ABSTRACT


Automated speech recognition (ASR) system is a machine speech recognition used for recognizing human’s speech in form of statement, command and question. Some well-known ASR systems such as Google and Microsoft have been around for quite a while and proven to be valuable applications (software) which possess almost the same performance with that of a human. This study aims to compare the word error rate (WER) as a benchmark of ASR system’s accuracy of Google and Microsoft. By using audio files retrieved from accent.gmu.edu, the writer manages to test the recognition performance of both Google and Microsoft ASR systems. Both ASR systems recognize two instance of English speech with different accents; one is in American accent and the other one is in Indonesian accent. The result shows that Google and Microsoft makes the same 8.695% WER on English speech with American accent. However, both ASR systems experience a significant rise in WER when recognizing Indonesian accented English speech which go up to 28.985% for Google and 33.333% for Microsoft. Based on the types of error theory proposed by Rani & Girija (2012), Google tends to make type 1 and type 3 errors on American accent and all types of error but type 2 on Indonesian accent. Microsoft, on the other end, is likely to make some type 1 and type 4 errors on American accent and all types of error but type 4 on Indonesian accent. There is definitely a difference in error pattern found in Google and Microsoft ASR systems, but Google has slightly better odds when dealing with Indonesian accent of English speech.

Keywords: automated speech recognition, word error rate, types of error, accent