VOICE CONTROLLED AI USING NVIDIA JETSON

BY PRAJWAL MUTALIK, SANDESH HR, MEGHANA M,

MD ATHEEQ AHMED, C SYED ARSHIYA ANJUM

Regd.No.: BU22CSEN0600200, BU22CSEN0600188, BU22CSEN0600187, BU22CSEN0600198, BU22CSEN0600175

SOFTWARE DEFINED VEHICLES

INTN2333

(Duration: 01 Feb , 2024 to 01 Apr, 2024)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Gandhi Institute of Technology and Management

(DEEMED TO BE A UNIVERSITY)

BENGALURU, KARNATAKA, INDIA

SESSION:2024-2025

CERTIFICATE

This is to certify that the mini project entitled “Voice Controlled AI using NVIDIA Jetson” is bonafide record of work carried out by the team submitted in partial fulfillment of Bachelor of Technology in Computer Science and Engineering

---

Dr. S.Sowmya
Assistant professor
Department of CSE,
GST, Bengaluru

ACKNOWLEDGEMENT

The satisfaction and euphoria that accompany the successful completion of any task would be incomplete without the mention of the people who made it possible, whose consistent guidance and encouragement crowned our efforts with success.

We consider it our privilege to express our gratitude to all those who guided us in the completion of the project.

We express our gratitude to Director Prof. Basavaraj Gundappa Katageri for having provided us with the golden opportunity to undertake this project work in their esteemed organization.

We sincerely thank Dr. Y. Vamshidhar, HOD, Department of Computer Science and Engineering, Gandhi Institute of Technology and Management, Bengaluru for the immense support given to us.

We express our gratitude to our project guide (Dr. Sowmya S), (Assistant professor), Department of Computer Science and Engineering, Gandhi Institute of Technology and Management, Bengaluru, for their support, guidance, and suggestions throughout the project work.

Contents

TITLE **PAGE NO.**

ABSTRACT

With the rapid advancement of AI-driven technologies, voice control is becoming a critical interface for hands free interaction, especially in Software-Defined Vehicles (SDVs) and smart automation systems. This project focuses on designing an efficient and real-time speech recognition system optimized for embedded platforms, leveraging NVIDIA Jetson AGX Orin for high-performance AI inference

The system integrates Porcupine for wake word detection and Vosk for speech-to-text conversion, ensuring low latency processing and high command recognition accuracy. Additionally, GPIO and CAN interfaces facilitate seamless hardware interaction, enabling integration with automotive and IoT-based applications. By utilizing edge computing, the system enhances response time, security, and offline functionality, making it ideal for real-world deployment.

One of the primary challenges in voice-based control systems is maintaining processing efficiency without compromising recognition accuracy. To address this, the Jetson AGX Orin optimizes neural network inference, ensuring robust command execution even in noisy environments. The system also employs advanced noise filtering techniques, enabling precise speech recognition in dynamic and multi-user scenarios.

Beyond SDVs, this technology has broader applications in industrial automation, robotics, and smart city infrastructure. By integrating AI-driven decision-making with voice control, the system enables intelligent, context-aware interactions. This research demonstrates the feasibility of deploying scalable, adaptive, and AIpowered voice interfaces on embedded platforms, paving the way for next-generation human-machine interaction.

INTRODUCTION

1.1 Introduction to NVIDIA Jetson AGX Orin for AI Applications

NVIDIA Jetson AGX Orin is a cutting-edge AI computing platform designed for **edge computing applications**, offering a balance of **high performance, power efficiency, and real-time AI processing**. It integrates a **powerful GPU, AI accelerators, and advanced processing cores**, making it ideal for applications that require low-latency decision-making, such as **autonomous systems, robotics, and intelligent surveillance**.

Unlike cloud-based AI solutions, which depend on constant network connectivity, **Jetson AGX Orin enables on-device AI inference**, ensuring **faster response times, enhanced security, and reliability**. This capability is particularly useful in real-time AI applications like **voice recognition**, where efficient speech-to-text conversion and wake-word detection are essential.

With support for industry-leading AI frameworks such as **TensorFlow, PyTorch, ONNX, and TensorRT**, developers can build and deploy advanced deep learning models optimized for various industries. Its compatibility with **CUDA and DeepStream SDKs** allows for efficient GPU utilization, enabling high-performance computing in **embedded systems, self-driving vehicles, industrial automation, and smart edge devices**.

Additionally, Jetson AGX Orin's **power efficiency** makes it an excellent choice for **battery-operated systems, drones, and portable AI solutions**, ensuring long-term sustainability in resource-constrained environments. Its **rugged design and advanced thermal management** further enhance its usability in harsh industrial conditions, making it a versatile platform for modern AI-driven applications.

