

ABSTRACT

Lip reading is a method to understand what the speaker said by reading the speakers lips movement. It also called visual speech recognition. Visual speech recognition will produce text from video only. The text is consisting of some words or even sentences that spoken by the speakers. The challenges often encountered in lip readings are variances in inputs, which make speech recognition becomes harder. The variances like facial features and different speed of speech can decrease the accuracy. Nowadays, deep learning provides promising results in extracting visual features. This also provide an opportunity for lip reading to give the better result. In order to be able to used video as the input, this works used 3D Deep Learning architecture. Besides that, out of vocabulary problem also make visual speech recognition harder to apply in the real world. System can only predict the words that exist in the dictionary. However, the vocabulary continues to grow each year, especially in Indonesian. It is almost impossible to fit all the words into system. That is the reason, this research used syllable based to handle that problem. Syllable based would give a chance to build new word that does not exist in dictionary. The combination of existing syllable is used to construct the new word. Because the data obtained too small for deep learning, the augmentation process was done 40 times. As a result, data after augmentation can reach 100% accuracy in the testing process. In the 10 OOV words that were tried, 8 of them could be recognized by the system.

Keywords: lip reading, visual speech recognition,deep learning, out of vocabulary