**Graduation Project Proposal**

**Audio Mood Visualizer: Automated Emotional Avatar Generation from Audio**

## Team (3-6):

|  |  |  |
| --- | --- | --- |
| **Member name** | **Department** | **e-mail** |
| 1. امنيه احمد السيد مصطفى | CS | 20201700142@cis.asu.edu.eg |
| 1. امانى السيد محمد محمد | CS | 20201700139@cis.asu.edu.eg |
| 1. الاء صلاح عبد الهادي رجب احمد | CS | 20201700134@cis.asu.edu.eg |
| 1. نور الهدى احمد مصطفى كمال | CS | 20201700931@cis.asu.edu.eg |
| 1. نورهان عادل عبد الراضي زكي | CS | 20201700941@cis.asu.edu.eg |
| 1. مي عادل عبد العزيز سبع | CS | 20201701213@cis.asu.edu.eg |

## Supervision team approval

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Signature** | **Department** | **e-mail** |
| **Dr. Hanan Hindy** |  | CS | hanan.hindy@cis.asu.edu.eg |
| **TA. Yomna Ahmed** |  | CS | yomna.ahmed@cis.asu.edu.eg |

## Introduction

“When dealing with people, remember you are not dealing with creatures of logic, but creatures of emotion.” Dale Carnegie [2]. Emotions play an extremely important role in human mental life [3]. Nowadays, most people are constantly interacting with virtual voice assistants such as Siri and Alexa, some of these interfaces ignore the user's emotional state which leads to these assistants being perceived as cold, socially inept, untrustworthy and incompetent [4]. Speech Emotion Recognition (SER) can be defined as extraction of the emotional state of the speaker from his or her speech signal. Because of this, SER is becoming an increasingly relevant task [3]. Speech is an important means of transmitting a wide variety of information, which also includes emotional cues. SER has become the subject of various research efforts in recent years. SER is faced with several challenges including the fact that it is a relatively small field. There is also a lack of official standards in the definition of emotion. Different listeners may have different views on the feelings of the same voice. In addition, a segment of speech often has emotional changes and strong subjectivity, which leads to the lack of universality of much research works. SER can be used for different purposes in many areas like Education, Security, Communication and Health.

Facial expression is one of the most expressive ways for human beings to deliver their emotion, intention, and other nonverbal messages in face-to-face communications. In face-to-face interactions, people express themselves through several different modalities, not just speech [1]. Therefore, in this project, we aim to visualize the emotions of the speaking person through the generation of an appropriate avatar that reflects the detected emotions. Avatars are highly distinctive and easily recognizable, they may be used in a range of applications as mentioned above...

## Motivation

An application that detects emotions from voice and visualizes these emotions through an avatar can be used in many fields. It can be used in the medical field for therapy, where it can help people who have social phobias to communicate easily with others. In games, it can deliver a virtual reflection of a player’s emotions while protecting their privacy. Such an application can also be beneficial in customer service where it can aid the call center workers in quickly recognizing the emotions of the calling customers. It can also be useful in E-learning and virtual meetings as it can help the teachers to detect the state of the students throughout the sessions. In addition, it also supports privacy as you can use a photo that isn't yours or by using a default avatar.

## Objectives

Apply recent advances in speech emotion recognition research.

Visualize the recognized emotions (Happy, Sad, Fearful, Calm, Angry) through the generation of avatars with facial expressions that correspond to the detected emotion.

Create an application where the user could communicate with people and all his/her emotions are delivered to them through an avatar without needing to open the camera.

Experiment with additional features to further personalize the avatars while preserving the identity of the original speaker.

## Main Modules/Functionalities

1. Extract Emotions (happy, sad, angry, fearful, calm) from Audio
2. Create Avatar based on emotions.
3. Create customizable Avatar.

## Work Plan

Survey recent research works in the field.

Select the appropriate datasets and techniques to be used.

Design the system architecture.

Implement the speech emotion recognition module.

Implement the avatar creation module.

Apply model enhancement and evaluation.

Experiment with additional features.

