

## **Abstract**

*Video as a media is widely use. On 2015, 70% from internet users access video and as predicted it would climb up to 82%[3]. With the large number of users and data, the informations could be significant. If those information could be processed, then it could be used to many needs. On 2010, Janakiraman et al. shows that speech recognition based on syllable could reduce calculation complexity and word error rate (WER) to 4,4% on TIMIT database and 21,2% on NTIMIT database [5].*

*Keyword : speech, syllable, audio, video, lips movement.*