Bahria University
Discovering Knowledge

FINAL YEAR PROJECT REPORT

REAL TIME SPEECH DRIVEN FACE ANIMATION SYSTEM

By

IZHAR US-SALAM KAYANI	**(36567)**
MUHAMMAD AHMER ASIM	**(36581)**
HASSAN SHAHAB ABBASSI	**(36563)**
RAKESH KUMAR BHIMANI	**(36598)**

SUPERVISED BY

MADAM ASIA SAMREEN

BAHRIA UNIVERSITY (KARACHI CAMPUS)

2017

ACKNOWLEDGEMENTS

We would like to thank everyone who had contributed to the successful completion of this project. We would like to express our gratitude to our research supervisor, Madam Asia Samreen for her invaluable advice, guidance and her enormous patience throughout the development of the research.

In addition, we would also like to express my gratitude to our loving parent and friends who had helped and given us encouragement.

REAL TIME SPEECH DRIVEN FACE ANIMATION SYSTEM

ABSTRACT

We have opted a paper "LIPNET: END-TO-END SENTENCE-LEVEL LIPREADING" as a base paper of our Final Year Project.

Lip-reading is the task of decryption text from the movement of a speaker's mouth. Ancient approaches separated the issue into 2 stages: planning or learning visual options, and prediction. Newer deep lip-reading approaches are end-to-end trainable (Wand et al., 2016; Chung & Zisserman, 2016a). However, existing work on models trained end-to-end perform solely word classification, instead of sentence-level sequence prediction. Studies have shown that human lip-reading performance will increase for extended words (Easton & Basala, 1982), indicating the importance of options capturing temporal context in an ambiguous communication. Intended by this observation, our project presents, a model that maps a video frames to text, creating use of spatial-temporal convolutions, a neural network, and therefore the connection temporal classification loss, trained entirely end-to-end. End-to-end sentence-level lip-reading model that at the same time learns spatial-temporal visual options and a sequence model.

TABLE OF CONTENTS