1.2 OVERVIEW OF NVIDIA JETSON FOR AI APPLICATIONS NVIDIA

Jetson AGX Orin is a high performance AI computing platform designed for edge computing applications. It integrates an advanced GPU, AI accelerators, and power-efficient computing capabilities. Jetson is widely used in robotics, autonomous vehicles, and smart devices. Its ability to handle real time deep learning workloads makes it an ideal choice for voice recognition applications.

NVIDIA Jetson AGX Orin is a high-performance AI computing platform designed specifically for edge computing applications, where low latency, high computational power, and energy efficiency are crucial. It integrates a powerful GPU, AI accelerators, and advanced processing cores, making it well-suited for demanding AI workloads, including computer vision, robotics, autonomous vehicles, and smart devices.

One of the key advantages of Jetson AGX Orin is its ability to handle real-time deep learning inference. Unlike cloud-based AI solutions that require constant network connectivity, Jetson AGX Orin enables on-device AI processing, reducing dependency on external servers and ensuring faster response times, enhanced security, and increased reliability. This makes it an excellent choice for applications where low-latency decision-making is essential, such as self-driving cars, industrial automation, and intelligent surveillance systems.

For voice recognition applications, Jetson AGX Orin stands out due to its optimized AI hardware and support for advanced speech processing frameworks. By leveraging dedicated AI acceleration cores, the platform can efficiently run speech-to-text models, natural language processing (NLP) algorithms, and wake word detection systems. This capability is particularly beneficial in Software-Defined Vehicles (SDVs), where voice commands are increasingly being used for hands-free control of vehicle functions.

Another critical feature of the Jetson AGX Orin platform is its scalability and flexibility. Developers can utilize NVIDIA's CUDA, TensorRT, and DeepStream SDKs to build highly efficient AI models tailored to specific applications. Its compatibility with frameworks like

TensorFlow, PyTorch, and ONNX further enhances its adaptability, allowing developers to deploy custom AI solutions across various industries.

Additionally, the platform's power efficiency makes it suitable for embedded and battery-operated systems, enabling AI-driven automation in drones, medical devices, and edge-based monitoring solutions. With its rugged design and advanced thermal management, Jetson AGX Orin can operate in harsh environments, making it ideal for industrial AI applications

SPEECH RECOGNITION & WAKE WORD DETECTION

Speech recognition involves converting spoken language into text using AI models. Wake word detection is the process of identifying a predefined word or phrase to activate the system. This project uses Porcupine for wake word detection, which is optimized for embedded devices with minimal latency and resource consumption.

Speech recognition is a fundamental technology in voice-controlled systems, enabling machines to interpret and respond to spoken commands. It involves converting spoken language into text using artificial intelligence (AI) models, which analyze audio input, extract linguistic features, and process them using deep learning techniques. Modern speech recognition systems rely on neural networks and natural language processing (NLP) models to achieve high accuracy in understanding human speech, even in noisy environments.

One of the key components of voice recognition systems is wake word detection, which identifies a specific predefined word or phrase to activate the system. This ensures that the device remains in a low-power standby mode until it detects the wake word, preventing unnecessary processing and enhancing energy efficiency. Wake word detection is crucial for real-time voice interfaces, as it enables hands-free operation, making the system responsive while avoiding accidental activations.

This project utilizes Porcupine, an advanced wake word detection engine optimized for embedded systems. Porcupine is designed to operate with minimal latency and low resource consumption, making it ideal for edge computing applications. Unlike cloud-based speech processing, which requires continuous internet connectivity, Porcupine performs wake word detection locally, ensuring faster response times, enhanced security, and reduced dependency on external servers.

Once the wake word is detected, the system transitions to full speech recognition mode, using Vosk, a lightweight yet powerful speech-to-text engine. Vosk supports real-time transcription, enabling the system to interpret commands, process natural language inputs, and execute

corresponding actions efficiently. By integrating Porcupine and Vosk, this project ensures seamless voice interaction with low-power consumption and realtime processing.

Wake word detection plays a critical role in Software-Defined Vehicles (SDVs), smart home automation, and industrial AI applications, where hands-free control enhances convenience and safety. By leveraging edge AI for wake word recognition and speech processing, this system reduces latency, improves reliability, and offers a more natural user experience.

I've expanded the sections to add more technical depth, context, and industry relevance while ensuring the content remains focused on **voice recognition and authentication** for SDVs.

VOICE COMMAND RECOGNITION

Voice command recognition is a specialized form of speech recognition that enables systems to identify and process predefined commands spoken by a user. Unlike general speech recognition, which aims to transcribe full sentences, voice command recognition focuses on detecting specific phrases that trigger predefined actions. This technology is widely used in automotive applications, smart assistants, and industrial automation, allowing hands-free control of devices and systems.

