and matching with the grammar that is already defined within a device. The use of speech recognition technology produces accuracy in recognizing spoken words

about 90 percent and has the word error rate of 9 percent by using the Google API service. [2] Based on this background, an application was built that could create an interview transcript using speech recognition technology. Thus, the authors take the topic to be a research entitled "Development of Interview Transcription Application Using Method Speech Recognition".

1.2 Problem Identification

Based on the above background explanation, it can be formulated a problem that is how to make an application that can transcribe interview using speech recognition method also there is a utilization of noise suppression.

1.3 Purpose and Objectives

The purpose of this research is to build an interview transcription application using speech recognition method. The expected goals of this study are as follows.

1. To facilitate the transcriber in making interview transcripts using speech recognition technology.
2. To help the transcriber to able to manage interview transcripts well.

1.4 Software Development Methodologies

Model in software development using a prototype model because it is in accordance with the stages undertaken in system development. The prototype model has several stages which are Requirements gathering, Quick design, Build a prototype, Customer evaluation of prototype, and Refine requirements incorporating customer suggestions. The prototype model is shown in Figure 1.1.

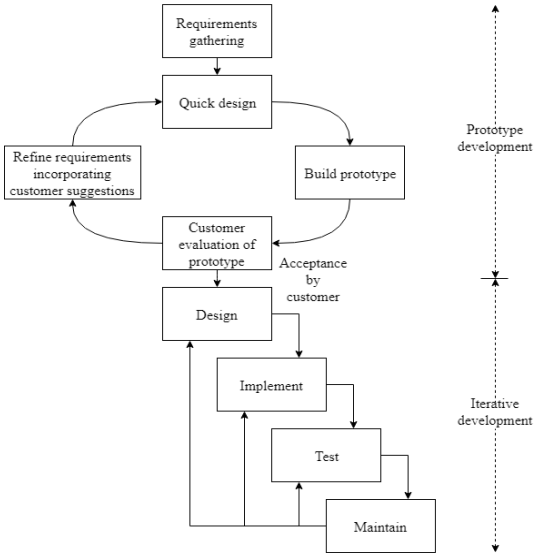


Figure 1.1 Prototype Model
(Reference : Fundamentals of Software Engineering, 2014)

2. CONTENT

2.1 Theoretical Base

Here are the theories used in the system to be built.

2.1.1 Interview

An interview is a process of communication interaction conducted by the interviewer and the interviewees on the basis of availability with reference to the intended purpose. The purpose of the interview process is to understand. In the process of understanding, required the ability to interview can run well. Ability in the process of understanding is the ability to motivate, verbal skills, listening ability, nonverbal skills, ability to draw conclusions, and writing skills. [3] The requirements that need to be done when going to interview in the context of research are as follows. [3]

1. Researchers need to pave the way and find the subject to be interviewed.
2. Researchers shall create a conducive communication to the subject of research.
3. Researchers need to keep the relationship in order to always be stable and have a positive climate.
4. Trust is something that is absolutely achievable.
5. Keep maintaining personal and professional distance with the subject of research.

Semi-structured interviews are more appropriate for qualitative research because researchers are given the freedom to ask questions and can manage the flow and set of interviews. Researchers rely on guidelines (guideline) interviews to dig the data. An example of a semi-structured interview is In-Depth Interview. [3]

2.1.2 Transcript

A transcript is a copy or note of each word from a source for a particular purpose. The transcripts are more reliable and accountable because everything discussed in the interview has been diverted into scripts that are easy to analyze. A scripting language is a language it is without any manipulation or language refinement. Resources are obtained from a sound or video recording and the contents are copied in writing. [3]

2.1.3 Speech Recognition

Speech Recognition is a process and related technology for converting a speech or speech into the form of appropriate words or entities through a specific algorithm implemented in the device or computer. The speech recognition app is included voice dialing, call routing, interactive voice

response, data entry and dictation, voice command and control, gaming, structured document, appliance control by voice, computer-aided language learning, content-based spoken audio search, and robotics. [4]

