International Journal of Autonomous Robotics and Intelligent Systems
Vol. 2, No. 1, June 2026 [53-62]

Design and Development of a Voice-Controlled Humanoid Robotic Mouth for Facial Expression Simulation

Rudzuan M. Nor[1,2,3*], H. Luqmanul Hakim[1], Y. Sofi[1,2], S. Nordiana[1,2], W.M. Nooriman[1,2] and S.A. Abdul Shukor[1,3]

[1]Mechatronic Department, Faculty of Electrical Engineering & Technology, UniMAP
[2]AI & Automation RG, Faculty of Electrical Engineering & Technology, UniMAP
[3]Center of Excellence for Intelligent Robotics and Autonomous Systems (CIRAS), UniMAP

Received 10 May 2026, Revised 22 May 2026, Accepted 21 June 2026

ABSTRACT

Human facial expression plays an important role in communication between humans and robots. This paper presents the design and development of a voice-controlled humanoid robotic mouth capable of generating lip and mouth expressions to enhance human-robot interaction. The proposed system integrates mechanical design, electronic control and embedded programming using an Arduino Uno microcontroller, HC-05 Bluetooth module, and five SG90 servo motors. Voice commands are captured through an Android smartphone application with speech recognition capability and transmitted wirelessly to the robot via Bluetooth. The compact robotic mouth, measuring 130 x 131 x 83 mm, was fabricated using 3D-printed Polylactic Acid (PLA) material and designed with five degrees of freedom to imitate natural lip movements. Four basic facial expressions, namely relax, smile, sad and surprise, were successfully implemented by assigning specific servo angle combinations. Experimental results demonstrated that the robotic mouth was capable of producing distinguishable and visually recognizable facial expressions with stable mechanical operation. Bluetooth communication remained stable up to approximately 8 meters under test conditions. The developed prototype demonstrates the feasibility of combining low-cost components and voice recognition technology to create an expressive robotic interface for future humanoid robot applications.

Keywords: Humanoid Robot, Facial Expression, Voice Control, Bluetooth Communication

1. INTRODUCTION

Social robots are gaining importance due to technological advancements and increasing human intelligence, leading to heightened interest in human-robot interaction (HRI) research. Humanoid robots, which closely resemble human anatomy, serve as the primary subjects for HRI development. These robots are continuously being enhanced to mimic human behaviour, such as running, making facial gestures and speaking. A survey revealed a significant increase in publications related to social robots from 2000 to 2020.

Humanoid robots find applications in various aspects of daily life, including healthcare, research, entertainment and demonstrations, making facial expression capabilities crucial. Human perception of HRI transparency improves when robots exhibit emotions, an essential element in human-robot relationships. For effective interaction, robots should dynamically change their emotions based on external inputs and adapt their behaviour accordingly. Notable robots like KOBIAN-R and GolemX-1 already showcase these capabilities.

*rudzuan@unimap.edu.my

In this study, a simple humanoid robot mouth with connected lips was utilized to perform the four common facial expressions: relax, smile, sad and surprise. The robot incorporated voice recognition input to produce precise expressive output. The study examines the mechanical, electrical and programming aspects and documents the observed outcomes. To contribute to this evolving field, this study aims to develop a lightweight, budget-friendly 5-DOF robotic mouth assembly that minimizes mechanical complexity while maintaining recognizable and realistic human expressions. Currently, the most popular robots are specialized for particular tasks like doing laundries, i.e. automatic washing machines, and automatic vacuum cleaners. However, these robots lack the ability to understand and respond to human emotions and empathy. To bridge this gap, developers are now focusing on creating robots that can comprehend human gestures, speech and facial expressions, enabling more natural interactions.

These human-like communication skills make such robots valuable for various applications, including home companions, nursing care services, storytelling, live advertising and interactive machines. Building a realistic humanoid robot poses challenges, especially in replicating complex human anatomy, leading to awkward movements in mechanical parts. This phenomenon, known as the “uncanny valley”, causes discomfort and uneasiness when humanoid objects almost resemble humans but not quite.

Despite the challenges in developing realistic robotic facial systems, humanoid robots with expressive mouth and lip movements offer significant advantages in applications such as interactive robotics, education, storytelling and human-robot communication. The proposed research focuses on the development of a low-cost, compact and voice-controlled humanoid robotic mouth capable of generating basic facial expressions for more natural human-robot interaction. Unlike complex humanoid facial systems that require a large number of actuators and sophisticated control mechanisms, the developed prototype utilises only five servo motors with five degrees of freedom (5-DOF) to simulate human-like lip and mouth movements. The system was designed using commercially available electronic components and fabricated using 3D-printed Polylactic Acid (PLA) material to achieve a lightweight structure with simplified manufacturing and assembly processes. The proposed design demonstrates the feasibility of producing recognizable facial expressions using a simplified mechanical structure and embedded control system.