Working Principle of Voice Command Recognition

Voice command recognition involves the following key steps:

1. **Audio Acquisition:** The system captures voice input through a microphone array and preprocesses the signal to remove noise and enhance clarity.

2. **Feature Extraction:** The audio signal is transformed into a spectrogram using signal processing techniques such as Mel-Frequency Cepstral Coefficients (MFCCs) or Wavelet Transforms.

3. **ClassificatCommand ion:** A machine learning model, such as a Hidden Markov Model (HMM), Support Vector Machine (SVM), or a deep neural network (DNN), is used to classify the extracted features and recognize the spoken command.

4. **Action Execution:** Once the command is recognized, the system maps it to a predefined action, such as adjusting climate control in a vehicle, opening an application, or initiating navigation.

Voice Command Recognition Using Python and MATLAB

Voice command recognition enables systems to detect and process spoken commands efficiently. Both **Python** and **MATLAB** offer powerful libraries for implementing AI-driven speech recognition solutions.

Python-Based Voice Recognition

Python provides various libraries for speech processing, including **Vosk** for offline transcription, **SpeechRecognition** for interfacing with cloud and local engines, **Librosa** for feature extraction, and **PyTorch/TensorFlow** for deep learning-based models.

MATLAB-Based Voice Recognition

MATLAB is widely used for speech signal analysis, leveraging tools such as the **Audio Toolbox** for processing, the **Deep Learning Toolbox** for training neural networks, the **Signal Processing Toolbox** for noise reduction, and **MATLAB Speech Command Recognition** for pre-trained classification models.

These tools enable efficient voice recognition across different platforms, supporting real-time processing and AI-driven automation.

Integration with NVIDIA Jetson AGX Orin

To optimize performance for embedded applications, the voice command recognition system is deployed on the NVIDIA Jetson AGX Orin, leveraging its GPU acceleration and AI capabilities. The combination of Porcupine (for wake word detection) and Vosk (for command recognition) ensures real-time processing with minimal latency.

By using both Python and MATLAB, developers can prototype in MATLAB for feature engineering and signal processing, then deploy optimized Python models on the Jetson AGX Orin for real-world execution.

1.3 Problem Statement

Existing voice recognition systems in vehicles are:

- **Limited in functionality**, allowing only pre-defined commands.
- **Dependent on cloud services**, causing **latency and security risks**.
- **Lack authentication**, making them **vulnerable to unauthorized usage**.

A **local, AI-powered voice control system** that operates independently of cloud infrastructure is required to provide **secure, real-time voice-based interaction** in SDVs.

1.4 Objectives

- **Develop an on-device voice recognition system** optimized for Jetson AGX Orin.
- **Implement a robust authentication mechanism** to verify user identity.
- **Enhance AI inference efficiency** using **CUDA and TensorRT**.
- **Ensure secure voice access** to vehicle functionalities.
- **Improve real-time performance** for seamless vehicle interactions.

LITERATURE REVIEW

2.1 Existing Systems

Current voice interfaces in vehicles include:

- **Google Assistant, Alexa Auto, Apple CarPlay**, which rely on **cloud-based processing**.
- **Proprietary in-car voice assistants (e.g., BMW Intelligent Personal Assistant, Mercedes-Benz MBUX)** that provide limited **local voice recognition**.
- **Biometric authentication in SDVs** is still in early stages, with limited adoption.

2.2 Challenges in Current Approaches

- **Cloud dependency** introduces **latency and privacy concerns**.
  Limited dataset generalization, causing recognition issues in different accents or environments.
- **No built-in authentication**, making them vulnerable to unauthorized access.

2.3 Summary of Related Work

- Studies show **on-device AI inference** reduces latency in voice recognition.
- **Machine learning models trained with MFCC and spectrograms** improve speaker identification.
- **Edge AI platforms like Jetson** outperform traditional cloud-dependent models.

IMPLEMENTATION

4.1 Voice Recognition & Authentication

This project involves the **end-to-end implementation** of a voice-controlled system, focusing on **real-time command processing and user authentication**.

4.1.1 Speaker Identification & Authentication

- A **voiceprint-based authentication system** is implemented, ensuring that only authorized users can control vehicle functions.
- **Deep learning models** trained on voice patterns **differentiate users based on speech characteristics**.

4.1.2 Feature Extraction (MFCC, Spectrograms)

- **Mel-Frequency Cepstral Coefficients (MFCCs)** and **spectrogram-based features** are extracted from voice samples.
- AI models use **pattern recognition** to verify **speaker identity**.

