MEKATRONIKA
JOURNAL OF INTELLIGENT MANUFACTURING AND MECHATRONICS
E-ISSN: 2637-0883
VOLUME 7, ISSUE 1, JANUARY, 2025, 19 - 28
DOI: https://doi.org/10.15282/mekatronika.v7i1.11801

RESEARCH ARTICLE

Development Algorithm and Assessing the Efficacy of Voice Control Robotic Prosthetic Hand

David Hao Lau Kin and Sheikh Muhammad Hafiz Fahami Zainal*

Faculty of Manufacturing and Mechatronic Engineering Technology, Universiti Malaysia Pahang Al-Sultan Abdullah, 26600 Pekan, Pahang, Malaysia

ABSTRACT - This research investigates the development and evaluation of a voice-controlled robotic prosthetic hand, offering a potentially more intuitive and user-friendly interface for individuals with upper limb differences. The system utilizes a Raspberry Pi as the central processing unit, leveraging the Python Speech Recognition library and the Google Cloud Speech API for speech-to-text conversion and command recognition. Five servo motors, controlled via a PCA9685 driver board, actuate the prosthetic hand's fingers, mimicking essential grasping and individual finger movements. The performance of the system was assessed through rigorous testing with three participants, focusing on metrics such as word recognition accuracy, command success rate, and overall system latency. Results demonstrated high recognition accuracy, exceeding 97% across all participants, confirming the effectiveness of the chosen speech recognition engine. Command success rates were also consistently high, indicating reliable translation of spoken commands into the intended hand movements. However, the "grip" command presented challenges due to phonetic similarities with other words, highlighting the need for further optimization in speech recognition. Analysis of the system latency revealed that audio capture and processing time on the Raspberry Pi was the dominant contributor to overall delay, suggesting potential benefits from exploring local speech recognition methods. The servo motor performance was consistently fast and accurate, confirming the viability of the mechanical design and control strategy. This research successfully demonstrates the feasibility of voice control for robotic prosthetic hands, providing a foundation for future development and highlighting the importance of addressing pronunciation variability, optimizing latency, and incorporating user feedback for improved usability.

ARTICLE HISTORY

Received	:	20th Oct. 2024
Revised	:	17th Dec. 2024
Accepted	:	26th Dec. 2024
Published	:	06th Jan. 2025

KEYWORDS

Voice control
Prosthetic hand
Raspberry Pi
Google Cloud API
Speech recognition

1. INTRODUCTION

The field of prosthetics has long sought to restore functionality and improve the quality of life for individuals with limb loss. Traditional prosthetic devices, such as body-powered and myoelectric prostheses, have provided valuable solutions but often present limitations, especially for individuals with higher-level amputations [1][8]. Body-powered systems rely on manual manipulation through cables and harnesses, while myoelectric control relies on sufficient muscle activity to generate signals, both posing challenges for certain users. Furthermore, these conventional methods can be costly, further limiting accessibility [2][9-10]. The integration of robotics into prosthetics has opened up new possibilities, but the development of more intuitive and user-friendly control methods remains an active area of research. Voice control, with its increasing prevalence and success in diverse applications [3][13-14], emerges as a promising alternative, potentially addressing the limitations of traditional approaches [4][11-12]. However, the application of voice control in prosthetic hands remains relatively underexplored [5-9].

This research aims to bridge this gap by developing and evaluating a voice-controlled robotic prosthetic hand prototype. The central challenge lies in accurately interpreting voice commands and translating them into precise and reliable hand movements, a complex task involving technological, physiological, and usability considerations. Through a systematic investigation using a cost-effective prototype and a small participant group, this study seeks to demonstrate the feasibility and efficacy of voice control as a viable and user-friendly alternative for prosthetic hand manipulation. The remainder of this paper is organized as follows: Section 2 briefly explains the various methodologies employed in voice-controlled prosthetic hand systems. Section 3 details implementation of the voice-controlled prosthetic hand developed in this research, focusing on the utilization of the Google Cloud Speech API. Section 4 presents the data collection methods and performance evaluation metrics, followed by a detailed analysis of the results. Section 5 summarizes the key findings and concludes the paper.