Deployment.

|  |  |  |
| --- | --- | --- |
| **Project Activities** | **Start Date** | **End Date** |
| Survey & Learning | 1/10/2023 | 15/11/2023 |
| *…* Requirement Specifications  Project Analysis  Use cases  ER Diagrams  Sequence Diagram  . . .  Project Design  Project Architecture  User Interface Design  . . .  Project Testing  Modules Testing  Modules Integration  . . .  Team Work  Weekly Team Member/Tasks table  Requirement Specifications | 16/11/2023 | 21/11/2023 |
| Project Analysis  Use cases, ER Diagrams, Sequence Diagram… | 22/11/2023 | 26/11/2023 |
| Project Design  Project Architecture  User Interface Design | 27/11/2023  1/12/2023 | 30/11/2023  4/12/2023 |
| Project Implementation (according to the [specified modules](#_Main_Modules/Functionalities))  Module#1: Speech emotion recognition.  Module#2:Avatar Generation based on emotions  Module#3:Additional Features | 5/12/2023  6/1/2024  7/2/2024 | 5/1/2024  6/2/2024  28/2/2024 |
| Project Testing  Modules Testing  Modules Integration | 28/2/2024  14/3/2024 | 13/3/2024  29/3/2024 |
| Project Documentation | 9/10/2023 | 30/5/2024 |

## C:\Users\future\AppData\Local\Packages\5319275A.WhatsAppDesktop_cv1g1gvanyjgm\TempState\011FDA2B42A3E8279E739AE6A1CE5450\WhatsApp Image 2023-10-08 at 12.24.46_15c53c0b.jpg

## References

# [1] S. Zhang, Z. Wu, H. M. Meng, L. Cai *"Facial Expression Synthesis Based on Emotion Dimensions for Affective Talking Avatar"* Modeling Machine Emotions for Realizing Intelligence, Volume 1, T. Nishida, L. C. jain, C. Faucher, Eds. 2010, pp.109-132.

[2] T.Turrecha “Conveying Emotions Through Avatars” 2 march,2021. [online]. Available: <https://www.meldcx.com/blogs/conveying-emotions-through-avatars>

[3] M.Selvaraj, R.Bhuvana, S.Padmaja Karthik “Human speech recognition” International Journal of Engineering and Technology(IJET), February 2016.[online]. Available: <https://www.researchgate.net/publication/299185942_Human_speech_emotion_recognition>

[4] L.Pepino, P.Riera, L.Ferrer “Emotion from Speech Using wav2vec 2.0 Embeddings.” Proc. Interspeech, pp.3400-3404. August 2021.[online]. Available: <https://www.researchgate.net/publication/354220704_Emotion_Recognition_from_Speech_Using_wav2vec_20_Embeddings>

# [5] I. Attar, K. Kadole, G. Karanjekar, R. Nagarkar, Prof. S. Attar” *Speech Emotion Recognition Using Machine Learning”* International Journal of Research Publication and Reviews, May 2022. [online]. Available: <https://ijrpr.com/uploads/V3ISSUE5/IJRPR4210.pdf>

# [6] J. Singh, L. Saheer “*Speech Emotion Recognition Using Attention Model”* International Journal of Environmental Research and Public Health, 14 Mar 2023[online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10049636/>

# [7] M. Asif, R. Ullah, W. Ali, F. Anjam, I. Ullah, T. khurshaid, S. Shah,” *Speech Emotion Recognition Using Convolution Neural Networks and Multi-Head Convolutional Transformer* ” MDPI(Multidisciplinary Digital Publishing Institute), 7 July 2023.[online]. Available: <https://www.mdpi.com/1424-8220/23/13/6212>

[8] C. Moreno, A. Corrales” *The emotions effect on a virtual character’s design–A student perspective analysis”* Universidad Europea de Madrid, Spain, 10 August 2022.[online]. Available:

<https://www.frontiersin.org/articles/10.3389/fcomp.2022.892597/full>

[9] S. Tamanna, S. Sultana ” *A Review of the Advancement in Speech Emotion Recognition for Indo-Aryan and Dravidian Languages”* *Advances in Human-Computer Interaction, 1 Dec 2022*.[online]. Available: https://www.hindawi.com/journals/ahci/2022/9602429/?utm\_source=google&utm\_medium=cpc&utm\_campaign=HDW\_MRKT\_GBL\_SUB\_ADWO\_PAI\_DYNA\_JOUR\_X\_Partners\_IET&gclid=EAIaIQobChMIwe7CwuS9gAMVy5JoCR2CYQxpEAAYASAAEgIfJPD\_BwE

[10] R. Jarina, E. Lieskovska, M. Jakubec, M. chmulik ” *A Review on Speech Emotion Recognition Using Deep Learning and Attention Mechanism” MDPI(Multidisciplinary Digital Publishing Institute), 13 May* 2021. [online]. Available: <https://www.mdpi.com/2079-9292/10/10/1163>

[11]S. Chamishka, I. Madhavi, R. Nawaratne, D. Alahakoon “*A voice-based real-time emotion detection technique using recurrent neural network empowered feature modelling”* *Multimedia Tools Applications, 3 June 2022* . [online]. Available: <https://www.researchgate.net/publication/361478200_A_voice-based_real-time_emotion_detection_technique_using_recurrent_neural_network_empowered_feature_modelling>