

## **Abstrak**

Penggunaan video sebagai media semakin marak. Pada tahun 2015, 70% dari pengguna internet mengakses video dan pada tahun 2020 diprediksikan akan mencapai 82% [3]. Dengan jumlah data dan pengguna yang besar tersebut, maka jumlah informasi yang dikandungnya signifikan. Apabila informasi tersebut dapat diolah, maka hasil olahan tersebut dapat digunakan untuk berbagai kebutuhan. Pada tahun 2010, Janakiraman dkk. menunjukan pengenalan ucapan berdasarkan silabel mampu menurunkan kompleksitas perhitungan dan *word error rate*(WER) menjadi 4,4% pada database TIMIT dan 21,2% pada database NTIMIT [5].

*Keyword : ucapan, silabel, audio, video, gerak bibir.*