KMUTT ภูมิปัญญา 65 ปี สู่วิถีความยั่งยืน | คณะวิศวกรรมศาสตร์ Faculty of Engineering | PROJECT DAY 2026 FACULTY OF ENGINEERING

INC06

SMART COMMUNICATION SYSTEM FOR SERVICE DELIVERY ROBOTS

CONTROL SYSTEMS AND INSTRUMENTATION ENGINEERING PROGRAM

MISS SARUNPORN SAMANWORAKUL , MR. PANTAKAN KIATPONGSA , PROJECT ADVISER: ASST. PROF. DR.PRANCHALEE SAMANPIBOON

SYSTEM ARCHITECTURE

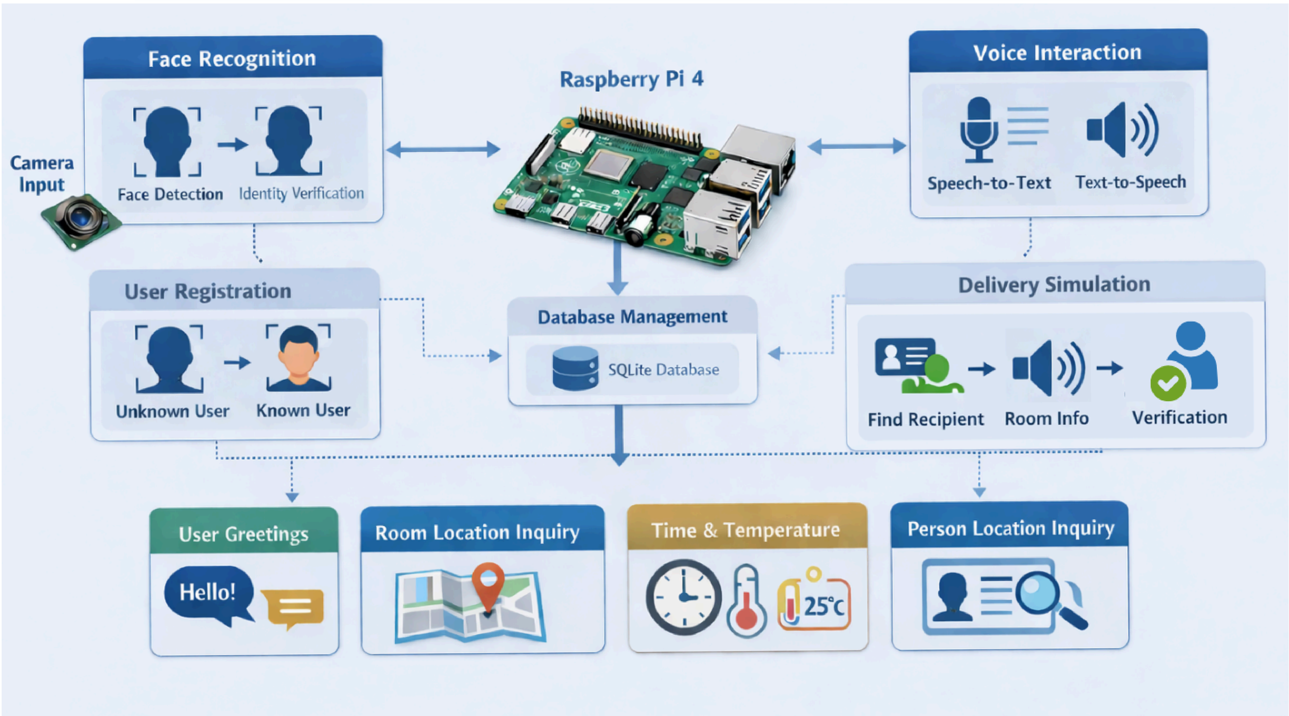

Figure 1 : System Overview

1|ABSTRACT

This project aims to develop an intelligent communication and interaction system for service delivery robots to enhance the efficiency of human-robot interaction. The development focuses on addressing response latency and improving the accuracy of recipient identification. The proposed system integrates three core components: a facial recognition system using LBPH algorithms and SQLite for secure user identification; a voice interaction system evaluated through comparative performance testing of Vosk, Groq, and Google Speech-to-Text; and a voice command interface for real-time information retrieval such as weather, time, and navigation status. The experimental results indicate that selecting optimal models significantly reduces communication latency, enabling real-time responses. Furthermore, the facial recognition system successfully manages user registration and authentication via the integrated database. These findings suggest that the smart communication system effectively enhances the naturalness of robot interactions and demonstrates strong potential for practical deployment in autonomous delivery services.