In recent years, speech recognition technology has changed the way humans work in interacting with computers. Dong Yu and Li Deng classify speech recognition applications to help humans interact with computers, that is: [5]

1. Human-Human Communication (HHC)

Speech technology can remove the barrier in the interaction between humans and humans. For example, people who cannot speak Japanese will find it difficult to live in Japan, this limitation can be reduced by using a speech-to-speech interpreter system. ASR technology can also be used to dictate short messages and recognize a teacher's conversation in order for students to easily find the information they need.

2. Human-Machine Communication (HMC)

Speech technology plays a major role in the development of interaction between man and machine. Applications that are often recognized are voice search, personal digital assistants (PDA), gaming, and living room interaction systems. Voice search makes it easy for users to search for information by voice. Personal digital assistants (PDA) is a system that can read the information stored in the device, one of which is Siri from Apple. Siri can accept commands entered by voice. Experience in playing games can increase when using speech technology. Speech technology can also control the existing device in the living room by voice.

The speech recognition architecture has four main components shown in Figure 2.**1**.

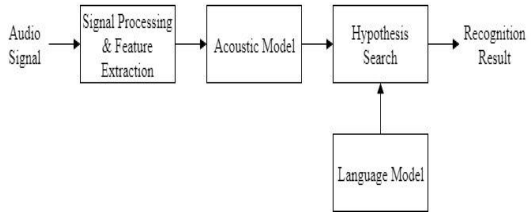


Figure 2.1 Speech Recognition Architecture
(Reference : Automatic Speech Recognition; A Deep Learning Approach, 2015)

2.1.4 Noise

Noise can be defined as an unwanted noise that can interfere with one's comfort. Noise can be in the form of a whisper, a sound of a vehicle, or a sound in the street. Processed noise levels using decibels with noise type babble noise, car noise, white noise, and street noise. [6]

Noise can be reduced by noise reduction or noise suppression process. Noise Suppression is the process of taking noise signals and reducing noise as much as possible during the running sound. The use of noise suppression can be aimed at VoIP or video conference applications and speech recognition applications. Implementation in speech recognition can be done in real-time by receiving voice input through a microphone and then running noise suppression so as to make a sound more easily heard and understood (denoised). The simple concept of noise suppression is shown in Figure 2.2.

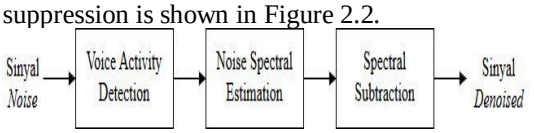


Figure 2.2 The Concept of Noise Suppression
(Reference : A Hybrid DSP/Deep Learning Approach to Real-Time Full-Band Speech Enhancement, 2017)

2.1.5 Web Speech API

The Web Speech API is a javascript web browser specification that allows developers to combine speech recognition and speech synthesis systems into a website. Developers can also use a set of lines of code to convert sound into text output (speech-to-text) and use speech recognition technology as input from a form, dictation, and voice control. Use of the Web Speech API must be permitted by the developer or user through a web browser where the browser shows the use of the API in the form of a notification. Currently, the Web Speech API supports 32 languages where one of them is Indonesian. [7]

2.1.6 RNNoise

RNNoise is a project that implements noise suppression by combining DSP, deep learning and recurrent neural networks (RNN) techniques to reduce noise. The main target of the RNNoise project is in the realm of speech recognition applications such as video-conferencing with low complexity. The design and training of RNNoise data are developed in Python using the deep learning library Keras. With the goal of creating a real-time system, RNNoise is implemented in C language. RNNoise provides a javascript module that can be run through the browser to reduce the sound noise recorded from the microphone in real-time. [6]

2.2 Analysis of Current System

Analysis of current system describes ongoing activities or activities at the research site. The current procedure is divided into three, that is:

1. Procedures on how clients provide research studies so that interviews can be implemented.

2. Procedures on how to transcribe the interview.
3. Procedures on how the moderator reports the transcript of the interview to the client.

