Design and Implement of Smart Voice Controlled Two-Wheeled Self-Balancer for Following and Avoidance

Liu Hao and Suchada Sitjongsataporn
[1]The Electrical Engineering Graduate Program, , Faculty of Engineering and Technology
Mahanakorn University of Technology International College (MUTIC)
Mahanakorn University of Technology, Bangkok, Thailand

[2]Department of Electronic Engineering, School of Electrical and Electronic Engineering (SEE)
Faculty of Engineering and Technology, Mahanakorn University of Technology, Bangkok, Thailand
Email: ssuchada@mut.ac.th

Abstract:

Self-balancing robots, traditionally used for entertainment or educational purposes, are now increasingly being explored for their potential to assist elderly people and individuals with limited mobility. Such robots can provide various functionalities such as transportation, monitoring, and interaction, improving the quality of life for users. However, existing two-wheeled self-balancing robots mainly focus on basic mobility and remote control through mobile applications, with limited emphasis on human-robot interaction and adaptability to dynamic environments. This paper proposes a solution to enhance the interaction capabilities of self-balancing robots by integrating voice control and intelligent following the obstacle avoidance features. Traditional remote control methods may not be intuitive for all users, especially elderly or disabled individuals, who could benefit more from voice commands that allow hands-free operation. Additionally, in the dynamic home environment, robots must be able to follow users while avoiding obstacles autonomously, ensuring both efficiency and safety. From the test data, it can be seen that the speech recognition system performs best in a quiet environment, with an identification accuracy rate of 95%. As the noise level increases, the system's identification rate gradually decreases, but still maintains an accuracy rate of over 90%, indicating that the system is reliable for use in a home environment. Thus, this study aims to design and implement an intelligent system that combines offline speech recognition with ultrasonic sensor-based following and obstacle avoidance, providing a more seamless and user-friendly experience.

Index Terms-- Two-wheeled self-balancing robot, voice interaction, offline speech recognition, ultrasonic sensor, following mode, obstacle avoidance.

I. INTRODUCTION

In recent years, the development of self-balancing robots has received significant attention in both research and practical applications, particularly in the fields of smart homes and assistive technologies [1]. These robots were originally designed for entertainment or educational purposes, but now they have gained increasing attention due to their potential in assisting the elderly and those with mobility impairments. Such robots can offer various functions such as transportation, monitoring, and interaction, thereby improving the quality of users' daily lives. However, the existing two-wheeled self-balancing robots mainly focus on basic mobility capabilities and remote control through mobile applications, with relative less emphasis on human- computer interaction and adaptability to dynamic environments.

This paper proposes a solution aimed at enhancing the interaction capabilities of self- balancing robots by integrating voice control and intelligent following obstacle avoidance functions. Traditional remote control methods may not be intuitive for all users, especially for the elderly or disabled groups, who need voice commands that can be operated without hands [2]. Moreover, in dynamic home environments, the robot must be able to follow the user while autonomously avoiding obstacles to ensure efficiency and safety. Therefore, this research aims to design and implement an intelligent system that combines offline speech recognition with following and

Received: 26 September 2025
Revised: 9 December 2025
Accepted: 11 December 2025
* Corresponding author: Assoc. Prof. Dr. Suchada Sitjongsataporn, Department of Electronic Engineering, School of Electrical and Electronic Engineering (SEE), Bangkok, Thailand. (Email: ssuchada@mut.ac.th).

obstacle avoidance functions based on ultrasonic sensors, providing a more seamless and user-friendly experience. By addressing these issues, this research provides new knowledge contributions for integrating advanced interaction technologies into low-cost, daily-assistive robots in this field. The robot model is shown in Figure 1.

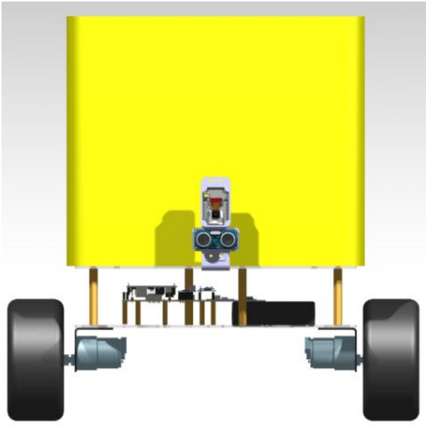


Figure 1 Three-dimensional digital model of a two-wheel balancing robot

The proposed self-balancing robot adopts a modular hardware and software architecture, enabling seamless integration of enhanced interaction functions (such as voice control and intelligent following/obstacle avoidance). The basic balance and motion control rely on the combination of traditional Arduino UNO, MPU6050 accelerometer/gyroscope, and TB6612 motor driver [3], while the interaction expansion is achieved by introducing additional modules, which expand the robot's perception capabilities and human-machine interaction system architecture as shown in Figure 2.

