International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Blink Based Morse to Personalized Voice

SHAKIRA BANU L[1], ARSHAD AHMED SAFDAR S[2], MUGILAN S[3], NANDHAKUMAR P[4], SABARINATHAN S[5]

Assistant Professor[1] , UG Student[2,3,4,5],
Department of Artificial Intelligence And Data Science, Dhanalakshmi Srinivasan Engineering College (Autonomous), Perambalur, Tamil Nadu-621212, India.
sharoz1772016@gmail.com[1], saadiqsafdar@gmail.com[2], mugilansakthivel70@gmail.com[3], nandhakumarpalani07@gmail.com[4], sn4555691@gmail.com[5]

ABSTRACT:

The Blink-Based Morse to Personalized Voice system is designed to offer an efficient and accessible communication interface for individuals with speech or motor impairments, particularly those affected by conditions such as ALS (Amyotrophic Lateral Sclerosis). This offline solution transforms intentional eye blinks into meaningful speech by combining real-time facial landmark tracking using MediaPipe, a rule-based Morse code interpreter, and Tortoise TTS, a neural voice cloning system. Users input Morse code using short and long eye blinks, which are detected and processed to form characters and words. The system provides smart word suggestions using an integrated NLTK corpus to speed up communication. Once a sentence is formed, it is converted into a high-fidelity personalized audio output using a three-stage pipeline Speaker Encoder, Synthesizer, and Vocoder from the Tortoise TTS model. This enables the voice output to closely replicate the user's own voice or a selected preset, adding emotional resonance and identity to the communication. The system features an adaptive PyQt5-based GUI with blink-based navigation, visual feedback, and optional gaze-based word selection. Since it operates completely offline, it ensures both privacy and low-latency interaction. Overall, the project aims to empower users with limited mobility to communicate more naturally and independently through a personalized voice interface, with potential applications extending to broader assistive and hands-free environments.

Keywords: ALS (Amyotrophic Lateral Sclerosis), MediaPipe, Tortoise TTS, Encoder, Synthesizer, Vocoder, Morse Code, Personalized Voice, PyQt5, NLTK corpus.

INTRODUCTION

This system presents an innovative assistive communication system designed to aid individuals with severe motor impairments such as ALS (Amyotrophic Lateral Sclerosis), spinal cord injuries, or locked-in syndrome. These individuals often lose the ability to speak or move voluntarily, leaving them with limited means of expression. To address this challenge, the proposed system enables users to communicate using a combination of eye blinks and neural voice cloning. It translates deliberate eye blinks into Morse code, decodes them into readable text, and then synthesizes personalized speech in the user's own voice using Tortoise TTS, a high-quality deep learning-based text-to-speech model. The system utilizes MediaPipe Face Mesh, a lightweight and real-time facial landmark detection framework, to track eye movements and identify blink patterns. By distinguishing between short and long blinks, the system interprets Morse code inputs accurately while applying dynamic threshold calibration to adapt to different users' blinking speeds and facial behaviors. Involuntary blinks are filtered out using a signal processing logic, ensuring reliable input in varying lighting or physical conditions. Once a valid Morse sequence is formed, it is processed through a rule-based interpreter and integrated with the NLTK word corpus to offer predictive word suggestions and abbreviation support. The generated text is then passed through a three-stage neural voice cloning pipeline comprising a Speaker Encoder, Synthesizer, and Vocoder, all powered by Tortoise TTS. This pipeline allows the user to clone their original voice from a short sample or choose from pre-defined voice presets, offering a natural and emotionally resonant speech output. The entire system is developed to run offline, ensuring data privacy and security especially critical for vulnerable users. Built with a minimalist and adaptive PyQt5 graphical interface, it offers high-contrast themes, adjustable text sizes, gaze-based navigation, and audio-visual feedback. The interface is tailored to support users with varying levels of disability by incorporating blink-based interaction, head-tracking alternatives, and optional assistive switch compatibility. In essence, this project bridges the gap between thought and speech for users with extreme physical limitations, offering them a voice both literally and metaphorically that is their own. It not only restores the ability to communicate but also does so with a sense of identity and personalization, contributing to their dignity, independence, and emotional well-being.