2.2.1 Interviewing Procedure

The process that occurs when the client provides a research study until the interview is performed is shown in the notation Business Process Model in Figure 2.3.

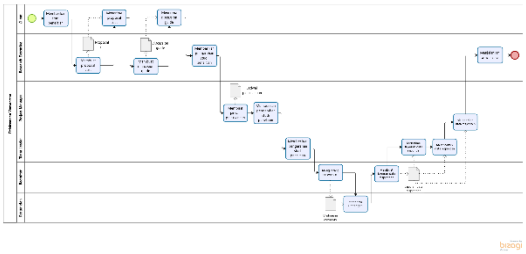


Figure 2.3 Interviewing Procedure

Explanation of BPMN for interviewing procedure is as follows.

1. The client provides research or research studies to the Research Executive.
2. Research Executive makes a cost proposal to conduct research and produce cost proposal documents then report it to Client.
3. The client receives proposal cost document.
4. Research Executive makes discussion guide about the research that will be done and generate the discussion guide document then report it to Client.
5. The client receives document discussion guide.
6. Research Executive provides research briefing to Project Manager.
7. Based on the direction of the Research Executive, the Project Manager creates an interview schedule and produces an interview schedule document.
8. The Project Manager provides guidance and information regarding the implementation of the interview to the Team Leader.
9. The Team Leader provides guidance to the Recruiter on the criteria of the respondent to be had.
10. Recruiter finds and invites respondents to interviews with interview invitation documents.
11. Respondents receive invitations to conduct interviews.
12. Recruiter makes a report of respondent data to be interviewed and produce document report of respondent data then report it to Team Leader.
13. Team Leader receives a report of respondent's data.
14. The Team Leader reports the respondent's data to the Project Manager that the interview is ready to be performed.
15. Project Manager reports the respondent's data to the Research Executive.
16. Research Executive runs the interview.

2.2.2 Making Transcript Procedure

The process that occurred during the interview until the formation of the transcript is shown in the notation Business Process Model in Figure 2.4.

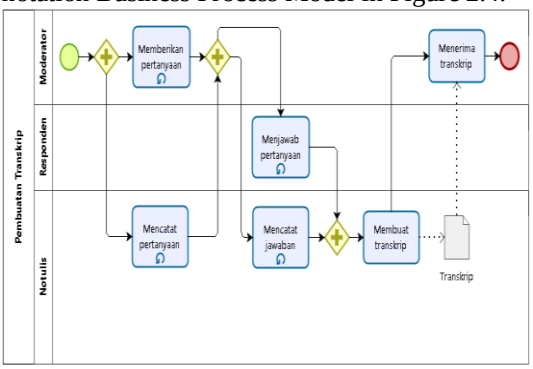


Figure 2.4 Making Transcript Procedure

Explanation of BPMN for making transcript procedure is as follows.

1. Moderator asks the respondent the question and the transcriber write the question where the two processes are running concurrently and repeatedly.
2. Respondents answer the questions and the note keeps answering the answers where both processes run concurrently and repeatedly.
3. Transcriber makes transcripts based on a composite of moderator questionnaires and respondent records to produce transcripts and then transcripts are given to the moderator.
4. Moderator receives interview transcript.

2.2.3 Reporting Transcript Procedure

The process of reporting transcripts is shown in the notation Business Process Model in Figure 2.5.

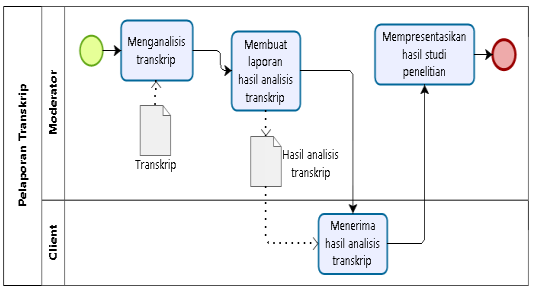


Figure 2.5 Reporting Transcript Procedure

Explanation of BPMN for reporting transcript procedure is as follows.