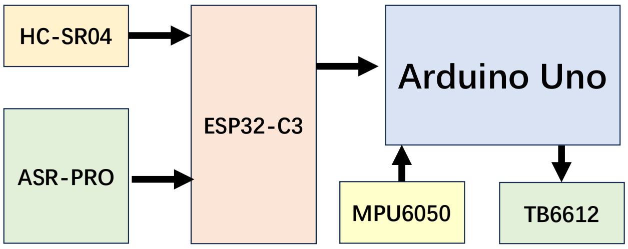


Fig. 2 Hardware structure diagram of the two-wheel balancing robot

II. HARDWARE DESIGN

This self-balancing robot adopts a modular hardware-software architecture to achieve seamless integration of balance control, voice control, intelligent following, obstacle avoidance and other enhanced interactive functions. The system can be divided into four core layers at the system level and the data flow is clear. Among them, the core architecture foundation relies on the classic combination of Arduino UNO, MPU6050 accelerometer/gyroscope, and TB6612 motor driver to ensure basic operation. The perception layer realizes environmental and command perception through MPU6050, HC-SR04 ultrasonic sensor, and ASR-PRO module[4]. The processing layer is handled by the ESP32-C3 microcontroller to process interaction data and generate advanced control instructions [5], while Arduino UNO is responsible for real-time balance control and motor driving to complete data parsing and instruction generation. The control layer precisely controls motor actions and mode switching logic to support the execution of various control modes, and the interaction layer provides human-computer interaction functions such as voice commands, user following, and obstacle avoidance. The overall data flow is the voice/audio signal, ultrasonic measurement data processed by ESP32-C3 to convert into advanced decisions and then transmitted to Arduino UNO, which then executes precise motor control. The layered design takes into account both balance stability and interaction flexibility.

A. HC-SR04 ultrasonic sensor

HC-SR04 is a low-cost and high-performance ultrasonic distance measurement sensor widely used in small and medium-sized electronic projects. Its core function is to calculate the distance to an object by transmitting and receiving ultrasonic signals. In the two-wheel self-balancing robot C-SR04, it is used for obstacle detection and following functions [6]. The robot measures the distance in front to determine whether it encounters an obstacle or to determine the position of the target object. This enables the robot to avoid collisions and maintain tracking of the target, enhancing its adaptability in complex environments. Its working principle follows the "time difference ranging method": when the Trig pin receives a high-level trigger signal sent by the microcontroller for at least 10 microseconds, the sensor automatically transmits 8 40kHz ultrasonic pulses; the ultrasonic waves encounter an obstacle and are reflected, captured by the receiving probe, and the Echo pin immediately outputs a high level.

The duration of the high level is proportional to the round-trip time of the ultrasonic wave from emission to reception; the microcontroller calculates the duration of the high level on the Echo pin, combined with the propagation speed of the ultrasonic wave in the air (approximately 340m/s), and can calculate the actual distance between the sensor and the obstacle using the formula "distance = (high level time × 340m/s) / 2". The effective measurement range of HC-SR04 is generally 2cm - 400cm, with a measurement accuracy of up to 3mm, and it can be quickly put into use without complex calibration. As shown in Figure 3.

Fig. 3 HC-SR04 [7]

B. ASR-PRO

ASR-PRO is a universal, low-power and high-performance speech recognition chip specifically developed for the development of low-cost offline speech application solutions. It features speech recognition, voiceprint recognition, speech enhancement, and speech detection functions. It performs strongly in echo elimination and environmental noise suppression, significantly enhancing the speech recognition effect. It has high-precision speech recognition capabilities and can process and interpret voice commands without the need for the internet.

In this paper, ASR-PRO is mainly used to receive users' voice inputs, such as initiating follow-up mode, stopping mode [8]. Its embedded design enables it to execute voice commands in low latency, making it the key hardware for achieving barrier-free interaction. It can perform offline speech recognition without a network connection, with an identification rate of over 98% [9], capable of achieving 10-meter long-range recognition, response time less than 0.1 seconds, and supports multiple global languages including Chinese, English, and Japanese shown in Figure 4.