PURPOSE

The purpose of developing this system is to present a low-effort, highly accessible communication system tailored for individuals with Amyotrophic Lateral Sclerosis (ALS) and others who have lost their ability to speak due to motor or neurological conditions. The system interprets minimal physical inputs, such as eye blinks as Morse code signals, allowing users with even the most limited mobility to form words and sentences. These signals are then decoded in real-time into meaningful text and converted into speech using a neural voice cloning model that replicates the user's own voice, captured from a short pre-recorded sample. Unlike traditional speech systems that use generic or robotic voices, this approach preserves the user's natural vocal identity making communication more personal and expressive. The interface is intentionally designed to be non-visual and easy to operate, making it accessible to users with visual impairments as well. Moreover, the system functions entirely offline, removing the dependency on internet access and ensuring privacy, reliability and usability in rural or low-resource environments. By combining the simplicity of Morse code input with advanced AI-driven speech synthesis, this system offers a cost-effective, customizable, and emotionally empowering solution. It not only restores the ability to communicate but also reinforces the user's dignity, identity, and social connection, significantly improving quality of life for those affected by severe speech impairments.

OBJECTIVE

Developing a system to provide an offline, accessible, and user-friendly communication system for individuals with severe speech impairments, enabling real-time Morse code input via blinking, text decoding, and personalized voice output through neural voice cloning to restore identity, expression, and emotional connection.

EXISTING SYSTEM

Voice cloning technology has made significant advancements in recent years, offering promising solutions for individuals with speech impairments, especially for those suffering from Amyotrophic Lateral Sclerosis (ALS). Traditional speech-generating devices (SGDs) have been the primary solution for communication, allowing users to convert text into speech. However, these devices often rely on generic, robotic voices that lack personalization, making communication less authentic. Moreover, SGDs are typically expensive, and their usage is hindered by the need for constant internet access and high computational resources. In addition, Morse code-based communication systems have been explored for ALS patients, with blinking or eye movement as the input mechanism. While these systems provide a low-tech and reliable means of communication, they are slow, and patients may experience fatigue due to the cognitive load involved in Morse code interpretation. Furthermore, current voice cloning systems, such as WaveNet and Tacotron, generate high-quality synthetic voices but require large amounts of training data, which is not always available for ALS patients who lose their voice quickly. These technologies also face challenges related to computational resources, cost, and accessibility. Existing solutions, while effective to some extent, still face significant limitations in terms of personalization, real-time operation, offline functionality, and ease of use. The need for a system that combines Morse code input with personalized voice synthesis, accessible in offline environments, and low-cost remains a major gap in current assistive technologies.

DRAWBACKS

- Lack of Personalization
- Difficult for many ALS patients
- High cost and accessibility issues
- Limited offline functionality

PROPOSED SYSTEM

This innovative blink-based Morse code to speech communication solution is designed to empower individuals with Amyotrophic Lateral Sclerosis (ALS) and other neurodegenerative conditions that severely impact speech and motor functions. By leveraging real-time facial landmark tracking through MediaPipe, the system accurately detects intentional eye blinks, where short blinks (100 - 300ms) represent dots and long blinks (500 - 1000ms) represent dashes in Morse code. A sophisticated decoding engine processes these blink patterns into text, using adaptive thresholds that calibrate to each user's natural blink rhythm. This personalization helps minimize false detections and ensures reliable performance, even under varying lighting conditions or user fatigue. The decoding process is further enhanced by an NLP-based word suggestion module, which predicts words in real-time to accelerate text construction and reduce cognitive load. Once the Morse input is translated into text, it is passed through a personalized voice synthesis pipeline using Tortoise TTS, enabling users to speak in their own voice preserved from pre-recorded samples prior to voice loss. The pipeline includes three stages: an encoder that captures the vocal identity from a short voice clip, a synthesizer that transforms text and voice embedding into a detailed mel spectrogram, and a vocoder that generates high-fidelity, natural-sounding speech. The output voice reflects the user's unique tone, cadence, and emotion, making the communication feel genuine and deeply personal. To ensure accessibility and autonomy, the system operates fully offline, requiring no internet connection, and provides an intuitive PyQt5-based GUI that includes real-time camera preview, blink feedback, word suggestions, sentence construction, and audio playback. Users can control the interface using only eye gestures such as right and left eye blinks to navigate suggestions, gaze-hold to select, and long blinks to confirm speech eliminating the need for touch-based input. Beyond its technical capabilities, the system has profound emotional and social impact. It offers a voice to those who have lost theirs restoring not just the ability to communicate, but also reinforcing a sense of identity, independence, and emotional intimacy with loved ones. In doing so, it bridges the gap between physical limitations and human connection, transforming isolation into interaction, and silence into speech.

