Speech Recorder and Translator using Google Cloud Speech-to-Text and Translation

[1]J.Y. Chan and [2]H.H. Wang
[1, 2] Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia
email: [1]jietying@outlook.com, [2]hhwang@unimas.my

Date received: 4 December 2020
Date accepted: 4 May 2021
Date published: 28 October 2021

Abstract - *The most popular video website YouTube has about 2 billion users worldwide who speak and understand different languages. Subtitles are essential for the users to get the message from the video. However, not all video owners provide subtitles for their videos. It causes the potential audiences to have difficulties in understanding the video content. Thus, this study proposed a speech recorder and translator to solve this problem. The general concept of this study was to combine Automatic Speech Recognition (ASR) and translation technologies to recognize the video content and translate it into other languages. This paper compared and discussed three different ASR technologies. They are Google Cloud Speech-to-Text, Limecraft Transcriber, and VoxSigma. Finally, the proposed system used Google Cloud Speech-to-Text because it supports more languages than Limecraft Transcriber and VoxSigma. Besides, it was more flexible to use with Google Cloud Translation. This paper also consisted of a questionnaire about the crucial features of the speech recorder and translator. There was a total of 19 university students participated in the questionnaire. Most of the respondents stated that high translation accuracy is vital for the proposed system. This paper also discussed a related work of speech recorder and translator. It was a study that compared speech recognition between ordinary voice and speech impaired voice. It used a mobile application to record acoustic voice input. Compared to the existing mobile App, this project proposed a web application. It was a different and new study, especially in terms of development and user experience. Finally, this project developed the proposed system successfully. The results showed that Google Cloud Speech-to-Text and Translation were reliable to use in video translation. However, it could not recognize the speech when the background music was too loud. Besides, it had a problem of direct translation, which was challenging. Thus, future research may need a custom trained model. In conclusion, the proposed system in this project was to contribute a new idea of a web application to solve the language barrier on the video watching platform.*

Keywords*:* Speech Recognition, Google Speech-to-Text; ASR; Google Cloud Translation;

1 Introduction

YouTube is the second most visited website in the year 2019 (Top Websites Ranking, 2019). The content on YouTube is localized in over 100 countries and can be accessed using 80 different languages (YouTube Offical Blog, 2019). There is a total of one billion hours of video viewed on YouTube every day worldwide. Hence, watching videos online has become an essential daily routine for many people who communicates in different languages. However, not all online videos provide multilingual subtitles. As a result, the audiences who speak a different language from the video could not understand the content of the video. Thus, this study implemented research and development on a web-based speech recorder and translator for the video watching platform.

To build a web-based speech recorder and translator, it uses Automatic Speech Recognition (ASR) technology. The process of Automatic Speech Recognition includes receiving audio input, analyzing the input patterns, and providing a text output (Lai & Yankelovich, 2007). There are diverse ASR applications nowadays. Based on an evaluation report that compares Google Cloud Speech-to-Text, Limecraft Transcriber, and VoxSigma, Google Cloud Speech-to-Text has the lowest Word Error Rate (WER) (Santo, 2017). There is also another research of