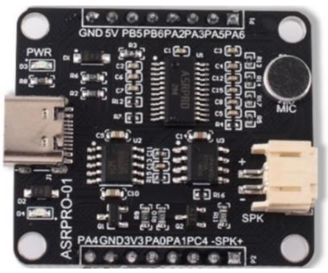


Figure 4 ASR-PRO [10]

C. Chassis Design

The chassis of this design adopts a double- layer structure. By combining the advantages of light weighting, ease of assembly and debugging, it ensures that the robot can meet the 2kg load capacity requirement while maintaining high stability and flexibility. The chassis material is made of corrosion- resistant acrylic sheets and nylon columns, which not only ensures the structural strength but also optimizes the use of internal space, facilitating the placement of components such as batteries, sensors, and control modules.

The design of the chassis also takes into account modularity, ease of maintenance, and scalability. In the future, it can be easily upgraded with hardware or added sensors. The overall layout is shown in Table 1.

Table I

Balanced robot chassis overall layout

Module Category	Module Name	Weight (g)	Placement Location	Balance Design Key Points
Balance Control System	Arduino uno	25	Right Wheel Axle	The correlation module centrally arranges the center of gravity to fall on the center line of the axle, effectively reducing the risk of static tilt.
	MPU6050 6-Axis IMU Sensor	12		
	TB6612 Motor Driver	10		
	12V-to-5V Power Module	15		
	PCB	27		
Power System	Battery	278	Left Wheel Axle	
Video Ranging System	ESP32-CAM Camera Module [11]	18	Front Chassis Edge	The two modules counteract gravity in front and back to improve balance and ease of commissioning
	HC-SR04 Ultrasonic Module	15		
Command Parsing System	ESP32-C3	14	Rear Chassis Depression	
	ASR-PRO Voice Module	15		

III SOFTWARE DESIGN

A. Voice Interaction Software Design

The software architecture of the voice interaction system is designed to ensure seamless communication between the offline speech recognition module (ASR-PRO), the ESP32-C3 microcontroller, and the Arduino UNO. The ASR-PRO is responsible for real-time speech feature extraction and pattern matching based on the pre-defined command library, enabling local recognition without reliance on external networks.

Upon detecting a valid command, the ASR-PRO transmits the recognition result via a UART interface to the ESP32-C3, where the software implements a parsing and decision-making layer. This layer maps the recognized speech token to specific robot control instructions, ensuring that commands such as “Forward,” "Backward," or "Obstacle Avoidance Mode" are translated into standardized control signals.

These signals are then relayed through serial communication to the Arduino UNO, which executes the corresponding motion by regulating the motor driver (TB6612) while maintaining dynamic balance using sensor feedback from the MPU6050.The program flowchart is shown in Figure 5.

The software workflow thus achieves a clear separation of roles: ASR-PRO handles recognition tasks, ESP32-C3 manages command logic and higher-level interaction states, and Arduino UNO ensures stability control. To enhance robustness, the software incorporates error-handling mechanisms, such as filtering ambiguous or low-confidence recognition results, and employs a dual-level command design that distinguishes between basic locomotion and autonomous behavioral modes.

This modular architecture not only optimizes computational load distribution but also improves system scalability, allowing new commands or modes to be integrated with minimal modifications. Compared with mobile-app-based control, the voice interaction software offers a hands-free, low-latency, and privacy-preserving interface, particularly beneficial for elderly or mobility-impaired users in home environments.

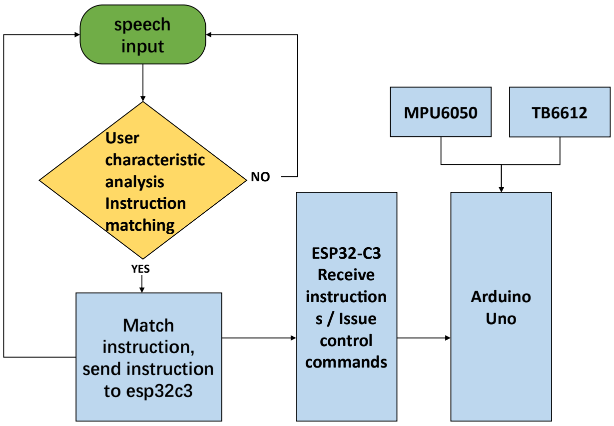


Fig. 5 Voice Interaction Software Design Flowchart

B. Obstacle Avoidance/Following Software Design

To ensure that the robot can operate safely and freely in the home environment, this system uses ultrasonic sensors HC-SR04 to achieve obstacle detection and obstacle avoidance functions [12]. When the robot is in the following mode, the ultrasonic sensors are used to detect the distance between the robot and the target (such as the user).

