Automatic Speech Recognition using HTK

William L. and Andrea T.

Data Mining and Machine Learning

# Recognition Result with Static MFCC Features

Executing the Perl script that trains and tests the given data gives us a result that is outputted into the *result* directory under the name *recognitionFinalResult*.*res*. The recognition results of the clean tests using a system of only clean training data with static features are shown in Fig. 1.

Graphical user interface, text, application, email

Description automatically generated

Fig. 1. Recognition result without extra features for clean data. No changes in -p flag value.

From Fig. 1, the results on sentence (SENT) and word (WORD) level are very different. These values indicate how well the entire words and sentences are recognised, so it can be deduced that this system is good at recognizing words, but bad at recognizing sentences, at a low accuracy of 63.54%. Additionally, at the word-level result, there is an imbalance between the terms (deletion) and (insertion). That can be changed by modifying *-p* flag value when executing the *HVite* command.

After testing, it was found out that a *–p* value of -15 gave the most balance between *D* and *I* as seen in Fig. 2. It was also discovered that by increasing the -*p* flag, *I* increase, and further increasing -*p* would also increase the difference between *D* and *I.* As expected, decreasing -*p* increases *D* and decreases *I*, allowing us to find a value where *D* and *I* are identical. The balance of *D* and *I* has also increased the accuracy at the sentence-level.

Graphical user interface, text, application, email

Description automatically generated

Fig. 2. Recognition result without extra features for clean data. -p flag = -15.

# Delta and Delta-Delta Features