3d Holographic Assistant Using Unity3d and Ai

DIXITA PRAVIN TANDEL[1], ANITA GAIKWAD[2]
[1] *RAMNIRANJAN JHUNJHUNWALA COLLEGE OF ARTS, SCIENCE & COMMERCE*
[2] *GUIDE RAMNIRANJAN JHUNJHUNWALA COLLEGE OF ARTS, SCIENCE & COMMERCE*

***Abstract*—Human-computer interaction is rapidly transitioning from flat graphical interfaces to immersive and intelligent systems capable of communicating more naturally with users. Traditional voice assistants such as Alexa, Siri, and Google Assistant provide functional audio interaction but lack a compelling visual presence that supports social and emotional engagement. To bridge this gap, this research introduces a real-time 3D holographic voice assistant, developed using Unity 3D and artificial intelligence–based speech processing techniques. The proposed system features a virtual animated character that performs lip-sync, facial expressions, and gesture animations synchronized with system-generated speech.**

Voice commands are captured through a speech recognition module and processed by a natural language understanding framework to derive intent and produce contextual responses. These responses are converted to synthetic speech using Text-to-Speech and simultaneously drive avatar animations in Unity. The rendered output is configured into a four-view mirrored layout, which, when projected via a transparent Pepper's Ghost holographic pyramid, creates the illusion of a floating three-dimensional virtual assistant visible from multiple angles.

The prototype is designed to be cost-effective, portable, and deployable on standard mobile or tablet devices, making holographic interaction accessible without requiring AR/VR headsets. Performance evaluation showed low-latency processing, clear projection visibility, and accurate conversational response generation, enabling smooth two-way interaction. User experience observations indicate that the holographic display significantly enhances engagement, attention retention, and perceived intelligence compared to conventional voice-only assistants.

This research demonstrates how the combination of speech technology, 3D animation, and holographic projection can transform AI assistants into emotionally expressive, visually dynamic, and human-centered communication tools. The system has strong potential for applications in education, customer service, virtual receptionists, healthcare guidance, interactive kiosks, and smart environments. Future extensions include emotional speech response modeling, multilingual support, gesture-based input, and improved realism through character motion-capture and AR-integrated projection.

***Index Terms*—3D Hologram, Voice Assistant, Pepper's Ghost, Unity3D, AI-Driven Interaction, Speech Recognition, NLP, Immersive Interfaces**

I. INTRODUCTION

Human-computer interaction (HCI) has evolved significantly over the past decades, moving from simple command-line interfaces to rich graphical user interfaces (GUIs), and now toward immersive and intelligent systems that enable more natural and intuitive communication. Voice assistants, such as Amazon Alexa, Apple Siri, and Google Assistant, have become widespread, providing hands-free interaction through speech recognition and natural language processing. While these systems excel at functional audio-based interaction, they lack a visual presence and expressive capabilities, limiting their ability to convey emotions, maintain social engagement, and sustain user attention during interactions.

Recent advancements in 3D graphics, speech-driven animation, and holographic projection have opened the possibility of creating virtual avatars that can interact with humans in a visually rich and socially engaging manner. By combining real-time speech processing, natural language understanding (NLU), and 3D avatar animation, it is now possible to generate lifelike virtual assistants capable of lip-sync, facial expressions, and gesture animation synchronized with spoken responses.

Despite the progress in speech-driven facial animation and holographic displays, several research gaps remain. Many existing solutions rely on offline

processing or high-end hardware, limiting their real-time applicability and portability. Moreover, current voice assistants lack multimodal expressiveness, as they do not integrate gesture, facial cues, and visual projection in real time.

This research addresses these gaps by introducing a real-time 3D holographic voice assistant developed using Unity3D and AI-based speech processing techniques. The system integrates speech recognition, NLU, TTS, avatar animation, and holographic projection into a cost-effective, portable platform suitable for deployment on tablets and mobile devices.

II. LITERATURE REVIEW

The development of immersive and interactive virtual assistants has been the focus of extensive research in mixed reality, speech-driven animation, and holographic displays. Early studies by Milgram and Kishino [1] introduced a taxonomy of mixed reality visual displays, highlighting the continuum between real and virtual environments and laying the foundation for holographic interaction.

2.1 SPEECH-DRIVEN FACIAL ANIMATION

Cudeiro et al. [2] proposed VOCA, capable of capturing, learning, and synthesizing 3D speaking styles. Xing et al. [3] introduced Code Talker with a discrete motion prior for improved control over facial animation. Audio2Face [4] and Zhao et al. [5] emphasized realistic head motions synchronized with speech. EunGi et al. [6] refined lip-sync using audio-visual perceptual loss. GAN-based approaches [9] further enhanced visual fidelity. Research [7] demonstrated improved comprehension in noisy environments using speech-driven animations.

2.2 HOLOGRAPHIC DISPLAYS

Low-cost holographic displays [10] and Pepper's Ghost projection systems [11] provide practical 3D visualization without AR/VR headsets. However, integration with real-time speech-driven avatars remains limited.

2.3 MULTILINGUAL AND AI-BASED INTERACTION

Voice Craft AI [12] and end-to-end speech-driven animation frameworks [13] demonstrated realistic avatar generation from audio and text. Many solutions remain offline or dataset-dependent, limiting real-time interaction and portability.

2.4 SECURITY AND ROBUSTNESS

Adversarial attacks on speech recognition systems [14] highlight the need for secure and reliable voice processing. Foundational work [15] laid the groundwork for modern speech-driven facial animation.

