

Design and Implementation of an Autonomous Driving Car Control System Based on Voice Interaction

Yangyang Xi
School of Electronic Information and Electrical Engineering
Yangtze University
Jingzhou, China

Abstract: To address the limitations of conventional remote-controlled miniature vehicle platforms, including a single human-machine interaction interface and cumbersome operation, this study proposes and develops an intelligent autonomous driving car system based on voice interaction. The system employs the STM32F103C8T6 microcontroller as the core processing unit and deeply integrates the LU-ASR01 embedded voice recognition engine to achieve efficient non-contact command control. Digitalized voice feature codes are transmitted to the main controller through UART asynchronous serial communication, and PWM (Pulse Width Modulation) technology is further utilized to drive the L293D dual H-bridge motor module, thereby enabling omnidirectional kinematic control of the vehicle, including forward motion, reverse motion, steering, braking, and multi-level speed regulation. Experimental results demonstrate that the proposed system exhibits a high command recognition accuracy, an end-to-end response latency strictly within 2 s, and excellent dynamic stability and robustness under complex testing conditions. This work provides a reliable design paradigm and theoretical basis for the engineering implementation of voice recognition technology in miniature intelligent mobile robots and embedded control systems.

Keywords: Voice Recognition; Autonomous Driving; STM32; L293D; Embedded System

1. INTRODUCTION

With the coordinated evolution of artificial intelligence algorithms and high-performance embedded microprocessors, voice recognition technology has become a core paradigm of next-generation intelligent human-machine interaction (HMI), and has been widely deployed in Internet of Things (IoT) home systems, in-vehicle infotainment platforms, and various intelligent terminal devices. Compared with traditional control approaches that heavily rely on physical buttons or radio-frequency remote controllers, voice interaction demonstrates remarkable advantages in terms of naturalness, efficiency, and hands-free operation. In the field of miniature mobile robots and intelligent vehicles, conventional solutions are increasingly constrained by lengthy operation paths and limited interaction dimensions. Therefore, introducing natural language processing and voice recognition mechanisms into the underlying vehicle control loop is of significant academic relevance and practical value.

In recent years, international technology enterprises have achieved substantial breakthroughs in acoustic model construction and complex semantic understanding, while domestic companies have also realized technological closure in the development of low-power offline voice recognition chips. This trend indicates that voice-controlled technology is progressing beyond the preliminary stage of simple keyword matching toward robust parsing of complex commands.

Against this background, this paper designs and implements a voice-controlled autonomous driving car system based on an STM32 embedded hardware architecture. By utilizing local computational resources to perform real-time decoding and execution of offline voice commands, the proposed system not only substantially improves user interaction experience, but also provides an innovative and practical solution for the development of low-cost intelligent hardware platforms.

2. Overall System Architecture Design

2.1 Overall System Architecture

To achieve highly reliable mapping of voice commands to vehicle control actions, this study constructs a closed-loop

embedded control architecture with multi-module collaboration. The system follows a top-down modular design philosophy and is divided into four layers: the perception input layer, the core decision-making layer, the execution output layer, and the power management layer. These subsystems are decoupled through standardized bus interfaces, thereby ensuring a clear hardware topology and high scalability.

From the perspective of signal flow and system dynamics, the core operating mechanism can be summarized as “acoustic signal acquisition – feature mapping – mechanical actuation.” Specifically, the LU-ASR01 voice front-end array captures spatial acoustic signals, and its internal DSP performs preprocessing operations, including adaptive noise suppression, endpoint detection, and Mel-frequency cepstral coefficient extraction. The matched command is encoded into a specific digital sequence and then transmitted to the STM32 controller through the UART interface. The microcontroller performs logical arbitration based on an internal finite state machine (FSM) and generates control levels with designated duty cycles. Finally, the L293D driver dynamically regulates the armature voltage and polarity of the DC motors according to the logical instructions, thereby achieving precise control of vehicle trajectory and speed.

Furthermore, to cope with extreme acoustic interference scenarios, such as severely degraded signal-to-noise ratio conditions, the system incorporates a redundantly designed physical-button intervention mechanism. This mechanism supports higher-priority hard-interrupt start/stop control and gear calibration, significantly improving the fault tolerance and engineering robustness of the system. The overall system block diagram is shown in Figure 1.

