



Visual Speech Recognition across Multiple Views in an Automotive Environment

by

Rajitha Navarathna



Overview

- Introduction
- Research Data
- Research Methodology
- Results
- Conclusion



Introduction



Overview

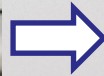


**Ultimate
Result ??**

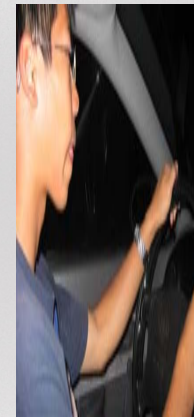




Voice Commands



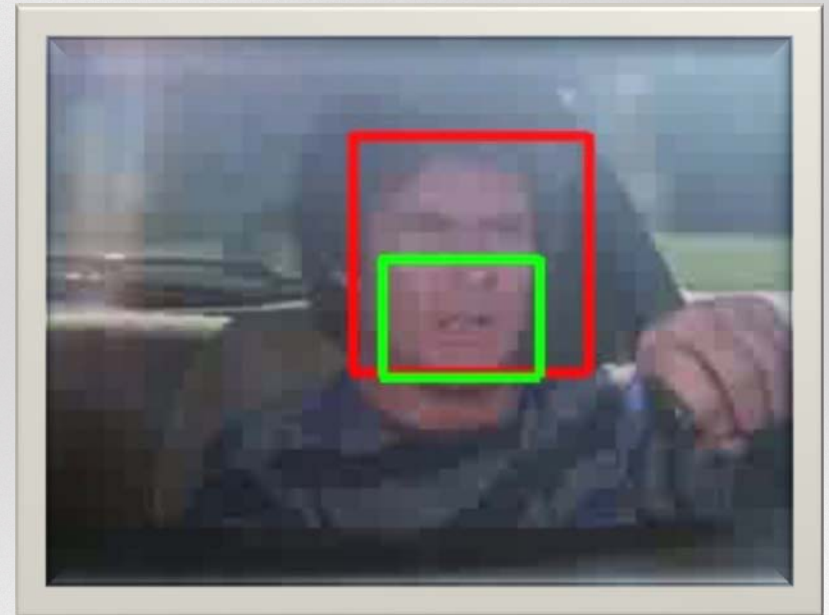
**But , vehicles
are in noisy
environments**





