**THESIS PROPOSAL**



**Mimic Human Speech in Bahasa Indonesia**  
**Using Speech Recognition and Speech Synthesis**

By

Valens Prabagita Ivan Susilo  
001201500075

A Thesis Proposal  
Submitted to the Faculty of Computing  
President University  
in Partial Fulfillment of the Requirements  
for the Degree of Bachelor of Science  
in Information Technology

Cikarang, Bekasi, Indonesia

August, 2018

**Mimic Human Speech in Bahasa Indonesia**  
**Using Speech Recognition and Speech Synthesis**

By

Valens Prabagita Ivan Susilo

Approved:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Dr. Tjong Wan Sen, S.T., M.T

Thesis Advisor

# **TABLE OF CONTENTS**

[**TABLE OF CONTENTS** 1](#_Toc523211369)

[**ABSTRACT** 2](#_Toc523211370)

[**INTRODUCTION** 3](#_Toc523211371)

[**PROBLEM STATEMENT** 4](#_Toc523211372)

[**RELATED WORK** 4](#_Toc523211373)

[1. Lyrebird 4](#_Toc523211374)

[2. Google Translate 4](#_Toc523211375)

[**PROBLEM DESCRIPTION** 6](#_Toc523211376)

[**SOLUTION STRATEGY** 7](#_Toc523211377)

[1. Application Overview 7](#_Toc523211378)

[2. System Development 7](#_Toc523211379)

[**EVALUATION** 9](#_Toc523211380)

[**RESOURCES REQUIRED** 9](#_Toc523211381)

[**TIMELINES** 10](#_Toc523211382)

[**SUMMARY** 10](#_Toc523211383)

[**BIBLIOGRAPHY** 11](#_Toc523211384)

# **ABSTRACT**

Speech recognition and speech synthesis are technologies that can be combined to produce many products related to speech and text including mimic speech. There is a few software related to sound mimic speech, and each of them have its own strength and weakness. Two of the main weakness is the mimic speech in English and the speech vocal just have one variant which is from the system itself. From that, this thesis will discuss about making an application that can mimic human speech in Bahasa Indonesia using speech recognition and speech synthesis. The application will be run on local machine that have connection to MongoDB and can be accessed through website. To use the application, the user can start by accessing the website, then, record the speech given random sentences. The system will recognize the recorded speech, and if the recorded is matched to the random sentences, the sound data will be stored to the database. Generating speech can be done after any sound data is stored.

# **INTRODUCTION**

“Ok Google, play some music”. “Siri, what should I eat for lunch?”. Everyday people use their virtual assistance to boost their activities. People very like to use it because they just asked to their device and then in seconds, the wish is granted. It seems like, people are talking to the computer. The truth is, speech recognition takes big role with the help of machine learning. Google Assistance, Apple Siri, Microsoft Cortana, Amazon Alexa, and others have thousands of speech data to be analysed with the machine learning and they easily add data by collecting people speech from the assistance with permission.

If speech recognition is the process to get data by analysed speech, the opposite of speech recognition is speech synthesis, the process to produce artificial speech. Therefore, speech recognition is known as speech-to-text and speech synthesis is known as text-to-speech. “Hey Cortana, read my email” command make virtual assistance generate speech from the email text. With each technology can produce any kinds software related to speech. Combine both of can produce many more. One of the combinations is mimic human speech.

# **PROBLEM STATEMENT**

I intend to develop website application which can be used to mimic speech in Bahasa Indonesia. The application can recognize the speech and generate speech from text.

# **RELATED WORK**

## Lyrebird

Lyrebird is website application, *https://lyrebird.ai*, that has 3 products: custom voice, vocal avatar, and vocal avatar API. Custom voice is a product to create speech based on real people’s speech, it can control the intonation, expression, and the emotion of the speech. Vocal avatar is a product to create own digital speech by read some English sentences, and then generate any sentences with own digital speech. Vocal avatar API is a product to provide API to use user’s own vocal avatar.