2|METHODOLOGY

The project combines advanced face recognition and voice interaction systems, utilizing a Raspberry Pi 4 as the central processing unit. The face recognition module employs OpenCV with the Haar Cascade algorithm for real-time detection. In contrast, the LBPH (Local Binary Pattern Histogram) algorithm extracts facial features to verify identities against an SQLite database. Simultaneously, the voice interaction module processes audio through the Vosk model for Speech-to-Text (STT) conversion, followed by Natural Language Processing (NLP) to analyze user intent and match it with predefined commands. An integrated logic controller synchronizes these inputs to facilitate secure identity verification and execute tasks within the delivery simulation. The system also generates feedback through Edge TTS, producing natural-sounding voice responses that enhance the human-robot interaction experience. Designed for delivery simulation, this system identifies recipients and room locations, ensuring secure delivery through final identity verification. Additionally, it includes interactive features such as user greetings, location inquiries, and real-time updates on time and temperature, all of which contribute to a seamless communication experience between humans and robots.

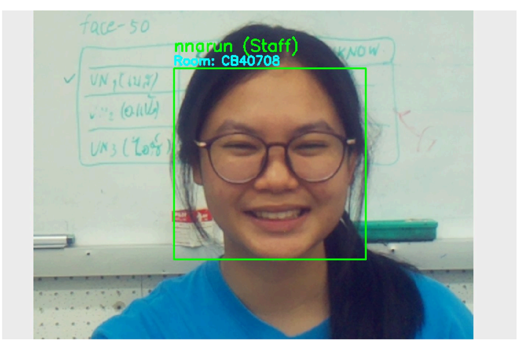

Figure 2 : Face Recognition using LBPH to verify identities

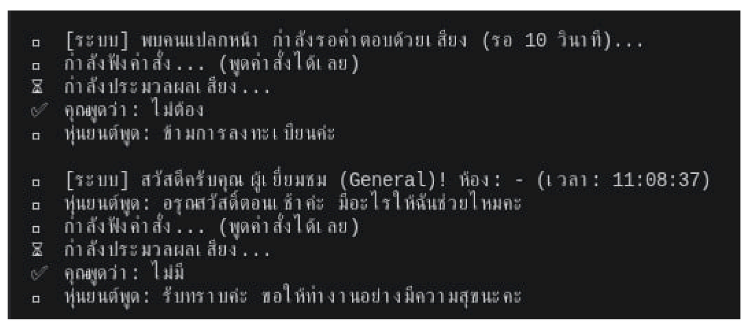

Figure 3 : Voice Interaction using NLP to analyze user intent

3|RESULT

The assessment of system performance was conducted across two primary components. In the Face Recognition segment, tests involving 10 participants were carried out, comparing datasets of 30, 60, and 90 images to evaluate the system's accuracy in recognizing registered individuals

Dataset Size	Correct	Incorrect	Unknown	Accuracy (%)
30 Images	21	2	27	42
60 Images	39	1	10	78
90 Images	48	0	2	96

Figure 4 : Table showing system accuracy across different dataset sizes.

Participant	Gender	Correct	Incorrect	Unknown
Subject 1	Female	5	-	-
Subject 2	Female	5	-	-
Subject 3	Female	5	-	-
Subject 4	Female	4	-	1
Subject 5	Female	4	-	1
Subject 6	Male	5	-	-
Subject 7	Male	5	-	-
Subject 8	Male	5	-	-
Subject 9	Male	5	-	-
Subject 10	Male	5	-	-
Total		48	0	2

Figure 5 : Table showing detailed recognition results for the 90-image dataset.