***CORRESPONDING AUTHOR | Sheikh Muhammad Hafiz Fahami Zainal** | ✉ hafizfahami@umpsa.edu.my

1.1 Related Work

Voice control has emerged as a promising alternative to traditional prosthetic control methods, offering the potential for greater intuitiveness and accessibility. Numerous researchers have explored various approaches to implementing voice-controlled prosthetic hands, employing different speech recognition technologies and control strategies. This section reviews notable research in this field, highlighting key methodologies and comparing their strengths and limitations. Several studies have utilized Voice Recognition Modules (VRMs) as the core of their speech recognition systems. VRMs are typically microcontroller-based modules (Arduino) that provide basic speech recognition capabilities. Samant [5] and Oppus [6] implemented prosthetic hand control using VRMs, demonstrating the feasibility of this approach. However, VRMs often rely on simpler recognition algorithms and may have limitations in accuracy, particularly in handling variations in pronunciation or background noise. Jafarzadeh & Tadesse [7] explored a different approach, utilizing Mel-Frequency Cepstral Coefficients (MFCCs) as features for speech recognition. MFCCs are widely used in traditional speech recognition systems and represent the spectral envelope of a sound. While effective, MFCC-based approaches might not be as robust as deep learning methods, especially in challenging acoustic environments or when dealing with a diverse range of speakers. Abdul-Nafa [8] integrated an Internet of Things (IoT) platform with Google Assistant for voice control of a prosthetic hand. Google Assistant employs advanced deep learning models for speech recognition, similar to the Google Cloud Speech API used in this study. However, their implementation focused on remote control through an IoT network, which could introduce latency issues. This research leverages the Google Cloud Speech API for voice command recognition. This API utilizes state-of-the-art deep learning models trained on vast amounts of data, offering high accuracy and robustness to variations in pronunciation and background noise. The cloud-based nature of the API provides access to powerful computational resources and continuous updates from Google, ensuring the system benefits from the latest advancements in speech recognition technology. However, the reliance on cloud processing introduces potential latency due to network communication, a factor that requires careful consideration in real-time control applications.

2. METHODOLOGY

The system uses a Raspberry Pi 4 Model B that acts as the central processing unit and the orchestrator of the prosthetic hand's movements as shown in Figure 1. It receives audio input from a USB microphone and the captured audio is then transmitted to the Google Cloud Speech API, a powerful cloud-based speech recognition service, for conversion into text. The transcribed text is then analysed and interpreted by the Raspberry Pi to identify specific voice commands. These commands are mapped to predefined hand gestures, allowing the user to control the prosthetic hand through spoken instructions.

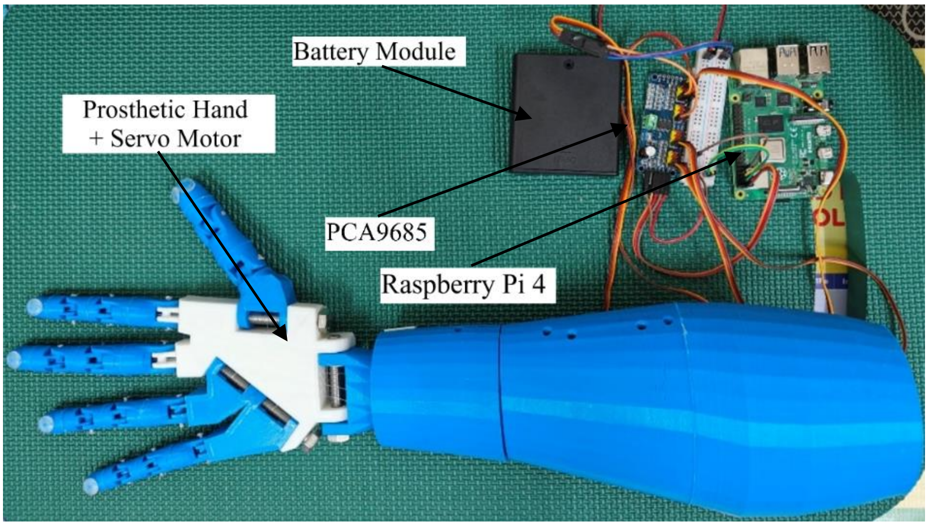


Figure 1. Prototype prosthetic-speech recognition

The core algorithm of the voice control system is summarized in the flowchart as shown in Figure 2. In summary, the process begins with capturing an audio input, which is then processed and converted to text using the Google Cloud Speech API. The system then verifies whether the converted command is valid. If the command is invalid, the user is prompted to try again. If valid, a motor control signal is generated, leading to the actuation of servo motors. The process then terminates after successful execution. This flow ensures accurate voice recognition and controlled motor response, making it suitable for applications like prosthetic hand control or automated systems.

