AntarVani – A Real-Time Neural Speech Decoder Using EEG and AI-driven Text-to-Speech Generation

Harsh Shirke
Email: shirkeharsh6@gmail.com

***Abstract*—This work presents AntarVani, a real-time neural speech decoding prototype designed to convert EEG signals into text and subsequently into natural speech. The system combines EEG feature processing, machine learning–based sentence prediction, and cloud-based text-to-speech synthesis to generate audio output in near real time. AntarVani aims to assist individuals with severe speech and communication impairments by enabling hands-free, voice-driven interaction using neural activity alone. The proposed pipeline demonstrates a functional, low-cost, and scalable brain-to-speech interface suitable for future assistive communication technologies.**

***Index Terms*—EEG, Brain-Computer Interface, Neural Speech Decoding, Text-to-Speech, Assistive Technology, Real-time Processing**

1. Introduction

Communication is a fundamental human requirement, yet millions of people affected by paralysis, ALS, stroke, or locked-in syndrome struggle to express basic needs. Assistive technologies such as eye-tracking keyboards or physical switches often demand effort, precision, or mobility that patients may not possess.

This project addresses the need for a frictionless communication channel by decoding a user's intent directly from brain activity. AntarVani is a prototype system that receives EEG signals, classifies the user's intended sentence using a trained neural model, and converts that output into speech using a cloud-based natural voice engine.

The motivation behind this work is to demonstrate a functional, real-time mind-to-speech pipeline that can operate on consumer-grade hardware and standard deep learning tools, while maintaining a responsive and natural interaction flow.

2. System Overview

AntarVani consists of three modules:

2.1. EEG Signal Acquisition & Feature Extraction

EEG data is captured from the user through a simulated or real EEG source. Raw wave values and preprocessing features such as frequency-domain attributes are provided to the backend model.

2.2. Machine Learning–Based Sentence Prediction

A trained Keras/TensorFlow model processes EEG features. A LabelEncoder maps model outputs to predefined sentence classes. The backend returns:

- Predicted sentence
- Model confidence
- Waveform samples

2.3. Text-to-Speech Using Murf Falcon API

Each new predicted sentence triggers a TTS request to Murf Falcon. The generated audio URL is saved by the backend. The frontend automatically fetches and plays the latest audio whenever a new sentence appears.

Together, these modules form a continuous pipeline:

EEG → Prediction → Text → Speech → Audio Playback

3. Methodology

3.1. EEG Data Processing

EEG samples and features are provided to the model through the `/predict` API endpoint. The backend performs:

- Input validation
- Feature normalization
- Model inference
- Confidence score extraction
- Mapping of numeric labels to sentence text

The system supports real-time polling through a dedicated `/latest` endpoint.

3.2. Neural Sentence Classification Model

A TensorFlow/Keras model was trained on sentence-level EEG feature datasets. **Key characteristics:**

- Input: numerical EEG feature vector and waveform metadata
- Output: class probabilities across predefined sentence categories
- Uses softmax activation for classification

- Model stored in .h5 format for lightweight deployment

Joblib-based LabelEncoder ensures correct mapping between class indices and natural-language sentences.

3.3. Text-to-Speech via Murf Falcon API

For every new predicted sentence:

1) The backend calls the Murf TTS API with:
   - Voice ID (e.g., en-US-natalie)
   - Text content of the predicted sentence
   - Audio format parameters
2) Murf returns a public audio URL.
3) This URL is made available at the `/audio` endpoint.
4) The React frontend auto-plays the audio through a persistent `<audio>` element.

This enables hands-free, real-time speech output without user interaction.

3.4. Frontend Interface

A React-based dashboard visualizes and interacts with the backend:

- Live EEG waveform graph
- Latest decoded sentence
- Confidence score indicator
- Command history panel
- Auto-playing synthesized speech
- Polls backend every second for updates

The interface is designed to emulate a professional ”neural decoding” display suitable for demonstrations and research presentations.

4. Results

The implemented prototype successfully demonstrates:

4.1. Real-time EEG-to-text decoding

- Continuous processing loop
- Smooth sentence updates
- Stable model inference pipeline

4.2. Immediate text-to-speech output

- Less than one second latency from new sentence detection to audio playback
- Natural, human-like voice using cloud synthesis

4.3. Responsive and intuitive UI

- Clear visualization
- Real-time waveform display
- Seamless audio feedback

The system runs on standard consumer hardware and does not require specialized neural decoding equipment for demonstration purposes.

5. Applications

AntarVani demonstrates the foundation for future assistive systems including:

- Communication support for non-verbal patients
- Voice restoration interfaces
- Brain-controlled assistive devices
- Silent-speech communication tools
- Thought-driven automation systems

Although this prototype is limited to predefined sentence categories, the pipeline architecture supports extension to larger vocabularies and deeper neural models.

6. Future Enhancements

With further research and funding, AntarVani can evolve significantly:

6.1. High-resolution neural acquisition

- Integration with medical-grade EEG or invasive electrodes
- Increased accuracy and vocabulary size

6.2. End-to-end deep neural decoding

- Direct imagined speech processing
- Transformer-based neural translation of brain activity

6.3. Wearable hardware

- Lightweight EEG headset
- On-device inference (edge AI)

6.4. Continuous speech generation

- Beyond sentence classification
- Real-time phoneme or word-level decoding

6.5. Adaptive personalization

- Individual calibration profiles
- Reinforcement learning for user-specific patterns

These advancements could transform the prototype into a fully deployable communication device.

7. Conclusion

AntarVani demonstrates a functional, real-time implementation of a neural speech decoder pipeline using EEG features, machine learning classification, and modern text-to-speech synthesis. The system showcases how multimodal AI can assist individuals with severe communication impairments by converting neural signals directly into spoken output. Although still a prototype, the architecture is scalable and provides a viable foundation for future brain-to-speech assistive technologies.

References

[1] Murf AI TTS Documentation – https://murf.ai/api

[2] TensorFlow Documentation – https://www.tensorflow.org

[3] EEG Signal Processing Basics – Standard academic resources

[4] FastAPI Documentation – https://fastapi.tiangolo.com

[5] Scikit-learn LabelEncoder Documentation