IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

AI-POWERED VIRTUAL COMPANION FOR EMOTIONAL SUPPORT

Rajalakshmi R
Assistant Professor, Department of CSE Coimbatore Institute of Technology
Coimbatore-641014

.

Pooja R, Preethi S, Priyanga K, Roshini B A
B.E., Department of CSE, Coimbatore Institute of Technology
Coimbatore-641014

Abstract

This research investigates the development of an AI-powered companion designed to provide emotional support to individuals facing challenging circumstances. The AI system interprets and responds to emotional cues through a sophisticated workflow that converts user input into contextually relevant interactions. The process involves generating empathetic textual responses to user inputs,transforming these responses into natural- sounding speech, and synchronizing the digital human's lip movements with the audio to enhance interaction realism. This seamless integration of speech synthesis, and lip synchronization aims to create an immersive and supportive user experience. This study contributes to the field of AI- driven emotional support by demonstrating the potential of advanced technologies todeliver meaningful and human-likeinteractions, promoting mental well-being and offering a novel approach to digital emotional assistance.

KEYWORDS: *Emotional Support Virtual Assistant (ESVA), Artificial Intelligence in Emotional Well-being, Digital Human Avatar, AI-driven Companionship*

I. INTRODUCTION

The rapid advancements in technology have significantly transformed various aspects of daily life, including how we address emotional well-being. Increasing instances of emotional distress, loneliness, and mental health challenges necessitate innovative solutions for support and companionship. This project focuses on developing an Emotional Support Virtual Assistant (ESVA) powered by Artificial Intelligence (AI). The ESVA aims to act as a digital companion, capable of understanding, empathizing, and engaging with users through human-like interactions.

Leveraging advanced technologies, including response generation and animations, the ESVA offers contextually relevant and empathetic responses. By integrating various technologies, the ESVA delivers natural-sounding speech and synchronized digitalhuman lip movements, providing an immersive and supportive user experience. The primary goal is to create a personalized, empathetic, and evolving virtual assistant that can adapt to individual user preferences and emotional states, offering meaningful companionship and emotional support.

II. LITERATURE SURVEY

a) A Smart 3D Assistant on the Web

This study introduces the SAMIR system, a sophisticated framework for developing intelligent web agents that feature a 3D animated face to convey expressions, coupled with a customized version of the ALICE chatterbot for conversational interaction. The system includes an XCS classifier to maintain coherence between conversations and facial expressions.

b) A Greedy Pursuit Approach for Fitting 3D Facial Expression Models

This research presents a novel method for fitting expression blendshapes, using a selective approach that reduces redundancy and employs an expression correlation map. The method ensures natural and semantically expressive facial representations without the need for regularization. Experimentalvalidation on public datasets shows that this approach surpasses baseline methods and state- of-the-art facial fitting techniques in achievinghigh-quality facial representations.

c) Engaging with intelligent voice assistants for wellbeing and brand Attachment.

The paper explores the integration of emotion recognition and speech recognition into virtual assistants, enhancing their flexibility and responsiveness. By detecting emotions through speech, virtual assistants canperform tasks more effectively and reduce the need for typing. This advancement aims to improve daily interactions by making virtual assistants more intuitive and user-centric.

d) Enabling Intelligent Environment by the Design of Emotionally Aware Virtual Assistant: A Case of Smart Campus

This study develops an emotionallyaware campus virtual assistant using a Deep Neural Network (DNN) and Chinese Word Embedding for better dialogue tolerance and semantic interpretation. The system classifies emotions directly through a convolutionalneural network, bypassing traditional methods.Presented in an app format, it offers a simple voice response interface, enhancing user convenience and cost-effectiveness.

e) Engaging with Intelligent Voice Assistants for Wellbeing and Brand Attachment

Drawing on self-determination and motivation theories, this study examines how intrinsic motives like autonomy, competence, and relatedness drive engagement with AI- powered intelligent voice assistants (IVAs). The research shows that consumer engagement leads to increased wellbeing, which mediates the relationship between engagement and brand attachment. This study contributes to both positive psychology and branding research by linking mental health and branding effects.

f) A Lip Sync Expert is All You Need for Speech to Lip Generation in the Wild

This research addresses the challenge of lip-syncing talking face videos to match arbitrary speech segments, particularly focusing on dynamic and unconstrained videos.While existing methods achieve accurate lip movements on static images or specific individuals seen during training, they struggle with arbitrary identities in dynamic contexts, leading to noticeable desynchronization. The authors propose the Wav2Lip model, which uses a robust lip-sync discriminator to enhance lip synchronization accuracy.