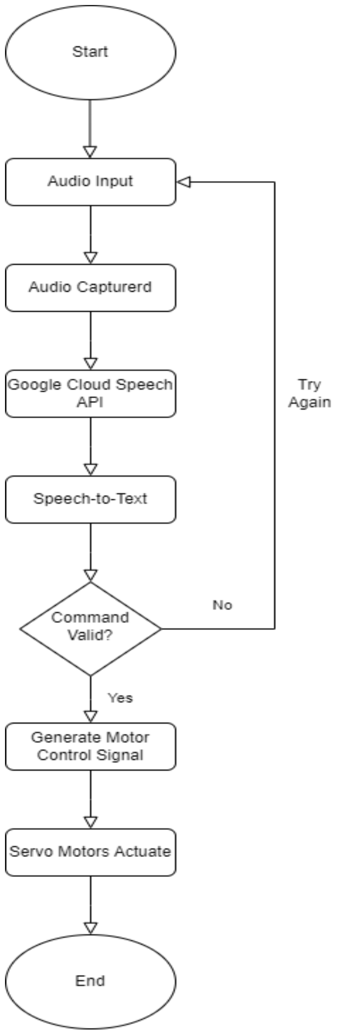


Figure 2. Flowchart of the process voice recognition

2.1 Hardware Components

The Raspberry Pi 4 Model B as shown in Figure 3 is a powerful single-board computer chosen for its processing capabilities, versatile connectivity options, and compact form factor. It serves as the brain of the system, running the Python code responsible for audio capture, speech recognition, command interpretation, and servo control. The PCA9685 as shown in Figure 4 is an I2C-controlled PWM driver board specifically designed for controlling servo motors. It expands the Raspberry Pi's limited (only 4) PWM output channels, providing 16 individual channels for precise and independent control of each servo motor in the prosthetic hand. Five Tower Pro MG996R servo motors as shown in Figure 5 are employed to actuate the fingers of the prosthetic hand. These servos provide a good balance of torque, speed, and affordability, making them suitable for this application. Their wide rotational range (180 degrees) allows for a variety of hand gestures.

Figure 3. Raspberry Pi 4 model B

Figure 4. PCA9685

The prosthetic hand itself is a 3D-printed model adapted from the open-source InMoov project. Its design was chosen for its articulation, providing a good range of motion for the fingers. The hand utilizes a tendon-driven mechanism, with strong and flexible nylon fishing lines acting as tendons to control finger movements. The tendons are attached to the servo motors through the servo horns, which pull on the lines to create the desired hand gestures.

Figure 5. Tower Pro MG996R Servo Motor

2.1 Software Implementation

The software controlling the system is written in Python, a versatile and widely used programming language well-suited for robotics and automation projects. The program leverages the Speech Recognition library to interface with the microphone and send audio data to the Google Cloud Speech API for transcription. The program operates within a continuous loop that listens for spoken commands. Once a voice command is recognized, the program analyses the transcribed text to identify the corresponding hand gesture. The command interpretation logic is implemented using a series of if-else statements, comparing the recognized text against a predefined list of valid commands. Each command is mapped to a specific set of servo motor positions, ensuring that the prosthetic hand performs the correct gesture.

2.1.1 Google Cloud Speech API Integration

The Google Cloud Speech API as shown in Figure 6 plays a crucial role in the system by providing powerful cloud-based speech-to-text conversion capabilities. The API uses advanced deep learning models, trained on vast datasets, to transcribe the captured audio into text even in the presence of background noise or variations in pronunciation. This robustness is crucial for a prosthetic hand control system, as users may have different speech patterns or use the device in diverse environments. This allows the system to interpret the user's spoken commands and translate them into prosthetic hand movements.

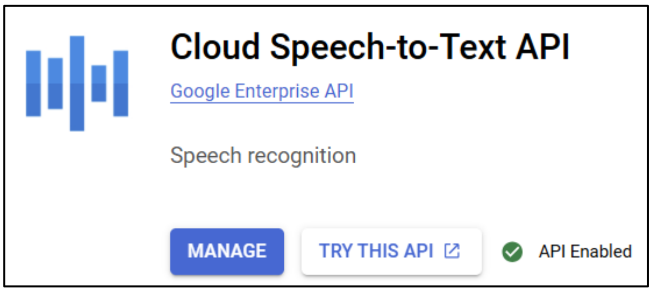


Figure 6. Google Cloud Speech-to-Text API enabled in cloud console

3. EXPERIMENTAL RESULT

3.1 Data Collection Methods

Three participants were recruited for the evaluation. Each participant was asked to utter a set of seven predefined voice commands fifteen times each with a total of 315 trials across the three participants. The commands chosen represented a range of common hand gestures: "grip," "relax," "pinky," "ring," "middle," "index," "ok" as shown in Table 1. These commands were selected for their relevance to basic hand functionality and their distinctiveness in terms of pronunciation to minimize potential recognition errors.

