Abstract

Automatic speech recognition (ASR) is a system to convert speech from recorded audio signal to text. In ASR process, syllabification is one of most important thing. Syllabification is the process to converting recognizable speech signal into syllabic speech units. In Bahasa, syllabification can be applied in phonemes or grapheme sequences. Syllabification that applied in grapheme sequence may hinder any problem such ambiguity and exception in word. This problem can be solved using an approach local statistics based method that is Fuzzy K-Nearest Neighbor in Every Class (FKNNC). This method will defined syllabic point in a word, based on the nearest distance between test data and neighbor for each class and fuzzy ambiguity information to optimize its performance. FKNNC is one of K-Nearest Neighbor (KNN) variants method. This method has been chosen because it can handle the ambiguity of classification generated by KNN. It also an easy and fast classification method and provides high accuracy. This method claim can solve syllabification problem that will be apply in grapheme sequence. Evaluating on 15 k name dataset and 60 k words datasets gives average results 7.928% and 2.484% for syllable error rate (SER).

Keywords: Indonesian syllabification, grapheme, Fuzzy K-Nearest Neighbor in Every Class.