Based on the set threshold, the robot can automatically adjust its speed and direction to maintain a safe distance from the target. When the target approaches, the robot will automatically stop to prevent a collision. Ensure to always follow the target and maintain an appropriate distance.

In the obstacle avoidance mode, the robot uses ultrasonic sensors to monitor the obstacles in front of it in real time [13]. When the detected obstacle distance is less than the predetermined threshold, the ESP32-C3 will decide whether to perform the obstacle avoidance action based on the sensor feedback.

If an obstacle is detected and the current position is not suitable for turning left or right, the robot will choose to bypass the obstacle and find the best path by turning left or right. This obstacle avoidance logic can dynamically respond to environmental changes and ensure the robot's smooth operation in complex home environments. The program flowchart shows in Figure 6.

III. SYSTEM TESTING

In order to ensure the feasibility and reliability of this system in real household environments, a system test was conducted on the robot after its design was completed. The significance of this test lies in verifying its stability, intelligence and interactivity in complex environments, thereby determining whether it can meet the application requirements of household service robots. The household environment is different from the laboratory environment, featuring narrow spaces, complex obstacles, and obvious noise interference.

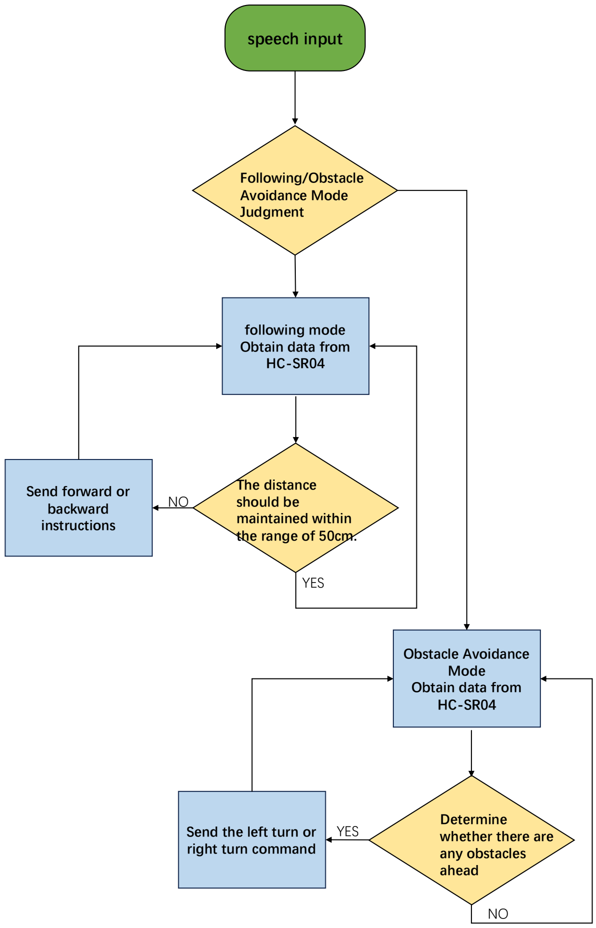


Fig. 6 Obstacle Avoidance/Following Software Design Flowchart

These factors may all affect the operation effect of the robot. Therefore, through multi-scenario and multi-angle tests, not only can the overall performance of the system be evaluated, but also potential deficiencies can be discovered and used as a basis for subsequent optimization.

This test is mainly divided into three parts as follows:

1) following function test
2) obstacle avoidance function test, and
3) voice control function test.

A. Following the functional test

The following function test is mainly used to verify whether the robot can follow the target stably in a dynamic environment and maintain an appropriate safe distance. According to the application scenarios of household service robots, the target may have actions such as uniform forward movement, sudden acceleration, sudden stop. Therefore, we designed multiple different test scenarios, including the target moving uniformly, suddenly stopping, accelerating and decelerating.