The system automatically logged data for each trial to a CSV file as shown in Table 2. The logged data included are written as follows:

a) Recognized Text: The text transcribed by the Google Cloud Speech API.
b) Command: The intended command spoken by the participant.
c) Recognition Time: The time taken for audio capture and transmission to the API.
d) Google Recognition Time: The time taken for the API to process the audio and return the recognized text.
e) Servo Response Time: The time taken for the servo motors to complete the intended hand movement.

Table 1. Visual representations of the commands

Command	Gestures	Command	Gestures
(a) Grip		(e) Middle	
(b) Relax		(f) Index	
(c) Pinky		(g) Ok	
(d) Ring			

Table 2. Example of raw data logged in CSV file

Trial	Recognized Text	Command	Recognition Time (ms)	Google Recognition Time (ms)	Servo Response Time (ms)
1	grip	grip	1953.62	673.16	7.04
2	relax	relax	1484.06	480.86	3.51
3	grape	Not Recognized	2295.25	698.96	0
4	middle	middle	1675.91	480.38	0.71
5	ring	ring	1548.74	651.08	1.51
6	pinky	pinky	844.28	344.15	0.71
7	pancake	Not Recognized	1335.27	614.33	0
8	index	index	1420.66	555.42	0.76
9	ok	ok	737.74	441.31	1.34
10	ok	ok	1441.25	480.76	3.32

3.2 Performance Evaluation Metrics

The description performance of evaluation metrics is written as follows:

a) Recognition Accuracy: This metric measures the percentage of spoken words correctly recognized by the Google Cloud Speech API, reflecting the accuracy of the speech-to-text engine.

b) Command Success Rate: This metric evaluates the percentage of spoken commands that resulted in the correct hand movement, indicating the system's ability to interpret commands and control the prosthetic hand reliably.

c) Servo Motor Performance: This metric assesses the responsiveness of the servo motors by measuring the time it takes for a servo to reach its target position after receiving a command. It reflects the speed and efficiency of the hand's actuation system.

d) Overall System Latency: This metric represents the total time elapsed between the participant speaking a command and the completion of the intended hand movement. It encompasses the combined latency of speech recognition, command processing, and servo actuation.

3.3 Analysis

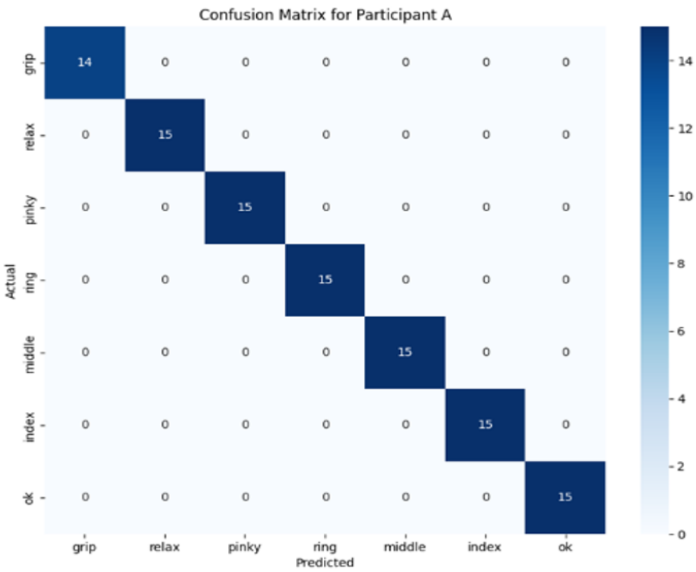
Assessing the accuracy of the speech recognition system is paramount to ensuring the reliability and usability of the voice-controlled prosthetic hand. To evaluate recognition accuracy, confusion matrices were generated for each participant, visually representing the system's ability to correctly classify spoken commands as shown in Figure 4. The overall word recognition accuracy for Participant A as shown in Figure 4(a) was calculated by dividing the sum of the diagonal elements of the confusion matrix (representing correctly recognized commands) by the total number of trials. This calculation resulted in an overall accuracy of 99%, indicating that the system correctly recognized 104 out of 105 spoken commands for this participant. An analysis of the confusion matrix reveals a very high recognition accuracy for Participant A. The system correctly classified all spoken commands except for a single instance where "grip" was misrecognized as "grape." This minor misclassification likely occurred due to the phonetic similarity between the two words. Moreover, could be due to sensor noise, feature overlap, or slight inconsistencies in gesture execution. Despite this, the algorithm model demonstrates high reliability, making it well-suited for prosthetic hand control with minimal errors.