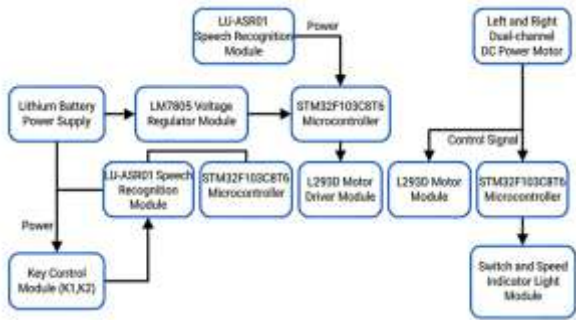


Figure 1: System Overall Hardware Architecture Block Diagram



Figure 2: Core Washing Finite State Machine

3. SYSTEM HARDWARE DESIGN

3.1 Main Control Circuit

The hardware layer serves as the physical carrier for real-time computation and stable output. The hardware topology proposed in this paper includes the microcontroller, voice front-end, bridge-type motor driver, human-machine feedback array, and linear voltage regulation network. The STM32F103C8T6 chip, based on the ARM Cortex-M3 core, is selected as the central decision-making unit of the system. Compared with traditional 8-bit CISC architectures, such as the 8051 series, this microcontroller provides a 72 MHz operating frequency, a Nested Vectored Interrupt Controller (NVIC), and abundant on-chip peripherals, including USART, 16-bit timers, and GPIO ports. These features make it highly suitable for the system requirements of concurrent interrupt handling and high-precision PWM waveform generation. In terms of pin assignment, PA9 (TX) and PA10 (RX) are configured as multiplexed push-pull output and floating input, respectively, and are dedicated to high-speed asynchronous communication with the voice module. The main control circuit is shown in Figure 2.

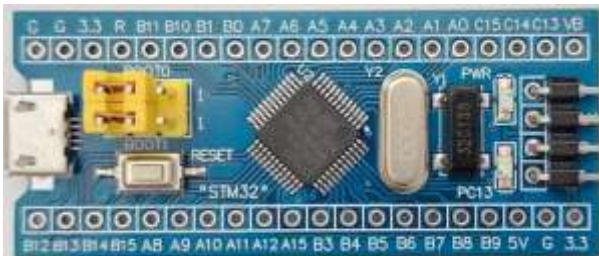


Figure 2: Main Control Circuit

3.2 Offline Voice Recognition Front-End

The perception layer employs the LU-ASR01 integrated offline voice recognition engine module. This module embeds a highly optimized neural-network-based acoustic model, enabling millisecond-level command retrieval even in offline conditions without cloud computing support. Its integrated ADC and signal-conditioning circuit effectively suppress high-frequency aliasing noise. Compared with the previous-generation LD3320 chip, the LU-ASR01 exhibits overwhelming advantages in dynamic recognition accuracy, acoustic robustness under background noise, and hardware packaging compactness, thereby substantially shortening the system development cycle.

3.3 Dual H-Bridge Motor Drive Network

The execution layer adopts the L293D integrated dual H-bridge driver IC for stepper and DC motor control. Based on the steering requirements calculated by the main controller, the MCU outputs a logic matrix to the L293D via GPIO ports to realize differential speed control of the two motors. During component selection, although the L298N provides a higher peak current capability, it suffers from excessive static power consumption and redundant packaging. In contrast, the L293D, with its compact DIP package, built-in clamp diodes, and low on-resistance characteristics, is more suitable for the stringent constraints on chassis space and power budget in miniature vehicle applications. The offline voice recognition front-end is shown in Figure 3.

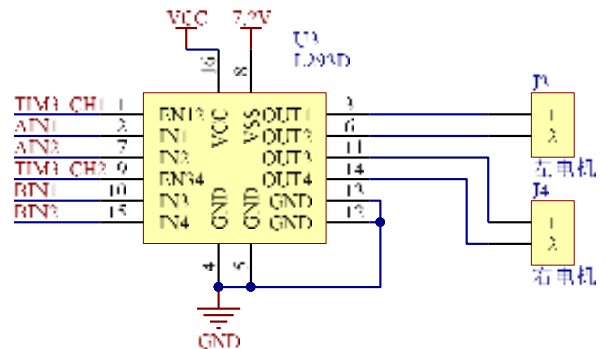


Figure 3: Offline Voice Recognition Front-End

3.4 Dual H-Bridge Motor Drive Network

The system deploys physical tactile buttons and a high-brightness LED matrix to form a human-machine feedback loop. The buttons trigger the external interrupt lines (EXTI) of the MCU to achieve ultra-low-latency gear switching. The LED indicator array, driven according to the current PWM output duty cycle, provides real-time visual representation of the vehicle's operating mode and dynamic speed level through logic mapping, thereby constructing a complete visual feedback mechanism.

To ensure that logical devices are not affected by back electromotive force ripples generated during instantaneous motor start-stop operations, the power management network adopts an LM7805 three-terminal linear regulator. The battery source voltage first passes through a π -type filtering network composed of a large-capacity aluminum electrolytic capacitor and a bypass decoupling ceramic capacitor, after which the LM7805 provides a stable 5 V DC reference supply. This configuration offers strong power support for the entire

system with a high power supply rejection ratio (PSRR).The dual h-bridge motor drive network is shown in Figure 4.

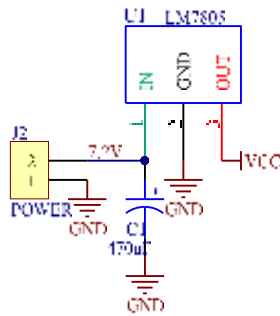


Figure 4: Dual H-Bridge Motor Drive Network

4. SYSTEM TESTING AND RESULTS

The robustness of the control algorithm and driver program is the key determinant of the temporal determinism of the system. In this study, the firmware is developed in the C language under the Keil μ Vision5 integrated development environment, and the overall software architecture adopts a foreground–background model consisting of a polling-based main loop and interrupt service routines.

