GRADUATION PROJECT PROPOSAL

PROJECT NAME	Human emotion recognition by optimally fusing facial expression and speech .	
PROJECT Supervisor	Dr:Eman Abd-El Ghafar	
Team Members	1. Mahmoud Mohamed Abd El Wahab	2. Amr Kamal Fouad
	3. Mohamed Ramadan Abo Ela	4. Marwa Adel Mustafa
	5. Mariam Maher Hassan	6.
	7.	8.
CONTACT NAME	Mahmoud Mohamed Abd El Wahab	
PHONE	+201099315074	
EMAIL	mahmoud170844@feng.bu.edu.eg	

SUMMARY	Emotion recognition is an important research field for Human-Computer Interaction. Emotion recognition has many applications. for example, instead of filling out a lengthy survey about how you feel at each point watching an educational video or advertisement, you can consent to have a camera watch your face and listen to what you say, and note during which parts of the experience you show expressions such as boredom, interest, confusion, or smiling. (Note that this does not imply it is reading your innermost feelings—it only reads what you express outwardly.) Other uses include helping children with autism, helping people who are blind to read facial expressions, helping robots interact more intelligently with people, and monitoring signs of attention while driving in an effort to enhance driver safety. Traditional speech emotion recognition algorithms cannot be effectively generalized since both training and testing data comes from the same domain, which have the same data distribution. In practice, speech data is acquired from different devices and in different recording environments. AudioVideo Emotion Recognition is now attacked with Deep Neural Network modeling tools. In papers, as a rule, the authors show only cases of the superiority in multi-modality over audio-only or video-only modality. However, there are cases superiority in uni-modality can be found. In our project . Our goal is to build multi model fusion emotion recognition using sound and video and achieve an acceptable accuracy.

PROBLEM STATEMENT	• Help computer decided to analyze human-emotion from sound and video • Emotion recognition can be used in different application like companies want to improve customers experience , helping blind people to read facial expressions and helping robots interact more intelligently with people and analyze their emotion .

EXPECTED CHALLENGES	- Unclear audio for speech signals. - Face occlusion and lighting issues. - Identifying facial features. - Fusion multi model video and audio.
PROPOSED SOLUTION	Our multi model emotion recognition from video and sound will work on video-audio dataset .we extract audio and video from dataset separately . In audio we have two approach the first one using signal processing and the second one using spectrogram . Then the output of one of the two approaches will be used as an input to the deep neural network to recognize emotion from sound. In video we will make face detection and will be used as an input to the deep neural network. then the outputs of this two models video and audio as an input to create multi fusion model to recognize emotions.
COMPETITORS/MAIN WORK DONE IN THE LITERATURE	in speech emotion recognition models ,speech Features extraction has two approach signal processing and spectrogram. then feed this Features into different deep-learning model like CNN-LSTM...etc.in video emotion recognition models used common techniques like CNN-RNN.then fusion this two models to detect emotion. papers achieved an acceptable accuracy using different datasets.
DESIRED OUTCOME	Enhance/improvement the model accuracy using various machine /deep learning techniques.
WHAT DO YOU NEED TO LEARN IF ANY	- Learning different machine /Deep learning techniques used in video and audio emotion recognition - Learning the best fusion technique for multi model video and audio.
- **EXPECTED TOOLS/TECHNOLOGIES NEEDED**	- RAVDESS dataset - Python - K-means,SVM machine learning algorithims - RNN ,CNN and LSTM Deep learning algorthims. - MFCC and Spectral centroid. - MFLS(mel spectrogram) - pytorch and tensorflow - google colab
PROJECT SCHEDULE/TIME PLAN	<table><tr><td>Date</td><td>Tasks</td></tr><tr><td>4 weeks</td><td>Search and Reading papers and about used techniques</td></tr><tr><td>3 weeks</td><td>Learning phase(machine learning)</td></tr><tr><td>1 weeks</td><td>Search and download different data sets</td></tr></table>

	3 weeks	Run different models with different data sets on speech plus learn deep learning
	2 weeks	Design emotion recognition speech model
	3 weeks	Run different models with different data sets on video plus learn deep learning
	2 weeks	Design emotion recognition video model
	1 weeks	Design fusion multi video and speech model
	5 weeks	Implementation phase
	1 weeks	Prepare the final report
• **PROJECTED TEAM AND RESOURCE REQUIREMENTS IF ANY**	• RAVDESS dataset • VidTIMIT dataset • BAUM-15 dataset	