While, for participant B as shown in Figure 4(b), the overall word recognition accuracy is 96.2%, with 101 out of 105 commands correctly recognized. While still high, this accuracy is slightly lower than that of Participant A (99.0%). Analyzing the confusion matrix, we observe that the "grip" and "pinky" commands were the most challenging for the system, each with two misclassifications. Interestingly, both misclassifications of "grip" were transcribed as phonetically similar words, "ripe" and "drip," suggesting that the speaker's pronunciation of "grip" might have contributed to these errors. Similarly, both instances of "pinky" were misrecognized as "pancake," indicating a potential consistency in the speaker's articulation of "pinky" that led to this specific misclassification. These findings highlight the importance of considering individual speaking styles and potential pronunciation variations when evaluating the performance of a voice recognition system.

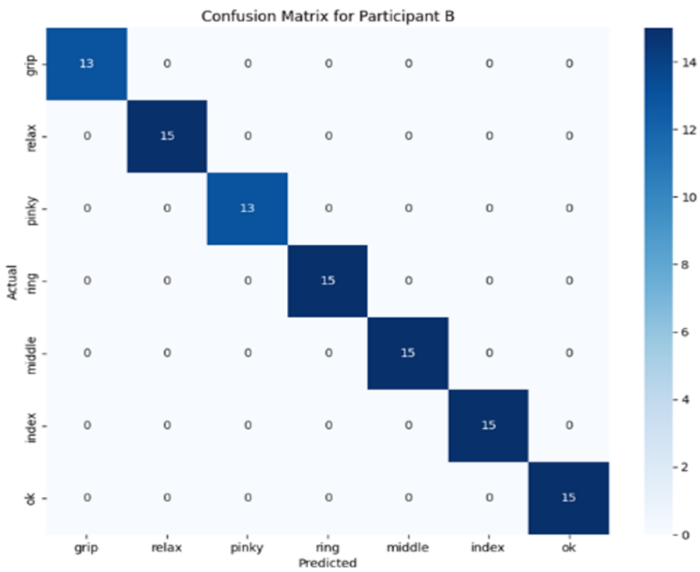
The overall word recognition accuracy for Participant C was calculated to be 98 %, with 103 out of 105 commands correctly recognized. This high accuracy signifies the system's capability to effectively transcribe spoken commands for this participant. Examining the confusion matrix reveals that only two commands posed a challenge for the recognition system: "grip" and "index." The single misclassification for "grip" was transcribed as "creep," likely due to the speaker's pronunciation. Similarly, "index" was once misrecognized as "Intex," again suggesting a potential influence of the speaker's articulation. These findings further emphasize the impact of individual speaking styles on voice recognition accuracy. These matrices revealed a consistently high level of accuracy across participants, with an average of 98 % of commands recognized correctly. While most commands were consistently recognized with high precision, the "grip" command presented occasional challenges, being misclassified as "grape" or "creep". This suggests a potential need for further optimization of the speech recognition model to account for phonetic similarities between words. Comparing this accuracy to existing research as shown in Table 3.

Table 3. Comparison of recognition accuracy with existing research

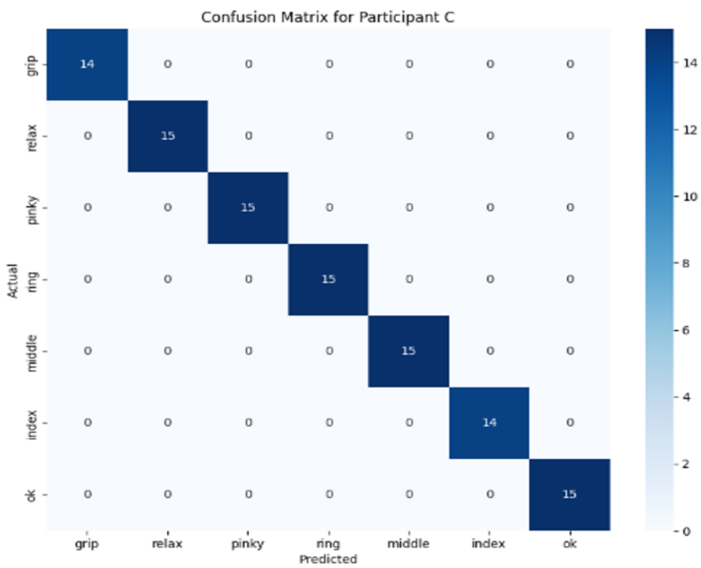
Source	Method	Accuracy (%)
Samant & Agarwal [5]	Voice Recognition Module	88.8
Oppus et al. [6]	Voice Recognition Module	89.6
Jafarzadeh & Tadesse [7]	Mel-Frequency Cepstral Coefficients	88.1
Abdul-Nafa et al. [8]	IoT with Google Assistant	97.0
This Research	Google Cloud Speech API	97.8