The findings indicate that the dataset comprising 90 images achieved the highest accuracy and minimized "Unknown" results, thanks to improved feature learning. However, the system's performance remains sensitive to factors such as lighting conditions, facial angles, and distance from the camera. Additionally, potential misidentifications can arise among individuals with similar facial features..

For the Voice Interaction component, the evaluation occurred in two phases: Speech Model Comparison aimed at assessing speech-to-text (STT) accuracy and latency, and Command Performance Evaluation focused on measuring the efficiency of intent classification and entity extraction. This ensures that the robot accurately interprets user commands and retrieves essential data for effective task completion.

MODEL	PERCENTAGE OF SENTENCE ACCURACY	PERCENTAGE OF KEYWORD ACCURACY	AVERAGE RESPONSE
VOSK	85 %	97.5 %	2.015 s
GROQ	80 %	85 %	1 m 16 s
GOOGLE	92.5 %	100 %	3 m 32 s

Figure 6 : Table showing the results of Speech Model Comparison

COMMANDS	PERCENTAGE OF KEYWORD ACCURACY	AVERAGE RESPONSE
Time	100 %	1.85 s
Temperature	100 %	2.01 s
Room Location Inquiry	80 %	2.02 s
Person Location Inquiry	70 %	2 s
Delivery Simulation	80 %	1.95 s

Figure 7 : Table showing the results of Command Performance Evaluation

The results revealed that Vosk is the optimal real-time model, delivering a remarkable 97.5% keyword accuracy with a response time of 2.015 seconds. Basic commands achieved 100% accuracy, while more complex tasks (such as inquiries and simulations) averaged between 70-80% due to challenges related to vocabulary and response latency across all functions.

In the delivery simulation, the integration of face recognition and voice interaction was assessed through four different scenarios to evaluate verification accuracy and system decision-making capabilities.

SCENARIOS	DETAILS	EXPECTED RESULT			
		Voice Interaction	Database	Face recognition	Delivery status
1	Normal Case	Valid	Matched	Verified	Delivered
2	Invalid Recipient	Valid	Unmatched	Not required	Rejected
3	Face not matched	Valid	Matched	Unverified	Pending
4	Voice Interaction Fail	Invalid	Not required	Not required	Rejected

Figure 8 : Table showing the scenarios in Delivery Simulation

In Scenario 1, the system achieves a "Delivered" status when the voice input, database information, and facial data are fully aligned, including room-based deliveries. Meanwhile, in Scenarios 2 and 4, the system triggers a "Rejected" status upon detecting unmatched data or voice input errors to preemptively prevent delivery failures. Scenario 3 results in a "Pending" status if facial verification fails, even when the initial inputs are valid. These findings underscore the effectiveness of the multi-level verification logic in promoting accurate and secure decision-making.

Note: The results presented in the table represent the Expected Result based on the system's decision-making logic design. However, testing with real users to determine the Actual Result has not yet been conducted at this stage of logic development.

4|CONCLUSION

Conclusion The development of this Smart Communication System for service delivery robots showcases the effective integration of facial recognition and voice interaction in delivery simulations. The system successfully processes voice commands, verifies database records, and authenticates recipient identities based on established criteria. Experimental results demonstrate a high level of reliability, underscoring the system's potential to enhance security and convenience in the future of automated delivery. Ultimately, this framework lays a scalable foundation for future robotic applications.

REFERENCES

[1] Personal Identification Using Embedded RaspberryPi-BasedFace RecognitionSystems (Jan,2025)
[2] An Automatic Face Recognition Using Local Binary Pattern Histogram (LBPH) Algorithm (Nov,2023)
[3] Voice Control Device using RaspberryPi (Feb,2019)
[4] Comparative Analysis of Vosk Toolkit and OtherSpeech RecognitionFrameworks for Custom Language Model Implementation (May,2025)