Visual Information

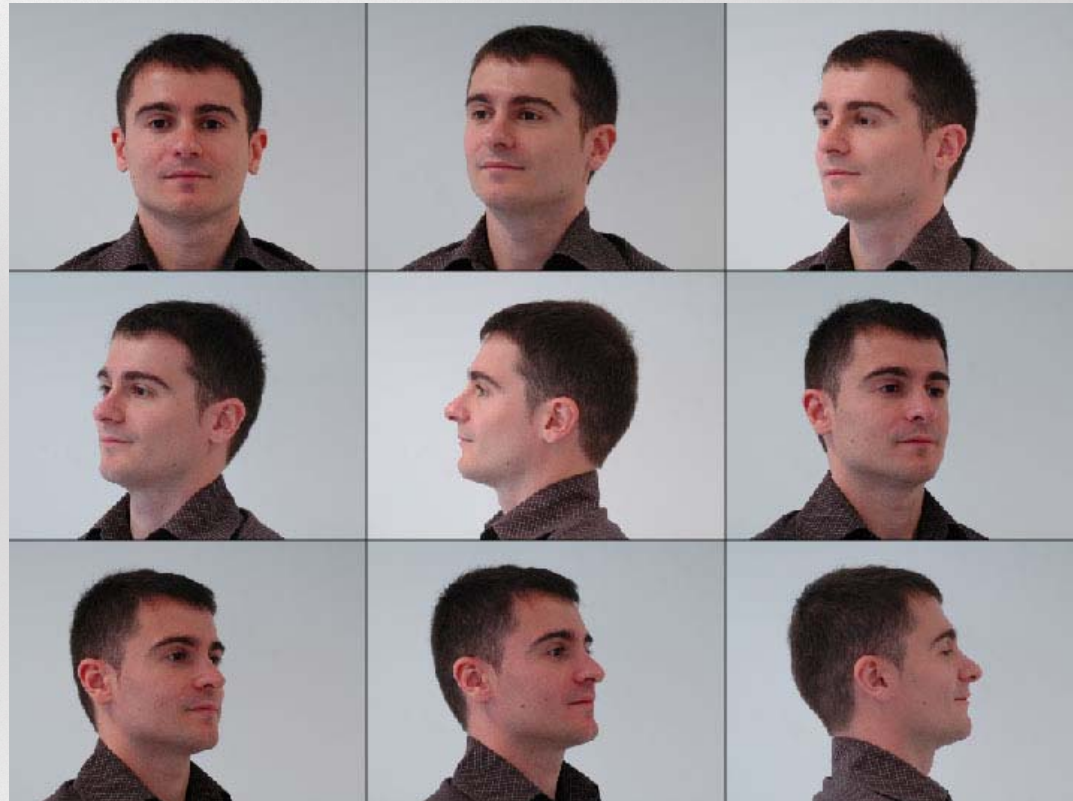
- Visual (Lip) Information can help





Camera View

- Problem
 - Position of the camera
 - What is the best position of the camera





Research Data



AVICAR Database

- Consist of 100 speakers
- Most are American-born English speakers





Driving Conditions



35 mph, Windows Up (35U)



35 mph, Windows Down (35D)



Car Stopped (IDL)