(a)



(b)



(c)

Figure 4. Confusion matrix of the commands across (a) Participants A, (b) Participants B and (c) Participants C

Its demonstrates that the Google Cloud Speech API employed in this research achieved higher recognition accuracy than other voice-controlled prosthetic hand systems, highlighting the effectiveness of this approach. This high recognition accuracy translated into similarly strong command success rates, with most commands achieving near-perfect execution across participants. However, the occasional misclassifications of the “grip” command did contribute to a slightly lower success rate for this gesture. The Figure 5 presents a comparison of command success rates among three participants in a study evaluating the effectiveness of a voice as proposed method.

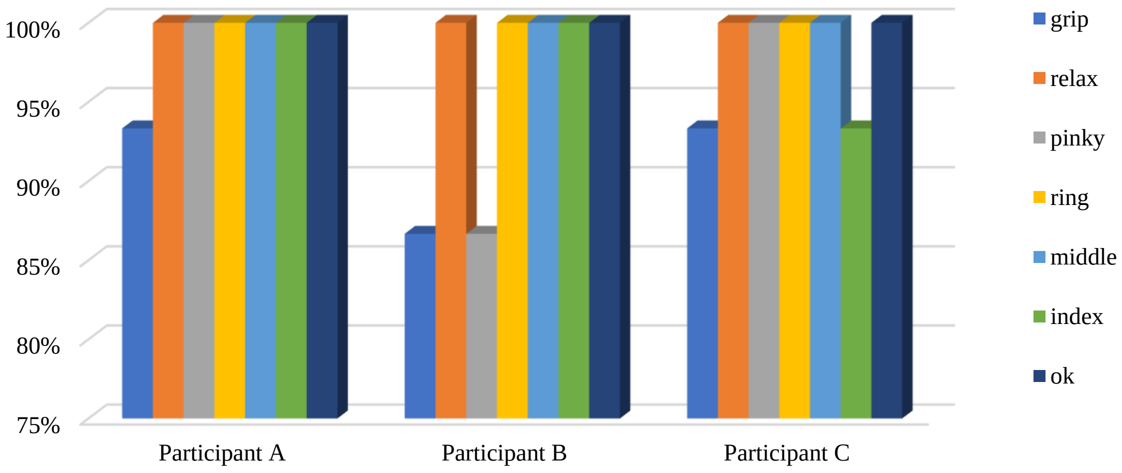


Figure 5. Comparison of command success rates across three participants

Whereby each of the bar represents the success rate of different voice commands ("grip," "relax," "pinky," "ring," "middle," "index," and "ok") for each participant. The results indicate consistently high success rates, mostly above 90%, with commands like relax, ring, middle, and ok achieving near-perfect accuracy. However, minor variations exist, with grip, pinky and index showing slightly lower success rates for some participants. These variations may be due to differences in speech recognition, user articulation, or system responsiveness. Therefore, its suggesting potential areas for refinement in both user training and system optimization.

While, the Figure 6 presents the average servo motor response time in milliseconds (ms) for different voice commands used to control a prosthetic hand across three participants. The results indicate that the "grip" and "relax" commands have the highest response times, exceeding 4 ms, while other commands such as "pinky," "ring," "middle," "index," and "ok" show significantly lower response times, generally below 1.5 ms. This suggests that more complex or multi-finger movements, like gripping and relaxing, require longer execution times compared to individual finger movements. Variations in response times among participants may be attributed to differences in speech input, system processing, or servo performance itself which is not taken consider in these studies.