1. Moderator analyzes transcripts obtained from the transcriber.
2. Moderator creates a report based on the results of the transcript analysis and produces the transcripts analysis documents and then reports them to the Client.
3. The client receives a report of transcript analysis result.
4. Moderator presents the results of research studies.

2.3 Functional Requirements Analysis

Functional requirements analysis is the stage of defining the processes that the system should provide. [8] The analysis used the DFD model with stages consisting of database analysis, a context diagram, level 1 DFD, and process specification.

2.3.1 Database Analysis

Database analysis used in the application to be built is Entity Relationship Diagram (ERD). ERD describes data in the form of entities, attributes, and connectivity between entities. [9] Figure 2.6 shows the ERD of the application to be built.

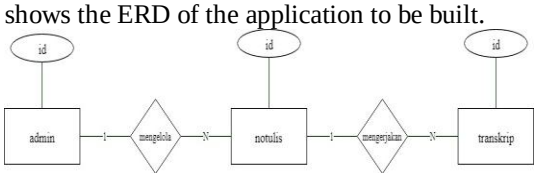


Figure 2.6 Entity Relationship Diagram

2.3.2 Context Diagram

Context diagram describes broadly the top level of all system data streams. [8] Figure 2.7 shows the context diagram of the application to be built.

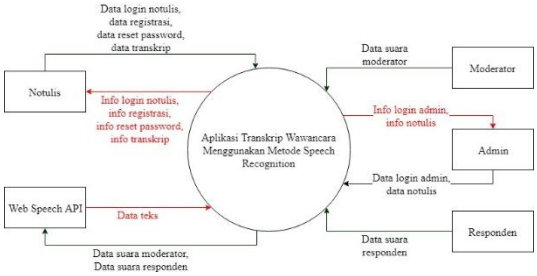


Figure 2.7 Context Diagram

2.3.3 Data Flow Diagram (DFD)

DFD shows the processing or functioning of the system and the exchange of data between the functions that exist on the system. DFD starts from level 1 DFD which is a process decomposition from a context diagram that describes the general processes that exist on the system. [8] Figure 2.8**Error! Reference source not found.** shows the level 1 DFD of the application to be built.

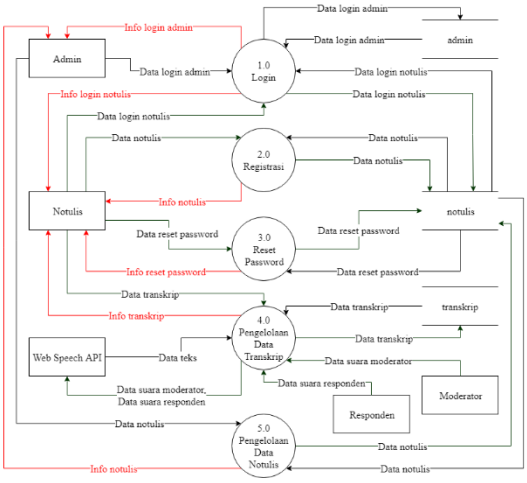


Figure 2.8 Level 1 DFD

2.4 Menu Structure Design

Menu structure design describes the menus are available and connected to the system to be built. Menu structure contained in the application to be built that is transcriber menu structure and admin menu structure.

2.4.1 Transcriber Menu Structure Design

Transcriber menu structure design is a menu design that will be used by the transcriber. Figure 2.9 shows the menu structure design for the transcriber on the application to be built.