In this test, the vehicle moved forward, stopped or accelerated at different speeds on a straight path, simulating various situations encountered in daily life. The car continuously detected the distance between the target and itself using ultrasonic sensors, and adjusted its speed in real time based on the behavior of the target [14] to maintain a safe distance of approximately 50cm between itself and the target. To ensure the diversity of the test, we set up different test scenarios, including the target moving at a constant speed, the target suddenly stopping, the target accelerating and decelerating, etc. The test data are shown in Table 2 and Figure 7.

Table II

Following the results of the functional test

Test Scene	Response of the vehicle	Average following error	Success rate
Move at a constant speed	The vehicle maintained a stable follow-up, with the error remaining within 5 cm.	±5cm	98%
Uniform deceleration	The car slowed down smoothly without any collision.	±6cm	96%
Suddenly stop	The car responded quickly and came to a stop within 0.8 seconds.	±7cm	94%
Accelerate forward	The car slightly slowed down before accelerating, and the error increased.	±9cm	92%

Figure 7 Follow the functional test

Based on the experimental results, the car was able to stably follow the target in most cases, performing best especially when moving at a constant speed or when the target suddenly stopped. In scenarios where the target was accelerating, the car's response was slightly slower, but it was still able to adjust its speed relatively smoothly. Overall, the car was able to successfully follow the target while ensuring safety.

B. Obstacle avoidance function test

where the robot must be able to promptly sense and avoid obstacles. Considering the common obstacles in household environments such as furniture and walls, the obstacle avoidance capability directly determines the safety and usability of the robot [15]. The obstacle avoidance test is mainly verified by placing obstacles on the path of the vehicle. The test examines the robot's reaction ability. The vehicle detects the distance of the obstacles in front in real time using ultrasonic sensors and adjusts its movement to avoid the obstacles. If the test as shown in Figure 8 is conducted.

In this test, the balancing robot uses the HC-SR04 ultrasonic sensor to identify obstacles in the environment and take corresponding obstacle avoidance actions. Meet the design requirements.

Figure 8 Obstacle avoidance function test

C. VOICE CONTROL FUNCTION TEST

Voice control is an important way to enhance the interaction experience between robots and users. In a home environment, voice control not only makes it convenient for users to operate but also enhances the interactivity between the robot and family members. The testing method involves continuously issuing 10 instructions (such as "move forward", "stop", "obstacle avoidance mode") in three environments: a quiet environment, normal household noise, and a noisy environment. We measured the recognition accuracy and response time in each environment to ensure the stability of the system under different noise disturbances. The test results are shown in Table 3 and Figure 9.

Table III Following the results of the functional test

Environmental noise level	Total number of instructions	Correct identification of numbers	Recognition accuracy rate	Average response time
30 dB (quiet)	10	10	100%	0.5s
40 dB (quieter)	10	10	100%	0.6s
50 dB (for home use)	10	9	90%	0.7s
60 dB (noisy)	10	9	90%	0.8s
70 dB (noise level)	10	8	80%	0.9s

From the test data, it can be seen that the speech recognition system performs best in a quiet environment, with an identification accuracy rate of 100%. As the noise level increases, the system's identification rate gradually decreases, but still maintains an accuracy rate of over 90%, indicating that the system is reliable for use in a home environment.

Figure 9 Voice control to turn right in a noisy environment

V. SUMMARY AND CONCLUSION

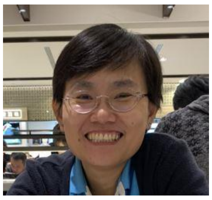
This study addresses the limitations of existing two-wheeled self-balancing robots—insufficient human-robot interaction and poor adaptability to dynamic home environments— by designing and implementing an intelligent prototype centered on ESP32-C3 (interaction hub) and Arduino Uno (balance control core). The hardware adopts a modular layered architecture: the perception layer integrates MPU6050 (posture detection), HC-SR04 ultrasonic sensor (following/ obstacle avoidance distance measurement), and ASR-PRO (offline speech recognition); the processing layer uses ESP32-C3 to parse data and generate control instructions; the control layer relies on Arduino Uno and TB6612 to achieve real-time balance and motor regulation; and the double-layer acrylic chassis optimizes component layout to ensure a centralized center of gravity and 2kg load capacity. In software, it realizes three core functions: offline voice interaction (98%+ recognition rate in quiet environments, 80%+ in 70dB noise), intelligent following (92%-98% success rate, ±5cm-±9cm average error), and ultrasonic-based obstacle avoidance.