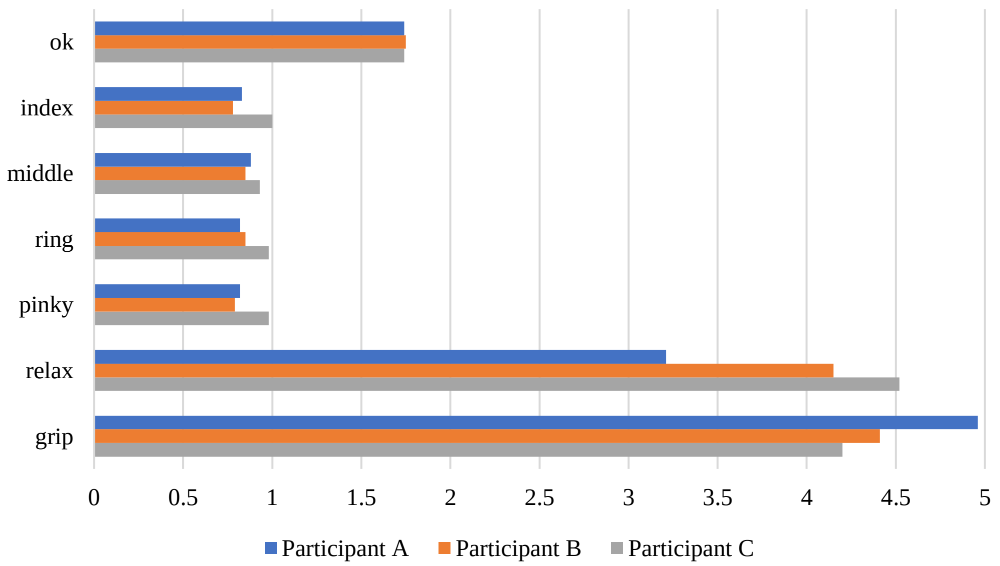


Figure 6. Average servo response time (ms)

Meanwhile, the Figure. 7 illustrates the overall system latency for different voice commands. The latency includes the time taken for voice recognition, processing, and servo actuation. The results show that commands like grip, ok and relax exhibit the highest latencies, exceeding 2000 ms, particularly for Participant C, suggesting variations in system response times among users. Other commands, such as index, middle, ring, and pinky generally show lower latencies, staying below 2000 ms. The observed variations may stem from differences in speech articulation, system processing efficiency, or mechanical execution delays.

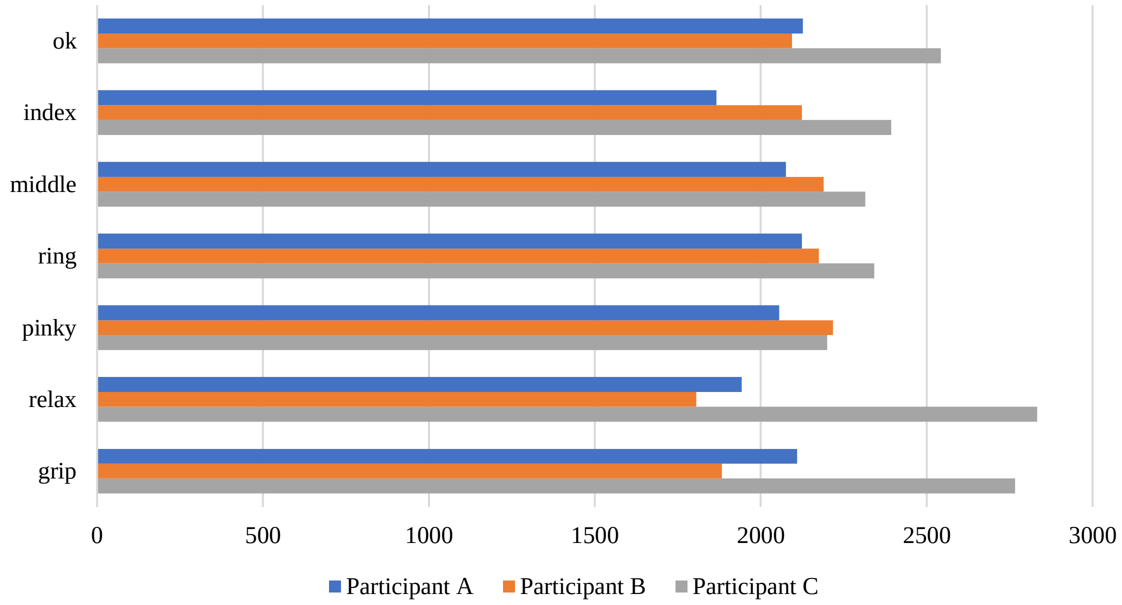


Figure 7. Overall system latency (ms)