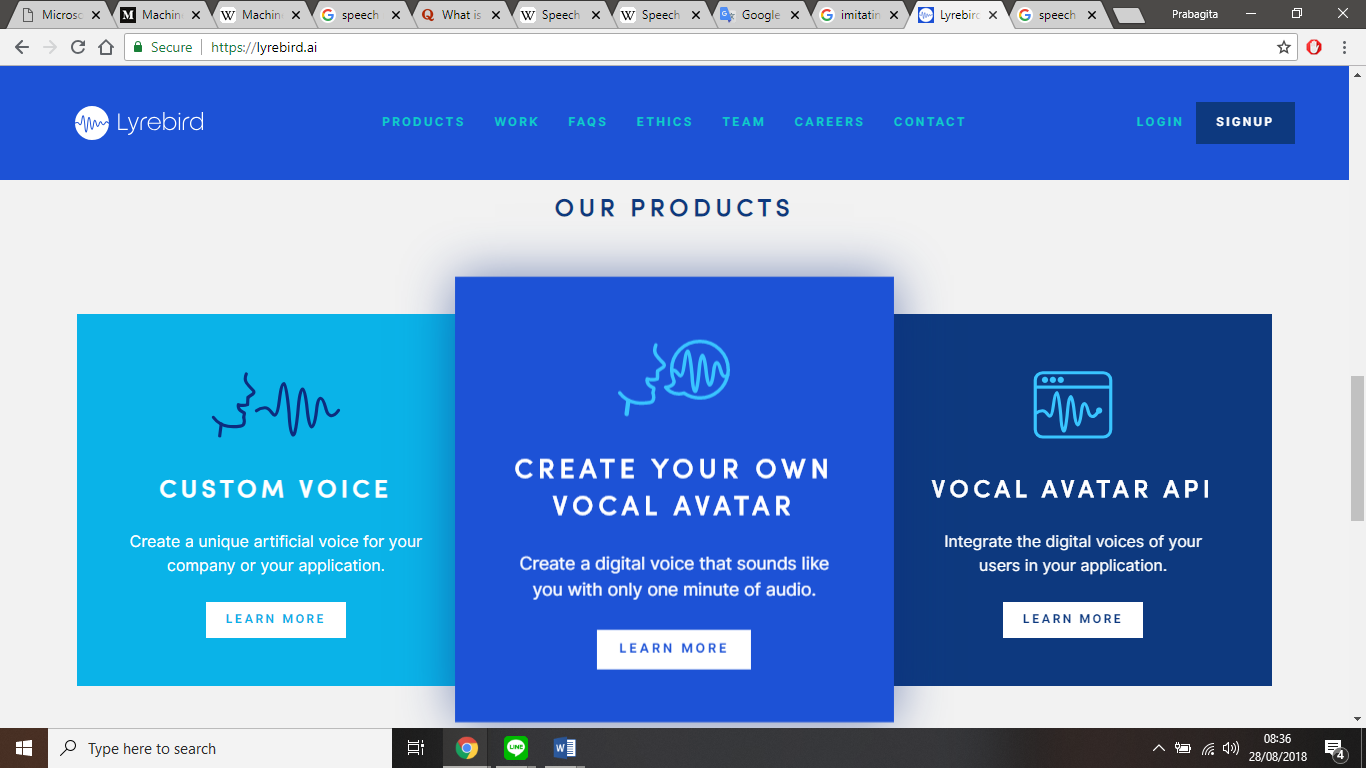


Figure 1. Screenshot of Lyrebird in the website

## Google Translate

Google Translate is an application to translate languages. It has feature to input text from speech and generate speech from text in any languages.