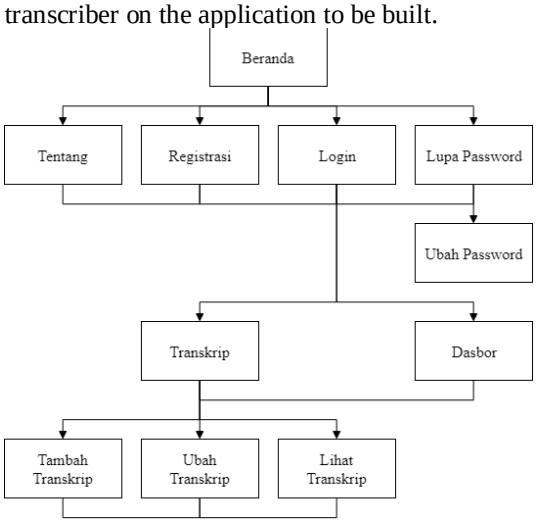


Figure 2.9 Transcriber Menu Structure Design

2.4.2 Admin Menu Structure Design

Admin menu structure design is a menu design that will be used by the admin. Figure 2.10 shows the menu structure design for the admin on the application to be built.

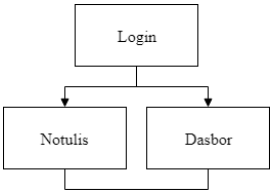


Figure 2.10 Admin Menu Structure Design

2.5 System Implementation

System implementation is a process how to implement the system based on the design of the previous system design stage. Web-based applications will be stored on the hosting server under the name Elrascribe.

2.5.1 Hardware Implementation

The hardware specifications used to run the Elrascribe application can be seen in Table 2.1.

Table 2.1 Hardware Implementation

No.	Hardware	Specification
1	Processor	Dual Core 2 Ghz
2	RAM	2048 MB
3	Operation System	Windows 7
4	Monitor	14 inch
5	Mouse / Keyboard	Standard
6	Microphone	Standard

2.5.2 Software Implementation

The software specifications used to run the Elrascribe application can be seen in Table 2.2.

Table 2.2 Software Implementation

No.	Software	Specification
1	Operation System	Windows 7
2	Internet	10 Mbps Speed
3	Web Browser	Google Chrome
4	Word Processor	Word 2010

2.5.3 Server Implementation

The server specifications used to run the Elrascribe application by online can be seen in Table 2.3.

Table 2.3 Server Implementation

No.	Server Requirements	Specification
1	Operation System	Linux
2	HTTP Server	Apache 2.4.29
3	PHP Version	PHP 7.2.2
4	MySQL Version	10.1.30-MariaDB
5	Network Protocol	HTTPS
6	Storage Space	500 MB

2.5.4 Interface Implementation

Interface implementation build interfaces based on the design at the interface design stage. Based on the interface design results, there are two interface pages namely interface for transcriber and interface for admin. Figure 2.**11** shows the implementation of the Elrascribe application.

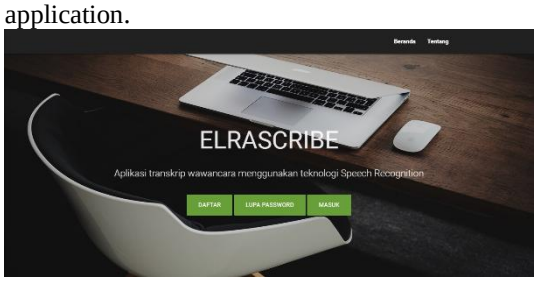


Figure 2.11 Home Interface Implementation

2.6 System Testing

System testing is the stage to find deficiencies or errors in the built system that is done to check whether the system can be run in accordance with the functionality that is designed and has been in accordance with what the purpose of the system development.

2.6.1 Black Box Testing

Black box testing is a system test that focuses on the functional specifications of the built system.

Based on the results of black box testing, it can be concluded that the functionality contained in the system that is built that Elrascribe application has been running well and in accordance with the expected functional requirements. Black box test results can be seen in Table 2.4.