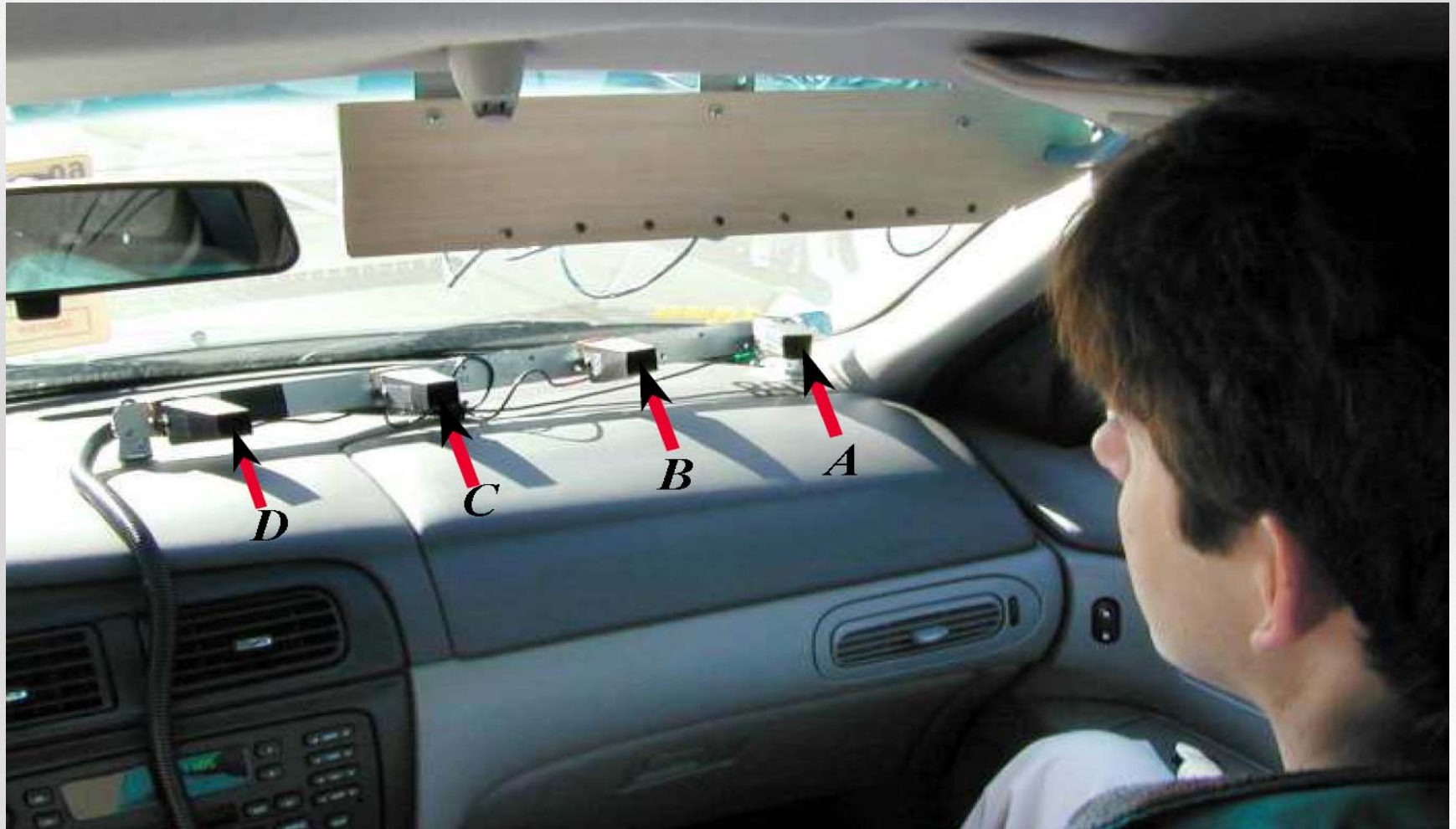




Research Methodology



Camera Setup



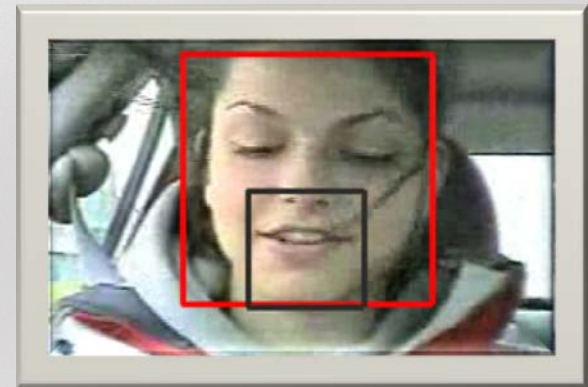


Visual Front End

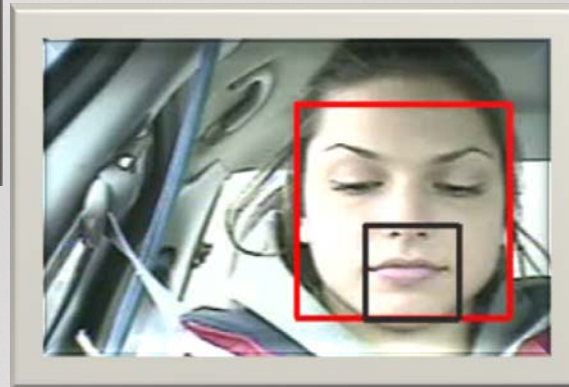
- Viola-Jones algorithm used to develop lip