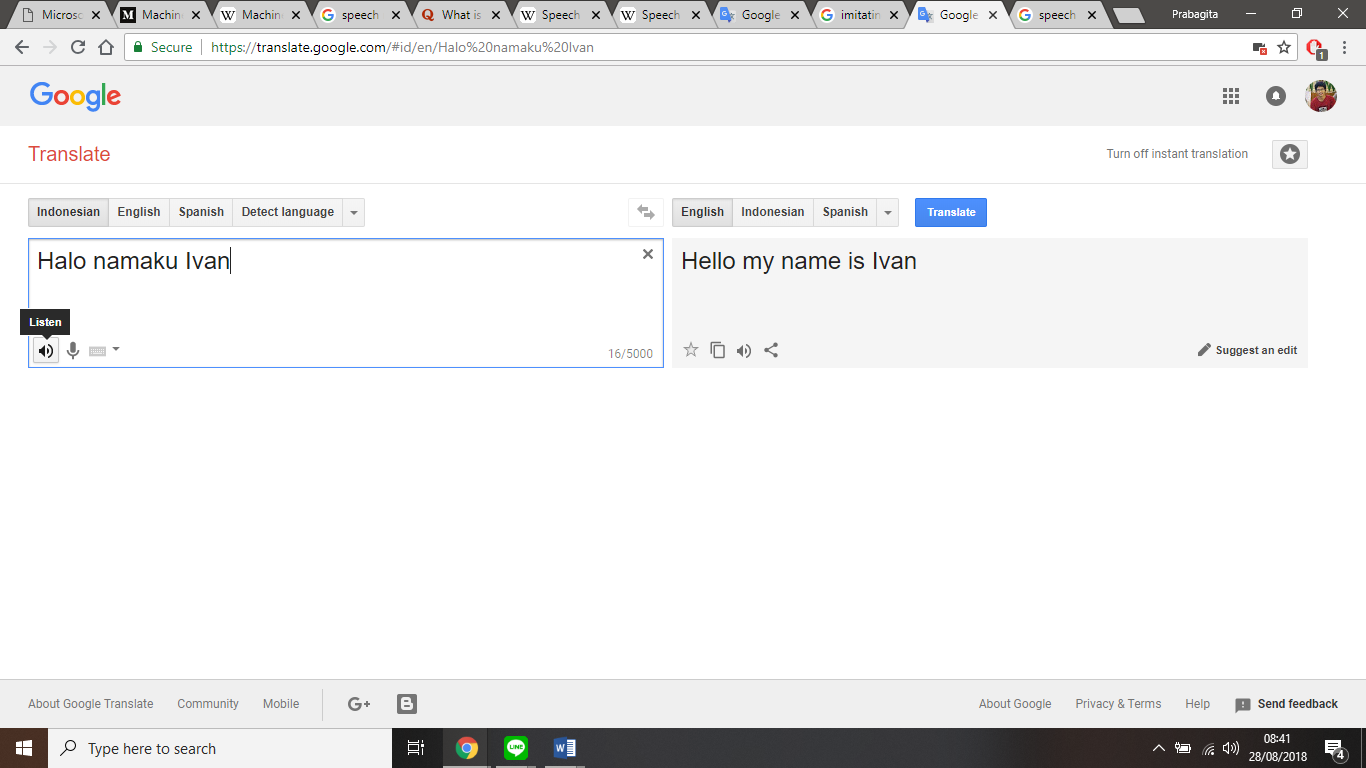


Figure 2. Screenshot of Google Translate in website

# **PROBLEM DESCRIPTION**

Lyrebird can mimic speech with its vocal avatar, but the speech is in English. In the other hand, Google Translate could mimic speech into any languages, but the speech vocal is from the Google Translate itself. Therefore, in this thesis I intent to build application that can mimic speech based on user’s speech vocal in Bahasa Indonesia.

This thesis will have approach using speech recognition and speech synthesis. The application will be trained by several data using machine learning. So that, the application will be able to recognize the speech. The application then provide sentences and user must record and say exactly as the sentences. If the application recognizes the user’s speech same as the text, the user’s speech data will be stored and used to generate speech from user text input.

**SOLUTION STRATEGY**

## Application Overview

The application will be run on local machine that have connection to MongoDB and can be accessed through website. The website has the following features:

* Record speech

This feature is used to record user speech given random sentences. The system will recognize the user record, and if the record is matched to the random sentences, the sound data will be stored to the database.

* Generated speech

This feature is used to generated speech from text based on which user data.

* ID selector

This feature is used to select which user speech data that will be used.

## System Development

The system development life cycles of the application will be using Rapid Application Development divided into 4 phases:

* Requirement Planning

This phase is purposed to collect all speech data from internet, especially YouTube. The collected data will be the first data for machine learning.

* User Design

This phase is purposed to build supervised machine learning to recognize speech from collected data. After the learning time is done, the machine will be tested. If the result is not accepted by criteria, it will need to collect more speech data and learning again. The process will be looping continuously until the machine has accepted by criteria. After that, it will be tested to generated speech manually, if the result is not accepted by criteria, it will need to polish the algorithm or add some speech data to machine learning to have better result.

* Construction

This phase is purposed to build website and its features. After it finished, it will do an integration test.

* Cutover

This phase is purposed to say that the application is ready to go live and defence.

# **EVALUATION**

This application will be evaluated in User Design and Construction. In User Design, it evaluates the machine learning result whether it accepted by criteria or not. After the machine learning is accepted, it will evaluate text-to-speech result whether it accepted by criteria or not. After text-to-speech is accepted, it will evaluate all the process from record the user speech until generate the speech in Construction. Finally, after all is accepted, the application is ready for defence.

# **RESOURCES REQUIRED**

1. Node.js, JavaScript Run-Time Environment.
2. MongoDB, NoSQL document-oriented database.
3. Git as the version control.
4. Visual Studio Code as its IDE.

All the resources required above is open source, so it can get from the internet for free.

# **TIMELINES**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No. | Activity Name | Duration | Start Date | End Date |
| 1. | Requirement Planning | 3 days | September 10, 2018 | September 13, 2018 |
| 2. | User Design | 6 weeks | September 13, 2018 | October 25, 2018 |
| 3. | Construction | 1 weeks | October 25, 2018 | November 1, 2018 |
| 4. | Cutover | 1 days | November 1, 2018 | November 2, 2018 |

# **SUMMARY**

This thesis is about website application that can mimic human speech in Bahasa Indonesia. To use the application, user can start by accessing the website, then record the speech given random sentences. The system will recognize the recorded speech, and if the recorded is matched to the random sentences, the sound data will be stored to the database. Generating speech can be done after any sound data is stored.

# **BIBLIOGRAPHY**

Lyrebird.ai. (2018). *Lyrebird • Ultra-Realistic Voice Cloning and Text-to-Speech*. [online] Available at: https://lyrebird.ai/ [Accessed 28 Aug. 2018].

Translate.google.com. (2018). *Google Translate*. [online] Available at: https://translate.google.com/ [Accessed 28 Aug. 2018].