Table 2.4 Black Box Testing of Add Transcript

Cases and Test Results (Data True)			
Input Data	**Which is Expected**	**Observation**	**Conclusion**
Proyek : Hugs IDI : 16 Tanggal : 2017-09-6 Hari : Rabu Waktu : 10:00 Moderator : Mesi Kriteria : Pria / 35 tahun Isi Transkrip :	The system will receive a data entry form added transcript and save the added data transcript to the database then display the message "Sukses! Data telah berhasil ditambah."	The system receives the data entry form added transcript and saves the added data transcript to the database then displays the message "Sukses! Data telah berhasil ditambah."	Accepted
Cases and Test Results (Data False)			
Input Data	**Which is Expected**	**Observation**	**Conclusion**
Proyek : Hugs IDI : enam belas Tanggal : 2017-09-6 Hari : Rabu Waktu : 10:00 Moderator : Mesi Kriteria : Pria / 35 tahun Isi Transkrip :	The system will refuse the contents of the data-added form transcript and display the message "Silahkan masukkan angka saja."	The system rejects the contents of the data form added transcript and displays the message "Silahkan masukkan angka saja."	Accepted
Proyek : Hugs IDI : 16 Tanggal : 2017-09-6 Hari : Rabu Waktu : 10:00 Moderator : Mesi11 Kriteria : Pria / 35 tahun Isi Transkrip :	The system will refuse the contents of the data-added form transcript and display the message "Silahkan masukkan huruf dan spasi saja."	The system rejects the contents of the data form added transcript and displays the message "Silahkan masukkan huruf dan spasi saja."	Accepted
Cases and Test Results (Data Empty)			
Input Data	**Which is Expected**	**Observation**	**Conclusion**
Proyek : IDI : Tanggal : 2017-09-6 Hari : Rabu Waktu : 11:00 Moderator : Kriteria : Isi Transkrip :	The system will refuse the contents of the data-added form transcript and display the message "Kolom ini tidak boleh kosong."	The system rejects the contents of the data form added transcript and displays the message "Kolom ini tidak boleh kosong."	Accepted

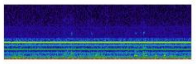
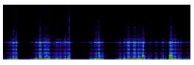
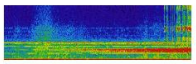
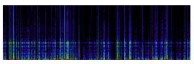
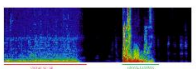
2.6.2 Noise Suppression Testing

The noise suppression test is a test of the moderator and respondent noise reduction process which results will be used at the speech recognition stage. The noise suppression test will

use the RNNoise module run through the Google Chrome web browser.

Based on the results of noise suppression testing, it can be concluded that the use of RNNoise module to reduce noise inbuilt applications can run well. The result of noise suppression testing can be seen in Table 2.5.

Table 2.5 Noise Suppression Testing

Aspect	Information	
Noise Suppression	None	RNNoise
Condition	A quiet room	A quiet room
Spectrogram		
Aspect	**Information**	
Noise Suppression	None	RNNoise
Condition	Vehicle sound	Vehicle sound
Spectrogram		
Aspect	**Information**	
Noise Suppression	None	RNNoise
Condition	Speech "terima kasih"	Speech "terima kasih"
Spectrogram		

2.6.3 Speech Recognition Testing

Testing speech recognition method by using service from Web Speech API is done by reading a text of the conversation between moderator and respondent that aims to know the accuracy of speech recognition process that converts speech into text form. To calculate the accuracy using the following calculation formula.

$$Accuracy = \frac{Total\ of\ recognized\ words}{Total\ of\ words\ tested}$$

The scenario or test plan of the speech recognition method will use the voice of a conversation between the moderator and the respondent. Testing is done by talking in accordance with existing text conversations using a microphone.

Based on the result of testing of speech recognition method, it can be concluded that speech recognition process found in built system has been run and has good accuracy but still not able to display punctuation marked on output. The results of speech recognition method can be seen in Table 2.6.