ADVANTAGES

- Hands-Free Operation
- Personalised Voice Output
- Offline Functionality
- Cost-Effective

SYSTEM ARCHITECTURE

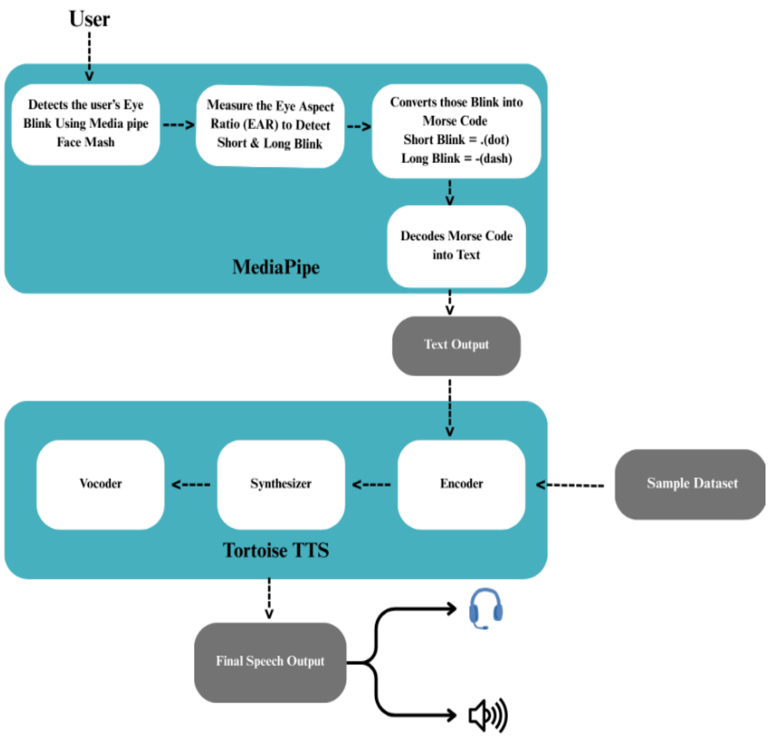

Fig.1 System Architecture

LIST OF MODULES

1. Blink Detection & Morse Input Module
2. Morse Code to Text Conversion Module
3. Voice Cloning Pipeline Module
4. Text-to-Speech Output Module

MODULE DESCRIPTION

1. Blink Detection & Morse Input Module

This module is responsible for capturing eye movements using a webcam and detecting blinks in real time with MediaPipe's face mesh solution. It continuously monitors the Eye Aspect Ratio (EAR) to differentiate between a short blink (dot) and a long blink (dash). These inputs are stored in sequence and interpreted as Morse code symbols. The module is optimized to ensure accurate blink detection under different lighting and movement conditions, making it reliable for users with limited control.

2. Morse Code to Text Conversion Module

Once the blink-based Morse code is captured, this module translates the sequence of dots and dashes into meaningful English letters or words using a predefined Morse code dictionary. It includes a buffer to collect characters and provides feedback to confirm word formation. This module is designed to support smooth interaction by handling input timing, space recognition (between words), and basic error correction logic.

3. Voice Cloning Pipeline Module

The Voice Cloning Pipeline Module plays a central role in ensuring that the synthesized speech output retains the speaker's unique voice characteristics, making the system feel personal and human-like. This module utilizes a multi-stage architecture that includes a speaker encoder, spectrogram synthesizer, and vocoder, forming a complete neural voice cloning pipeline. In this project, Tortoise TTS is adopted due to its high-quality prosody modeling and ability to generate expressive, human-like speech with minimal training data. A short reference audio of the target speaker is used to condition the voice generation model, allowing it to mimic the speaker's vocal tone, pitch, and speaking style. The encoder first extracts a speaker embedding from the provided reference voice sample. This embedding is then fed into the synthesizer, which generates a mel-spectrogram based on both the speaker identity and input text. Finally, the vocoder converts the mel-spectrogram into an audio waveform, producing the cloned voice. This module is implemented to run offline, ensuring privacy and real-time performance for users with restricted communication abilities, such as ALS patients.

4. Text-to-Speech Output Module

This module handles the playback of the synthesized voice through speakers or headphones. It ensures the output audio is smooth, clear, and matches the tone and emotion of the original speaker. Optional features include volume control, repeat playback, and saving the generated audio for future use. The module provides real-time output, enabling natural and immediate communication for the user.