System tests in simulated home environments confirm technical feasibility: the prototype maintains stable balance on various surfaces, voice control meets daily noise resistance needs, and following/ obstacle avoidance functions ensure safe operation around furniture and narrow spaces. Practically, it enhances accessibility for the elderly and mobility-impaired via hands-free voice control and automatic following, while offline recognition avoids privacy risks, and modular hardware supports maintenance and upgrades. However, limitations exist: slight lag in following accelerating targets, reduced voice recognition in >70dB noise. Future optimizations will integrate multi-sensor fusion (ultrasonic + visual tracking), adaptive noise cancellation[16], and multi-directional obstacle detection to improve performance. Overall, this research provides a feasible low-cost solution for household assistive robots[17], laying a foundation for their popularization in smart homes.

REFERENCES

[1] Y. Zhang, “Implementing Bluetooth Low Energy for Smart Home Devices,” Journal of Smart Home Technologies, vol. 3, pp. 23-28, 2020.
[2] P. Zhang, “Voice-Controlled Robotics for the Elderly: A Case Study,” Journal of Robotics and Human Interaction, vol. 19, pp. 33-40, 2020.
[3] J. Smith, “Development of a Two-Wheeled Self-Balancing Robot Using PID Control,” Journal of Robotics and Automation, vol. 8, no. 2, pp. 45-53, 2019.
[4] S. Lee, “MPU6050 Sensor Calibration and Data Fusion for Attitude Estimation,” Sensors, vol. 20, no. 5, pp. 1342-1355, 2020.
[5] H. Wang, “ESP32-Based Wireless Communication System for Home Automation,” IEEE Access, vol. 7, pp. 123456-123465, 2019.
[6] A. Kumar, “Design of Autonomous Robot for Obstacle Detection and Avoidance,” Journal of Robotics Systems, vol. 8, no. 1, pp. 15-25, 2021.
[7] https://www.robotics.org.za/HC-SR04?search=HC-SR04.
[8] P. Turner, “Improving Voice Recognition for Assistive Robotics,” Journal of Human-Robot Interaction, vol. 14, no. 2, pp. 20-35, 2021.
[9] D. Wilson, “A Review of Speech Recognition in Mobile Devices,” Mobile Computing and Communication, vol. 15, no. 4, pp. 108-115, 2019.
[10] https://www.ebay.com/itm/295431820154?_skw= ASR-PRO.
[11] R. Brown, “Performance Analysis of ESP32-CAM in IoT Video Surveillance Applications,” International Journal of Internet of Things and Its Applications, vol. 12, no. 3, pp. 67-78, 2020.
[12] L. Nguyen, “Obstacle Avoidance Techniques for Autonomous Vehicles,” International Journal of Autonomous Systems, vol. 13, no. 2, pp. 50-62, 2020.
[13] A. Lopez, “System Design of an Autonomous Mobile Robot with Obstacle Avoidance and Following Modes,” Journal of Robotics Research, vol. 11, no. 1, pp. 11-19, 2021.
[14] G. Zhang, “Real-Time Object Detection for Autonomous Robots,” Journal of Robotics and Artificial Intelligence, vol. 9, no. 2, pp. 75-89, 2018.
[15] K. Zhao, “Real-Time Obstacle Avoidance for Service Robots,” Robotics and AI Journal, vol. 21, no. 1, pp. 123-130, 2021.
[16] L. Turner, “Sensor Fusion for Enhanced Robot Navigation,” International Journal of Robotics and Automation, vol. 22, no. 3, pp. 76-84, 2019.
[17] J. Kim, “Design of Low-Cost Two-Wheeled Robots for Home Automation,” IEEE Robotics and Automation Magazine, vol. 5, pp. 40-50, 2017.

Liu Hao graduated from Kunming University of Science and Technology in Yunnan, China in 2022 and works at Chery Automobile. He is currently pursuing a master's degree in electrical engineering (electronics) at Mahanakorn University of Technology, Thailand. His research interests include embedded systems and automatic control systems.



Suchada Sitjongsataporn received her B.Eng. (first-class honors) and D.Eng. degrees in Electronic Engineering from the Mahanakorn University of Technology, Bangkok, Thailand in 2002 and 2009, respectively. She has worked as a lecturer at the department of electronic engineering, Mahanakorn University of Technology since 2002. Currently, she is an Associate Professor in electronic engineering, the Assistant President for Research and Dean of Graduate College in the Mahanakorn University of Technology. Her research interests are in mathematical and statistical models in the area of adaptive signal processing for communications, networking, embedded system, and image and video processing.

.