During the system reset and initialization phase, the firmware sequentially completes the configuration of the system clock tree (RCC), GPIO mode remapping, timer prescaler settings, and UART baud-rate calibration. After initialization converges, the MCU enters the idle loop and waits for peripheral-triggered events.

When the voice module completes local acoustic matching and generates a serial receive interrupt (RXNE), the MCU enters the interrupt service routine to extract the data frame, after which the state machine parses the command macro definitions. For example, when an “accelerate” command is parsed, the MCU synchronously updates the global target-speed variable and invokes the duty-cycle update register function. When a “brake” command is detected, the logic enable terminal of the L293D is immediately disabled to achieve emergency stopping through electromagnetic damping. Meanwhile, the button-scanning function adopts a non-blocking timer-based debounce algorithm, thereby completely eliminating the risk of thread lock caused by mechanical chattering. The program flow diagram is shown in Figure 4.

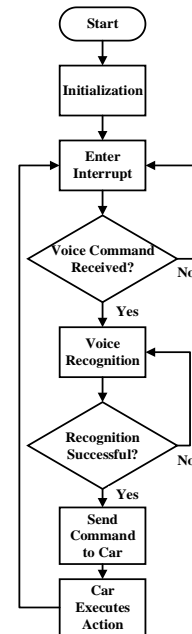


Figure 5: Program Flow Diagram

5. Experimental Verification and Performance Evaluation

To objectively evaluate the control precision and dynamic response performance of the proposed system, a closed testing environment was established to conduct system-level integration debugging and experimental measurements.

First, static electrical parameter calibration was performed. Oscilloscope probes were used to monitor the electrical characteristics at key nodes, verifying that the voltage fluctuation at the LM7805 regulation node was strictly limited within ± 50 mV, thereby ensuring logic-level integrity. At the same time, the frame loss rate of the serial communication link was verified to be zero.

Dynamic performance testing focused on the latency and accuracy of the voice control chain. Experimental results indicate that, under an ambient background noise level below 60 dB, the system achieved a qualified recognition accuracy for the predefined command library covering multiple vocal characteristics. By recording the time difference between command issuance and vehicle displacement using a high-frame-rate camera, the end-to-end system response latency was measured and confirmed to be strictly constrained within 2000 ms. The drive mechanism operated smoothly without obvious resonance, and the speed-level switching exhibited favorable linearity, fully satisfying the expected design constraints. The physical prototype is shown in Figure 4.



Figure 6: Physical Prototype

6. CONCLUSION AND FUTURE WORK

In this design, the LU-ASR01 voice recognition module is used to collect voice commands, convert analog signals into digital signals through internal processing, and send them to the STM32 microcontroller via a serial communication interface. The microcontroller issues corresponding commands to control the motor drive module's working state, realizing the car's autonomous driving function. Compared to traditional voice recognition cars, this system has more stable performance and more sensitive voice command recognition. However, due to time constraints, the system still has shortcomings, such as failing to accurately recognize voice commands when the car moves too far away or when external noise is loud. Future work will focus on improving recognition accuracy in noisy environments and extending the operational distance.

7. ACKNOWLEDGMENTS

The authors would like to thank the laboratory staff for their technical support and the experts who contributed to the development of this project.

8. REFERENCES

- [1] J. W. Yao, "Research and Design of Voice Recognition and Voice Synthesis in Human-Machine Interaction of Smart Pump Station Platform," Ph.D. dissertation, Univ. of Science and Technology of China, Hefei, 2021.
- [2] L. H. Zhang and G. Wang, "Research on the Application Difference between 51 Microcontroller and Arduino—Taking Electronic Clock Design as an Example," *Industrial Control Computer*, vol. 31, no. 10, pp. 137-139, 2018.
- [3] H. L. Liao, T. Ma, S. X. Feng, et al., "Design and Implementation of Unmanned Exploration and Rescue Vehicle Based on Alibaba Cloud and Raspberry Pi," *Internet of Things Technology*, vol. 12, no. 01, pp. 81-84, 89, 2022.
- [4] Z. T. Lin, "Design of Voice Control System for Smart Car Based on Android Platform," Ph.D. dissertation, Qingdao Univ. of Science and Technology, Qingdao, 2020.
- [5] W. C. Chen, H. Zhang, and J. T. Zhang, "Application of ESP8266-WiFi Module in Smart Car Control," *Industrial Control Computer*, vol. 32, no. 07, pp. 134-136, 2019.

[6] S. Y. Chen, L. Mu, and X. W. Wang, "Design of Warehouse Alarm Smart Car Based on Arduino," *Equipment Manufacturing Technology*, no. 06, pp. 30-33, 52, 2021.

[7] K. Ussen, C. M. Wa, and B. Osman, "Speech recognition datasets for low-resource Congolese languages," *Data in Brief*, vol. 52, 2024.