4. CONCLUSION

This research successfully developed and evaluated a voice-controlled robotic prosthetic hand, demonstrating the feasibility and potential of this innovative approach to prosthetic control. The system, based on a Raspberry Pi, a USB microphone, and the Google Cloud Speech API, achieved high recognition accuracy (averaging 98.7% across three participants) and fast servo motor response times. While the average system latency (1806 ms to 2833 ms) highlights the need for further optimization, particularly in reducing initial audio processing time, the results provide strong evidence that voice control can be a viable and user-friendly method for controlling prosthetic hands. The system's accuracy, exceeding that of existing systems, showcases the effectiveness of the Google Cloud Speech API. Despite the limitations of a limited command set and a controlled testing environment, the findings have significant implications for the future of prosthetic technology. Developing more intuitive and accessible control methods like voice control holds the potential to significantly improve the quality of life for individuals with limb differences, empowering them with greater autonomy and independence. Future work should address the identified limitations by exploring local speech recognition for reduced latency, expanding the command set, and conducting more extensive user testing in real-world scenarios to further evaluate the system's usability and practicality.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to Universiti Malaysia Pahang Al-Sultan Abdullah, for the support on this research project.

REFERENCES

[1] S. M. Engdahl, D. H. Gates, "Differences in quality of movements made with body-powered and myoelectric prostheses during activities of daily living," *Clinical Biomechanics*, vol. 84, p. 105311, 2021.

[2] P. S. Keerthi, S. Al Mamun, R. A. Tonima, S. A. Shanta, A. Ahmed, M. S. R. Zishan, "Design and implementation of a human prosthetic hand," In *2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*, pp. 640-643 2021.

[3] A. Singh, S. Srivastava, K. Kumar, S. Imran, M. Kaur, N. Rakesh, et al., "Iot-based voice-controlled automation," In *International Conference on Innovative Computing and Communications: Proceedings of ICICC 2021*, vol. 1, pp. 827-837, 2022.

[4] J. Pradeep, A. Jamna, R. Sasikumar, "Low-cost voice-controlled prosthetic arm with five degrees of freedom," *IETE Journal of Research*, vol. 69, no. 7, pp. 4047–4052, 2023.

[5] P. Samant, R. G. Agarwal, "Real-time speech recognition system for prosthetic arm control," Master dissertation, Thapar University, Patiala, India, 2014.

[6] C. M. Oppus, J. R. R. Prado, J. C. Escobar, J. A. G. Mariñas, R. S. J. Reyes, "Brain-computer interface and voice-controlled 3D printed prosthetic hand," In *2016 IEEE Region 10 Conference (TENCON)*, pp. 2689-2693, 2016.

[7] M. Jafarzadeh, Y. Tadesse, "End-to-end learning of speech 2D feature-trajectory for prosthetic hands," In *2020 Second International Conference on Transdisciplinary AI (TransAI)*, pp. 25-33, 2020.

[8] A. D. Abdul-nafa, S. S. Ahmed, Y. S. Ahmed, O. M. Alsaydia, "Control robotic hand depending on voice commands and IoT," *Nano Biomedicine and Engineering*, vol. 15, no. 3, pp. 288–295, 2023.

[9] M. H. M. Ariff, H. Zamzuri, N. Zulkarnain, M. H. C. Hasan, S. M. H. Fahami, "Robust composite nonlinear feedback for nonlinear steer-by-wire vehicle yaw control," *Bulletin of Electrical Engineering and Informatics*, vol. 8, no. 1, pp. 214-222, 2019.

[10] X. Meng, Y. Song, "Application research and implementation of voice control system based on android speech recognition," *Journal of Physics: Conference Series*, vol. 1865, no. 4, p. 042122, 2021.

[11] L. Bai, "Research on voice control technology for smart home system," *Proceedings of the Asia Conference on Electrical, Power and Computer Engineering*, pp. 1-7, 2022.

[12] MD Vu, H Wang, Z Li, G Haffari, Z Xing, C. Chen, "Voicify your UI: Towards android app control with voice commands," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 7, no. 1, pp. 1-22, 2023.

[13] S. A. Li, Y. Y. Liu, Y. C. Chen, H. M. Feng, P. K. Shen, Y. C. Wu, "Voice interaction recognition design in real-life scenario mobile robot applications," *Applied Sciences*, vol. 13, no. 5, p. 3359, 2023.

[14] R. Baruah, S. Mondal, B. Das, "A machine learning based approach to control device using recognition system," *Journal of Emerging Technologies and Innovative Research*, vol. 11, no. 4, pp. 96-104, 2024.

[15] X. Chen, "Research on voice control and analysis technology for the elderly at home" *Proceedings of the 2nd International Conference on Intelligent Design and Innovative Technology (ICIDIT)*, pp. 152-162, 2023.