Table 2.6 Speech Recognition Testing

Total of Words	173
Total of Recognized Words	166
Total of Wrong Words	6
Total of Unknown Words	1
Accuracy (%)	95.9

2.6.4 Voice Pause Testing

In the voice pause test, with voice input to be recognized through the microphone, the word recognition process can run. However, when the word recognition process does not recognize the incoming sound from the microphone, the speech recognition engine will be dismissed so that the transcriber must run the process again.

Based on the test sound interval then the results obtained is the service of speech recognition will wait for the incoming voice for 6 seconds and the service will inactive when there is a pause for 7 seconds. The result of voice pause test can be seen in Table 2.7.

Table 2.7 Voice Pause Testing

Pause Duration (s)	Service Status
1	Active
2	Active
3	Active
4	Active
5	Active
6	Active
7	Inactive

2.6.5 Internet Speed Testing

The application requires an internet network to be able to display web pages and recognize the word of the sound coming through the microphone. Internet speed testing is done by using Chrome DevTools available in the Google Chrome browser to see the duration of the process of loading pages and data sent during the word recognition process. The page being tested is a page of an addition of transcript data because this page includes the process of making an interview transcript using speech recognition method. The data sent is a voice with the words "Thank you very much for your time" obtained through the microphone.

Based on the results of internet speed testing it can be concluded that the faster the speed of the internet used then the duration to load the page will be faster. Judging from the data sent can be said internet speed does not affect the process of recognition of words because the recorded data is a measure of the number of words to be recognized. Internet speed test results can be seen in Table 2.8.

Table 2.8 Internet Speed Testing

Internet Speed	Load Page Time	Data Sent
1 Mb/s	17 seconds	708 bytes
2 Mb/s	10 seconds	656 bytes
4 Mb/s	6 seconds	706 bytes
8 Mb/s	4 seconds	615 bytes

3. CLOSING

3.1 Conclusion

Based on the results of analysis, design, and testing, it can be concluded is as follows.

1. Interview transcripts application using speech recognition method can facilitate the

transcriber in making the transcript of the interview.
2. Interview transcripts application using speech recognition method can help the transcriber to be able to manage transcripts well.

3.2 Suggestion

Suggestions for development of interview transcripts application using speech recognition method are as follows.

1. The application can recognize more unknown words.
2. Add punctuation to the output of speech recognition results.
3. The application can differentiate between moderator and respondent voices.

BIBLIOGRAPHY

[1] Bahasakita, "Notula Rapat," PT Bahasa Kinerja Utama, 2015. [Online]. Available: http://bahasakita.co.id/. [Accessed 18 Oktober 2017].

[2] V. Kepuska and G. Bohouta, "Comparing Speech Recognition Systems (Microsoft API, Google API And CMU Sphinx)," *Journal of Engineering Research and Application,* vol. 7, no. 3, p. 6, 2017.

[3] M. Haris Herdiansyah, Wawancara, Observasi, dan Focus Groups (Sebagai Instrumen Penggalian Data Kualitatif), Rajawali Pers, 2013.

[4] J. Li, L. Deng, R. Haeb-Umbach and Y. Gong, Robust Automatic Speech Recognition, Elsevier, 2015.

[5] D. Yu and L. Deng, Automatic Speech Recognition, London: Springer-Verlag London, 2015.

[6] J.-M. Valin, "A Hybrid DSP/Deep Learning Approach to Real-Time Full-Band Speech Enhancement," p. 5, 2017.

[7] G. Shires and H. Wennborg, "Web Speech API Specification," 19 Oktober 2012. [Online]. Available: https://dvcs.w3.org. [Accessed 5 Oktober 2017].

[8] R. Mall, Fundamentals Of Software Engineering, 4th ed., PHI Learning, 2014.

[9] Fathansyah, Basis Data, Bandung: Informatika Bandung, 2015.