4.1.3 Model Training and Deployment

- **Pretrained deep learning models** fine-tuned for voice authentication.
- **TensorRT acceleration** ensures high-speed inference on Jetson AGX Orin.

4.2 Placeholder for Wake Word Detection & Speech-to-Text

- The wake word detection system is implemented using a convolution neural network(CNN) in MATLAB, leveraging its deep learning and signal processing toolboxes for accurate recognition
- Speech-to-text component is developed in MATLAB,utilizing advanced signal processing and the deep learning model for high accuracy.

TOOLS & TECHNOLOGIES USED

5.1 NVIDIA Jetson AGX Orin

- **CUDA-powered GPU acceleration** for real-time AI inference.
- Supports **TensorFlow, PyTorch, and ONNX models** for deep learning.
- **Dedicated AI cores** for high-performance speech recognition.

5.2 Python Libraries

- **Vosk** – Open-source speech recognition toolkit.
- **Porcupine** – Wake word detection library.
- **TensorFlow & PyTorch** – Model training and inference.

5.3 MATLAB for Feature Extraction

- Used for analyzing **MFCC, spectrograms**, and other audio features.

5.4 GPIO & CAN Interfaces

- Integrates **voice commands with vehicle control mechanisms**.

EXPERIMENTAL RESULTS & PERFORMANCE ANALYSIS

6.1 Accuracy and Processing Speed

- **Voice recognition model performance** tested under various conditions.
- **Authentication success rate** analyzed for different users.

6.2 Latency Analysis

- **On-device vs. cloud-based inference** comparison.
- **Optimized model execution** for faster response times.

6.3 Comparative Study with Other Systems

- **Google Assistant, Alexa Auto, and in-car voice assistants** benchmarked.
- **Jetson-powered models** analyzed for **low-latency AI inference**.

CONCLUSION & FUTURE WORK

7.1 Summary of Findings

This project demonstrated the **potential of on-device AI inference** for real-time voice recognition and authentication in **software-defined vehicles (SDVs)**. By leveraging **NVIDIA Jetson AGX Orin**, we achieved:

Reduced latency – Processing voice commands directly on the Jetson device eliminates the delays associated with cloud-based models.
Enhanced security – The implementation of **speaker authentication** ensures that only authorized users can issue vehicle commands.
Optimized AI performance – Using **TensorRT acceleration**, we improved inference speed, making real-time execution feasible for SDVs.

However, due to **hardware and software limitations**, we could only implement the **baseline version of voice authentication**. Further enhancements require additional **computational resources and improved datasets** for training.

7.2 Potential Improvements

To fully realize a **robust and production-ready system**, several improvements are necessary:

1. Advanced Speech Models for Higher Accuracy

- The current implementation uses **basic speaker authentication**, but **deep neural networks (DNNs)** such as **Convolutional Recurrent Neural Networks (CRNNs)** or **Transformer-based models** (like Wav2Vec) could significantly **improve accuracy and speaker differentiation**.

- Training these models requires **higher computational power and a larger dataset**, which were unavailable during this phase of development.

2. Hardware Upgrades for Better Performance

- **Higher-quality microphones**: Additional **directional microphones** or **multi-mic arrays** would improve **noise reduction** and enhance **speech clarity**.

- **Jetson Xavier or higher-end GPUs**: While Jetson AGX Orin is powerful, future versions could leverage **more powerful AI accelerators** for better deep learning inference.

3. Dataset Expansion & Real-World Testing

- The accuracy of voice authentication depends on the **quality and diversity of the training data**.

- Due to resource constraints, we used **a limited dataset**; larger datasets covering **different accents, noise conditions, and real-world environments** are required for **robust performance**.

4. Integration with Additional Vehicle Control Mechanisms

- Currently, the system is limited to **voice authentication**; future improvements should enable **more complex vehicle controls**, such as **adjusting settings, navigation, or emergency overrides**.

5. Energy Optimization for Real-Time Deployment

- Running AI models on **edge devices** like Jetson **consumes significant power**. Future improvements should explore **low-power AI optimizations**, making the system **efficient for long-term SDV deployment**.

7.3 Real-World Applications

- **Secure hands-free vehicle control** for autonomous fleets.

- **AI-driven voice interfaces** for smart automotive applications.

REFERENCES

1. NVIDIA Corporation, *Jetson AGX Orin Developer Kit* [Online].
2. Hinton, G. et al., "Deep Neural Networks for Speech Recognition," *IEEE Transactions on Audio, Speech, and Language Processing*, 2012.
3. R. Prasad et al., "Speaker Recognition Systems in Edge AI," *International Conference on Embedded AI Computing*, 2023.
4. TensorFlow Developers, *TensorFlow Speech Recognition Mode-ls* [Online].