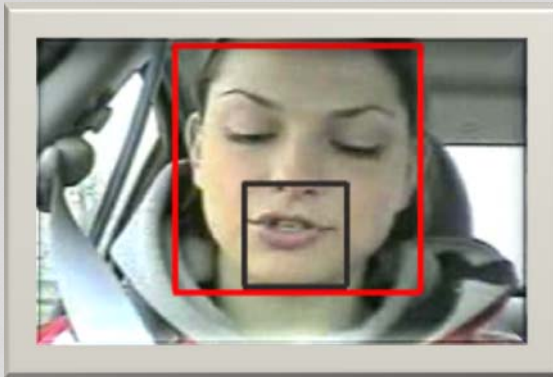
35U



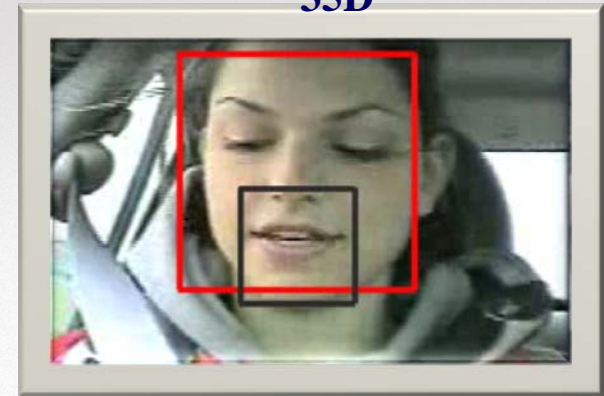
35D



IDL



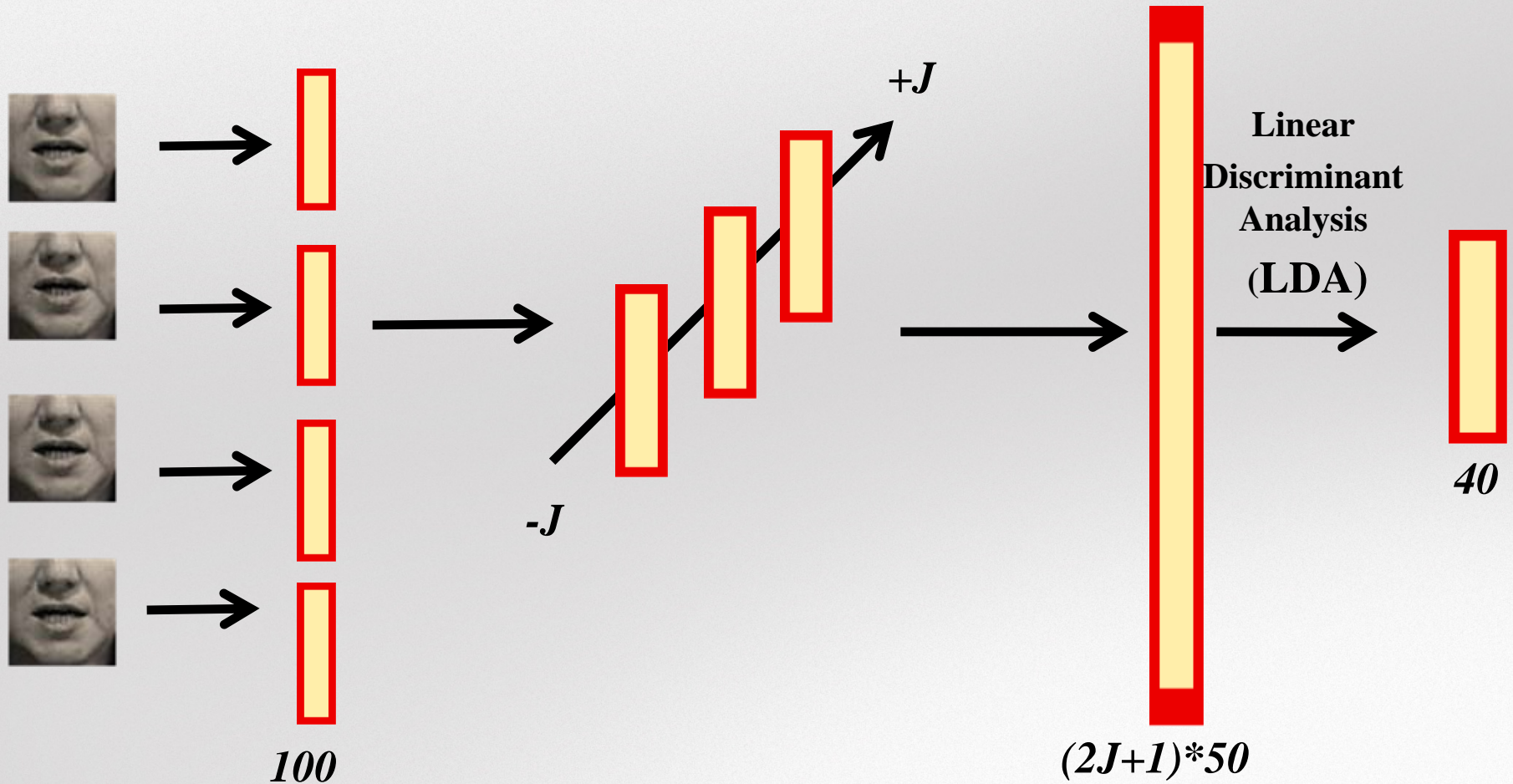
55U



55D

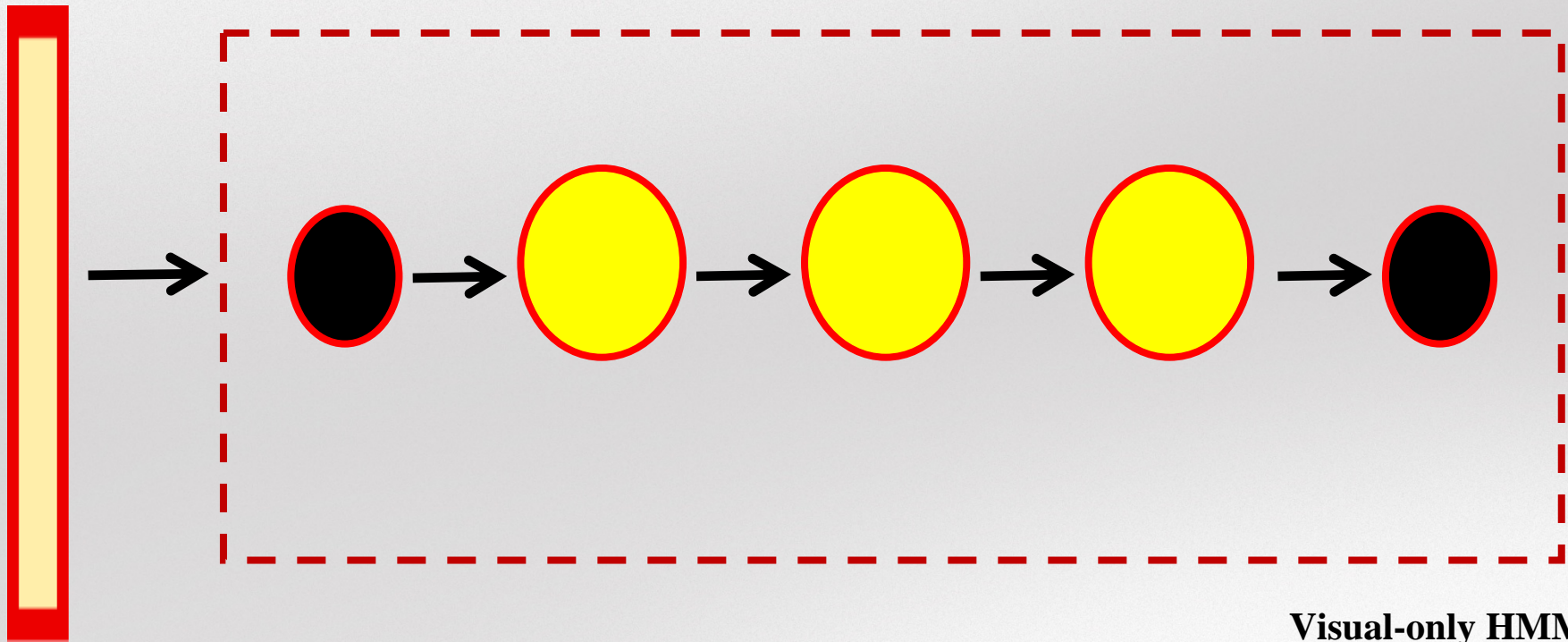


Visual Feature Extraction





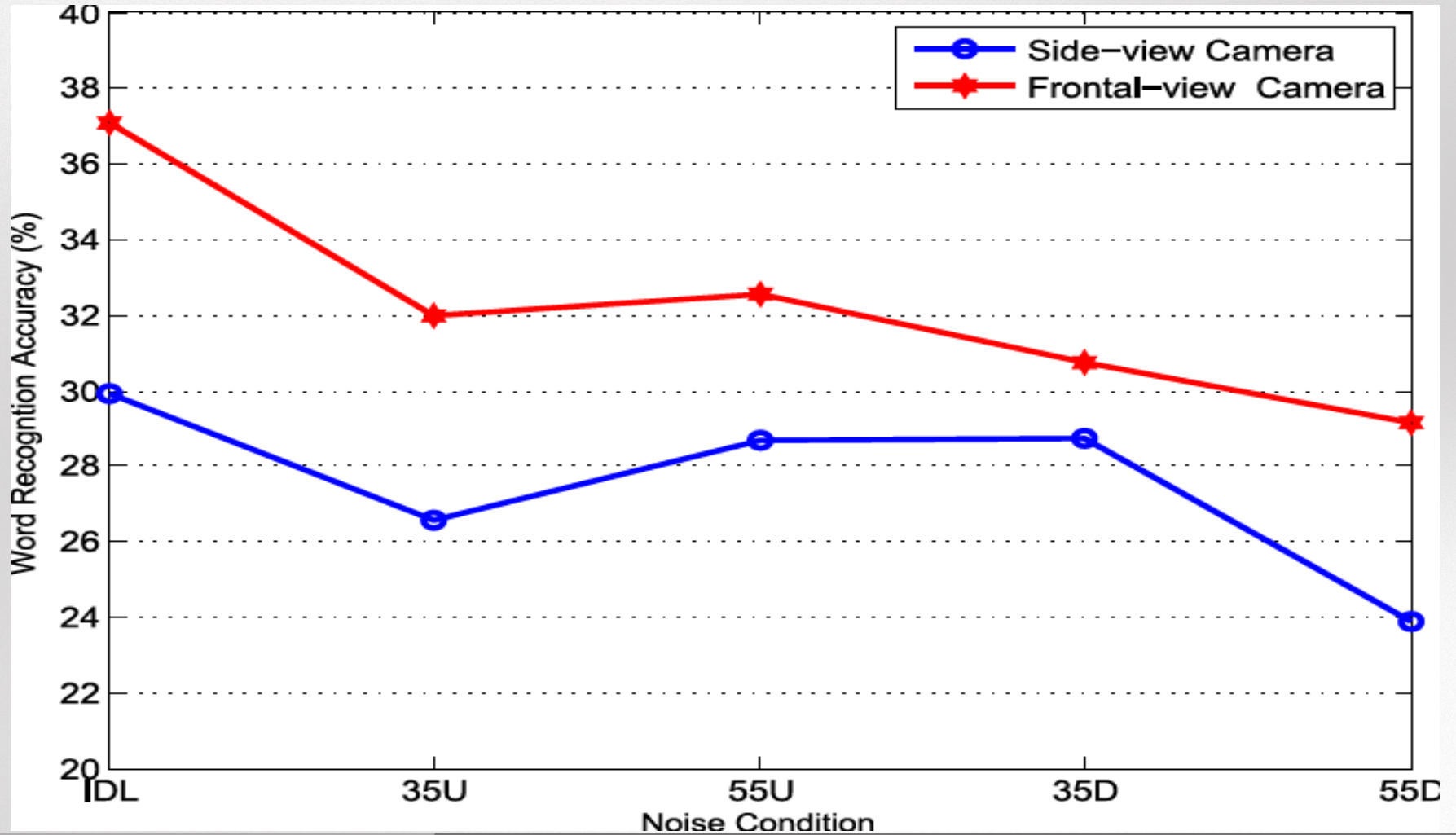
Visual-only Speech Recognition



**Visual-only HMM
Training/Testing**



Results





Conclusions



Summary/Conclusion

- We found that the visual-only speech recognition performance **varies** according to the
 - Driving conditions due to visual noise (i.e illumination, scale, etc ...)
 - Position of the camera



Summary/Conclusion

- Our experiments demonstrated that better visual-only speech recognition performance can be obtained from **frontal views**
 - It is **cost effective** to develop an efficient human car interaction system



Thank you