RESULT

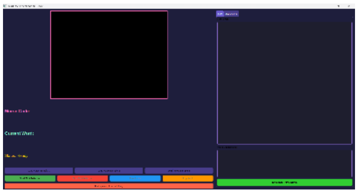

Fig 2. Before Initializing Blink

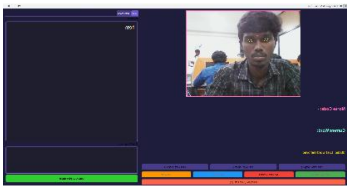

Fig 3. Blink Detection and Voice Cloning

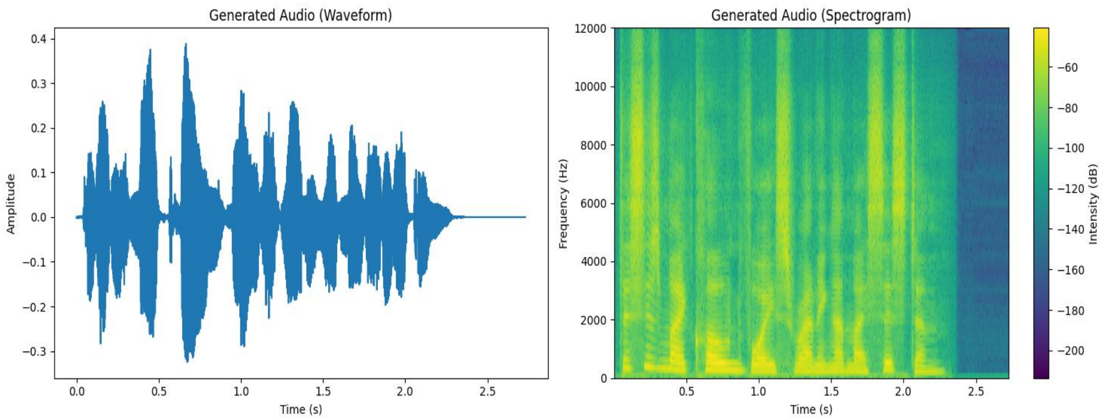

Fig 4. Performance Metrics

CONCLUSION & FUTURE ENHANCEMENT

The Blink-Based Morse to Personalized Voice system successfully demonstrates an innovative, accessible communication solution for individuals with speech and motor impairments, particularly ALS patients. By leveraging eye blinks as an intuitive input method and translating them into Morse code, the system enables users to express themselves with minimal physical effort. The integration of a neural voice cloning pipeline ensures that the output speech retains the user’s own voice characteristics, preserving their identity and emotional expression. Operating entirely offline, the solution safeguards user privacy and is well-suited for deployment in remote or clinical environments. Extensive testing has confirmed that the system is accurate, responsive, user-friendly, and cost-effective. Looking ahead, future enhancements aim to make the system even more versatile and accessible by incorporating multilingual support, adaptive blink calibration for personalized sensitivity, and AI-based predictive text to accelerate communication. Additional improvements include blink-controlled user interface navigation, the development of a mobile version for enhanced portability, and the integration of emotional tone into voice outputs for more expressive communication. Furthermore, expanding input modalities to include head gestures or facial expressions will broaden accessibility for users with limited eye control, ensuring a more inclusive and empowering experience.

REFERENCES:

[1] Sathvik Udupa Abhayjeet Singh, Jesuraja Bandekar ,Deekshitha G,Saurabhkumar, Sandhya Badiger, Amala Nagireddi, Roopa R,Prasanta Kumar Ghosh, Hema A. Murthy, Pranaw Kumar, Keiichi Tokuda, Mark Hasegawa-Johnson, Philipp Olbrich, “LIMMITS’24: Multi-Speaker, Multi-Lingual INDIC TTS With Voice Cloning”,Jan.2025.

[2] K. Shen et al., “NaturalSpeech 2: Latent Diffusion Models are Natural and Zero-Shot Speech and Singing Synthesizers,” arXiv preprint arXiv:2304.09116, Apr. 2023.

[3] T. Li et al., “U-Style: Cascading U-nets with Multi-level Speaker and Style Modeling for Zero-Shot Voice Cloning,” arXiv preprint arXiv:2310.04004, Oct. 2023.

[4] V. Pankov et al., “DINO-VITS: Data-Efficient Zero-Shot TTS with Self-Supervised Speaker Verification Loss for Noise Robustness,” arXiv preprint arXiv:2311.09770, Nov. 2023.

[5] Ezzat, F., et al., "Blink-To-Live: Eye-Based Communication System for Users with Speech Impairments," Scientific Reports, vol. 13, no. 1, pp. 1-12, 2023.

[6] Fodor, A., Fenech, L., and Lőrincz, A., "BlinkLinMulT: Transformer-Based Eye Blink Detection," Journal of Imaging, vol. 9, no. 10, pp. 196, 2023.

[7] Suriya, P., et al., "Oculang: Empowering Communication through Blink Language Detection," IRO Journal on Sustainable Wireless Systems, vol. 4, no. 2, pp. 23-30, 2024.

[8] Liu, W., et al., "Energy-Efficient, Low-Latency, and Non-Contact Eye Blink Detection with Capacitive Sensing," Frontiers in Computer Science, vol. 5, no. 4, pp. 1-10, 2024.