III. PROPOSED METHODOLOGY

The methodology for designing and implementing the AI-powered digital companion system for emotional support

revolves around the integration of several key modules. These modules include the User Input Module, Text Processing Module, Audio Generation Module, Viseme Generation Module, Synchronization Module, and Digital Human Rendering Module. Each module playsa crucial role in ensuring a seamless and engaging user experience.

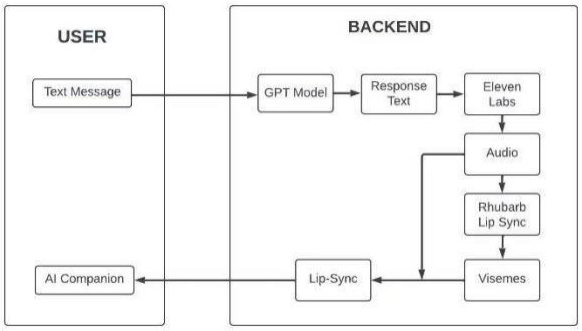

Fig1. System Architecture

Fig1 shows the workflow of the system. The process begins when the user inputs a text message, which is forwarded to the backend. The GPT model processes this input and generates a relevant and contextually appropriate response in text form. This response text is then sent to the Eleven Labs Text-to-Speech (TTS) API, which converts it into natural-sounding audio. The generated audio is further processed by Rhubarb Lip Syncto produce viseme data, which maps phonemes to corresponding mouth shapes for accurate lip synchronization. The synchronization module then combines the audio and viseme data to ensure that the digital human's lip movements are perfectly aligned with the spoken audio. Finally, the AI companion renders the synchronized speech and lip movements in a visually engaging manner, creating a lifelike interaction that enhances the emotional support experience for the user. This comprehensive workflow ensures that the digital companion can effectively understand user inputs, generate empathetic responses, and deliver them in a natural and realistic manner, providing users with timely and accessible emotional support. The functionality of each module is explained below.

The User Input Module serves as the entry point for user interaction, capturing inputin text form and formatting it for processing. Itutilizes technologies such as React for frontend development and Tailwind CSS for UI design, enabling a user-friendly interface for inputting queries or messages.

The Text Processing Module interpretsthe user's input, generating relevant responses using the GPT-3.5 Turbo API. Implemented with Express.js middleware, this module manages the flow of data and handles API calls efficiently, ensuring accurate understanding and response generation.

The Audio Generation Module convertsthe text responses into natural-sounding speechusing the Eleven Labs Text-to-Speech (TTS) API. Also, Node.js is employed to manage APIinteractions and facilitate data flow from text generation to speech synthesis, ensuring high- quality audio output.

The Viseme Generation Module analyzes the audio generated by the Audio Generation Module, producing viseme data that maps phonemes to corresponding mouth shapes for accurate lip synchronization. Implemented with Rhubarb Lip Sync and Python, this module enables realistic mouth movements aligned with spoken words.

The Synchronization Module aligns the audio output with the viseme data to ensure perfect lip movements of the digital avatar. A custom synchronization logic is employed, integrated within the frontend using React and

Three.js to guarantee smooth synchronization between audio and visuals.

The Digital Human Rendering Module handles the visual representation and animationof the digital avatar, incorporatingsynchronized lip movements and audio. Utilizing Three.js for 3D rendering and animation within a React application, this module brings the digital companion to life, enhancing the overall user experience with lifelike interaction.

IV. IMPLEMENTATION

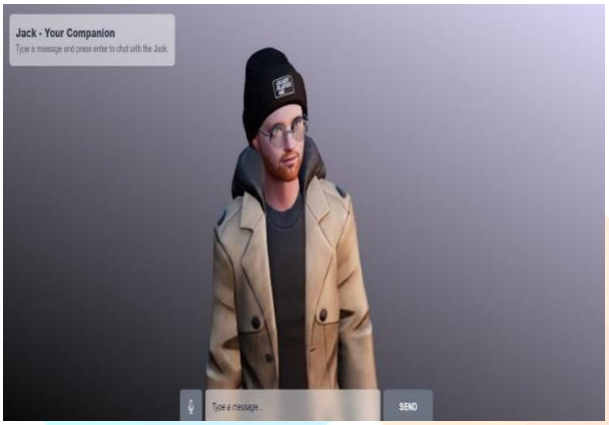

Fig2. User Interface

Fig.2 shows the user interface of the system. Here the users can give their own queries so that the model gives appropriate responses.