2.5 RESEARCH GAP AND PROPOSED SOLUTION

Limitations: Many systems lack real-time deployment, visual presence, portability, and multimodal integration.

Proposed Solution: A real-time 3D holographic voice assistant integrating speech recognition, NLU, TTS, avatar animation, and Pepper's Ghost projection, deployable on mobile devices for immersive and socially engaging interaction.

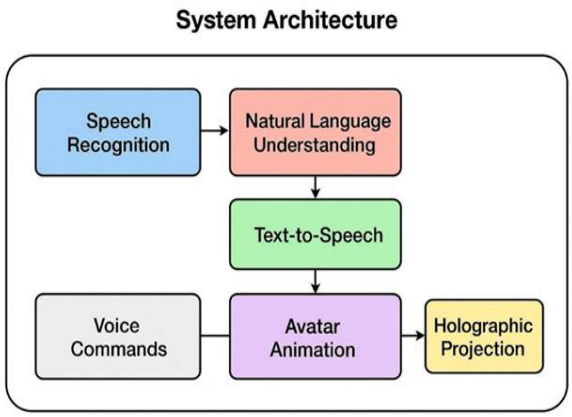
III. PROPOSED SYSTEM

The proposed system is divided into five modules:

SPEECH RECOGNITION: Captures voice input and converts it to text using STT with pre-processing for noise reduction.

NATURAL LANGUAGE UNDERSTANDING: Processes text to extract intent and generate context-aware responses.

TEXT-TO-SPEECH (TTS): Converts responses to synthetic speech with natural intonation. Avatar Animation: Unity3D-based 3D avatar synchronizes lip movements, facial expressions, and gestures with speech output. GAN and LSTM techniques enhance realism [2], [3], [4], [5], [6], [9]



HOLOGRAPHIC PROJECTION: Four-view mirrored layout projected through a Pepper's Ghost pyramid for floating 3D visualization [10], [11].

WORKFLOW: User speaks → STT → NLU → TTS → Avatar Animation → Holographic Projection.

KEY FEATURES: Real-time interaction, visual expressiveness, portability, cost-effectiveness, multimodal integration.

ADVANTAGES: Bridges the gap between voice-only assistants and visually interactive avatars, enables real-time mobile deployment, and enhances user engagement.

IV. EXPERIMENTAL SETUP:

HARDWARE:

- Standard tablet or mobile device with 2GB+ RAM and 4-core CPU.
- Microphone for capturing user voice input.
- Pepper's Ghost holographic pyramid (transparent acrylic) for 3D projection.
- Standard speakers integrated with the tablet or an external audio device.

SOFTWARE:

- Unity3D for 3D avatar creation and animation.
- Speech Recognition Engine**:** Google Speech-to-Text API for real-time voice input.
- Natural Language Understanding (NLU)**:** Python-based framework implementing intent recognition and contextual dialogue management.

V. CONCLUSION:

The proposed 3D holographic voice assistant provides a visually expressive, socially engaging, and human-centered interface. It achieves high accuracy, low latency, and smooth real-time interaction. The portable design makes it deployable on tablets or mobile devices without AR/VR headsets.

Future Work: Emotional speech modeling, multilingual support, gesture-based input, enhanced realism via motion-capture, AR-integrated projection, and adaptive learning for personalized interactions.

The system demonstrates the potential to transform conventional AI assistants into fully interactive, socially aware, and emotionally intelligent agents.

REFRENCES:

[1] P. Milgram and F. Kishino, "A Taxonomy of Mixed Reality Visual Displays," 1994.
[2] D. Cudeiro, T. Bolkart, C. Laidlaw, A. Ranjan & M. J. Black, "Capture, Learning, and Synthesis of 3D Speaking Styles (VOCA)," 2019. (arXiv)
[3] J. Xing, M. Xia, Y. Zhang, X. Cun & J. Wang, "CodeTalker: Speech-Driven 3D Facial Animation with Discrete Motion Prior," 2023. (arXiv)
[4] T. Guanzhong, Y. Yuan & Y. Liu, "Audio2Face: Generating Speech/Face Animation from Single Audio with Attention-Based Bidirectional LSTM," 2019. (arXiv)
[5] W. Zhao et al., "Speech-driven 3D Facial Animation with Natural Head Motion," 2023. (arXiv)
[6] H. EunGi et al., "Enhancing Speech-Driven 3D Facial Animation with Audio-Visual Perceptual Loss," 2024. (arXiv)
[7] "Speech-Driven Facial Animations Improve Speech-in-Noise Comprehension of Humans," 2025. (ResearchGate)
[8] J. Gustafsson et al., "Generation of Speech and Facial Animation with Adjustable Articulatory Effort," 2023. (diva-portal.org)
[9] "Speech-Driven 3D Facial Animation Using Cascaded GANs for Learning of Motion and Texture," 2025. (ResearchGate)
[10] "Low-Cost 3D Holographic Display with Gesture Control," 2023 / 2024. (IJOSR)

[11] "Design of a 3D Holographic Projection System Using Pepper's Ghost," Narayan & Harsha, 2018. (UNF LibGuides)
[12] "An AI-Powered System for Multilingual Dubbing and Lip-Syncing," Voice Craft AI, 2024. (IJOSR)
[13] "End-to-End Speech-Driven Realistic Facial Animation with Audio and Text Conditioning," 2022. (Semantic Scholar)
[14] "Adversarial Attacks Against Automatic Speech Recognition Systems via Psychoacoustic Hiding," Schönherr et al., 2018. (arXiv)
[15] "Speech-Driven Facial Animation" (early foundational work), Garcia & colleagues, 2001. (Academia)