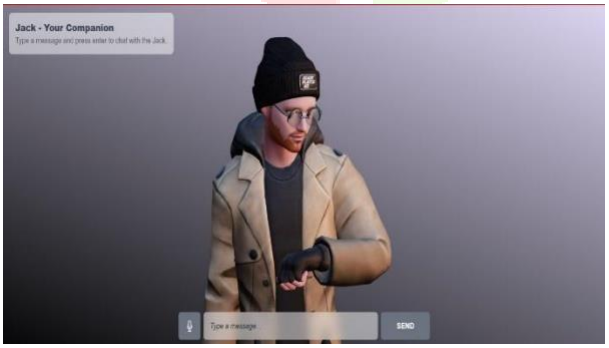

Fig3. User Interface-2

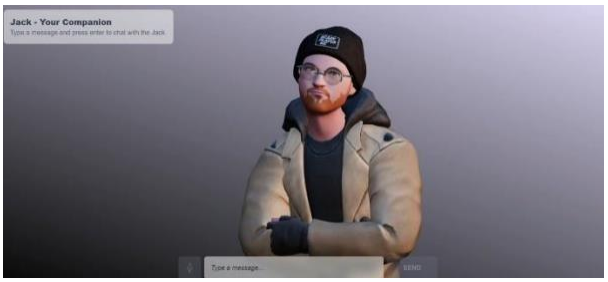

Fig4. User Interface-3

Fig3 and Fig 4 shows different animations of the avatar. Hence, we can see thehuman like interactions with the avatar enhancing the user experience.

V. CONCLUSION

In conclusion, our architecture providesa robust framework for an AI-powered emotional support system, seamlesslyintegrating response generation, text-to-speech conversion, and 3D avatar rendering. This system facilitates personalized and empathetic interactions with users, fostering a sense of companionship and understanding. By leveraging advanced technologies and ensuring cohesive module integration, our digital companion effectively delivers timely and realistic emotional support, enhancing theoverall user experience.

VI. FUTURE WORK

The future work on the system focuses on improving emotion recognition accuracy, refining response naturalness, and expanding 3D avatar capabilities for more immersive experiences. Exploring additional modalities such as facial expression recognition and gesture detection will enhance the system's ability to understand and respond to users' emotional cues, advancing the effectiveness and accessibility of AI-driven emotional support. Additionally, efforts will be made to achieve more precise lip synchronization and render the system in a mobile app, further increasing its accessibility and usability for users.

VII. REFERENCES

1.Prof.Y.D.Choudhari "Virtual Assistant with Emotional Recognition," Published on may-2021

2.PO-SHENG CHIU, JIA-WEI CHANG,MING-CHE LEE AND DA SHENG LEE-"Enabling Intelligent Environment by the Design of Emotionally Aware VirtualAssistant," published on April- 2020

3. Catherine Prentice, Sandra maria Correia Loureiro, Joao Guerrero ."Engaging with intelligent voice assistant for wellbeing and brand attachment," published on March- 2023

4. F. Abbattista, G. Catucci, G. Semeraro and F. Zambetti. "A Smart 3D Assistant on the Web," volume No.2 published on April 2014.

5. JIWOO KANG AND SANGHOON LEE , (Senior Member, IEEE)."A Greedy Pursuit Approach for Fitting 3D Facial Expression Models,". National Research Foundation of Korea(NRF) funded by the Korea Government in 2020.

6. Rud Rabha Mukhopadhyay, C V Jawahar, K R Prajwal, Vinay P. Namboodiri "A Lip Sync Expert Is All You Need for Speech to Lip Generation In The Wild "published on 23 August 2020.

7. C. Raghavendra; P. Nagarani; R. Suneetha Rani. "Speech Emotion Recognition: Node.jsand Web Speech API", published on march 2023.

8. Renuka S. Deshmukh Vandana Jagtap- "Software API and database for emotion recognition" ,by International Conference on Intelligent Computing and Control Systems (ICICCS).

8. Theodoros Giannakopoulos; Spiros Dimopoulos-"Using Oliver API for emotion-aware movie content characterization" By IEEE published on October 2019.

9. DinmukhamedMukashev-"Facial expression generation of 3D avatar based on semantic analysis" published on august 2021