1.1 Literature Review

In 2022, R. Stock-Homburg published a survey focusing on emotions in human-robot interactions. The survey identified six primary emotions suggested in the facial action coding system (FACS). The study encompassed various robots and emotional expression modes, including facial, as shown in Figure 1, bodily, and both. The findings revealed that humans can correctly recognize approximately 50% of a robot's emotions on average. Notably, for high arousal emotions like happiness and anger, the average recognition rates were even higher [1]. Meanwhile, T. Kishi et al. conducted a survey involving a walking humanoid robot named KOBIAN-R, which possesses 65 degrees of freedom (DOF). The robot has the ability to dynamically change its emotional state while carrying out actions in response to external stimuli. It incorporates visual tracking with eyes, head and upper body, as well as bipedal locomotion, emotion expression through facial patterns, and whole-body movements [2].

Similarly, X. Ke et al. developed a humanoid head robot called SHFR-III with 22-DOF. The robot is controlled using a field-programmable gate array (FPGA) and utilizes servos to actuate facial movements. It is equipped with a human-computer interactions interface featuring four types of work patterns, including facial expression recognition, face recognition, gender recognition and expression representations [3].

Figure 1. Variety of existing robotic faces [1]

Moreover, G. Trovato et al. have developed a facial expression generator specifically designed for the new 24-DOF head of the humanoid robot KOBIAN-R. This generator can produce a vast array of over 600 thousand facial cue combinations, including appropriate neck movements. The recognition rate of expressions produced by their system is comparable to that of the most common facial expressions [4]. Additionally, T. Kishi et al. have worked on a new head for a similar biped walking robot, KOBIAN-R. This robot head is capable of expressing the six basic emotions while being small and light enough to be mounted on the bipedal walking robot [5]. In 2019, M.E. Reyes et al. created a robotics face called GolemX-1, which features a minimalist design with only 8-DOF. They utilized a 3-D software modeler for its design. The system they developed can express six universal expressions, including an additional one called "worried" on top of neutral, happy, angry, sad, surprise and fear. According to their results, anger is the most easily recognized universal facial expression, and it has a regulatory effect on human actions, capturing human attention in unclear situations during tasks [6].

V. Seib et al. introduced an animated cartoon-like robot face, providing a user-friendly interaction with robots. This robot can execute seven distinct facial expressions, with its lips synchronized to synthesized speech using a text-to-speech system [7]. Similarly, H.-J. Hyung et al. developed a Korean lip-syncing system for robots. The robots can perform real-time lip-syncing for various words and sentences using ten mouth shapes. In a survey where sound was not provided, 80 human subjects guessed the intended words solely based on the visual mouth shape and evaluated the naturalness of the mouth shape and the robot's lip-sync timing. Interestingly, the percentage of correct guesses for the human announcer was higher for women compared to men [8]. Furthermore, Y. Usui et al. designed an anthropomorphic face robot featuring a soft mouth mechanism embedded with artificial facial muscles. This robot's face was built using a soft actuator, a silicone mouth and a dummy skull. They successfully created eight basic facial expressions, including lip puckering and dynamic mouth opening [9]. In addition, Harshani et al. developed humanoid robot faces with moving lips and text-to-speech ability. The authors constructed a robotic model capable of simulating jaw movements. By utilizing audio and text analysis technology, they enhanced accuracy based on phonetic symbols of the English language.

Moreover, they incorporated additional lip features with driver mechanisms to coordinate lip movements with the generated voice [10]. From here, it can be seen the vast increment and importance of facial expression for robot.

2. METHODOLOGY

This section describes the methodology used to design and develop a humanoid robotic mouth capable of expressing emotions through voice commands. The project integrates mechanical design, electronic control and embedded programming into a single prototype system. The methodology was structured into several stages, namely system design, material selection, hardware construction, software development and final system integration.

In general, the overall system operation begins when the controller is powered on and all servo motors are set to their default neutral positions. The robotic mouth then remains in standby mode while waiting for user input. Voice commands are spoken through an Android smartphone application equipped with speech recognition capability. The spoken words are converted into text strings and transmitted wirelessly to the robot using Bluetooth communication. The microcontroller receives the data, interprets the command, and actuates the servo motors to generate the corresponding facial expression. If no valid command is detected, the system continues waiting for a new input.

Several materials and components were selected based on functionality, weight, flexibility and cost-effectiveness. The mechanical structure was fabricated mainly using Polylactic Acid (PLA), which is a lightweight thermoplastic suitable for 3D printing applications. Flexible cable protectors were used to imitate lip motion due to their elastic characteristics. The electronic control system employed an Arduino Uno microcontroller as the main processing unit because it provides sufficient PWM output pins for controlling multiple servo motors. Bluetooth communication was achieved using the HC-05 module and connected using serial communication lines (TX and RX), enabling two-way data transfer between the smartphone and microcontroller. Meanwhile, five SG90 micro servo motors were used as actuators for jaw and lip movements. Power was supplied through a 7.4V LiPo battery, and voltage regulation was performed using an LM2596 buck converter. The block diagram in Figure 2 shows a general overview of the components and their interactions as they execute input reading, data processing and output actuation. Autodesk Fusion 360 software was used for 3D modelling, whereas Arduino IDE was used for software programming. The main hardware and software components used in the development of the humanoid robotic mouth system are summarized in Table 1.

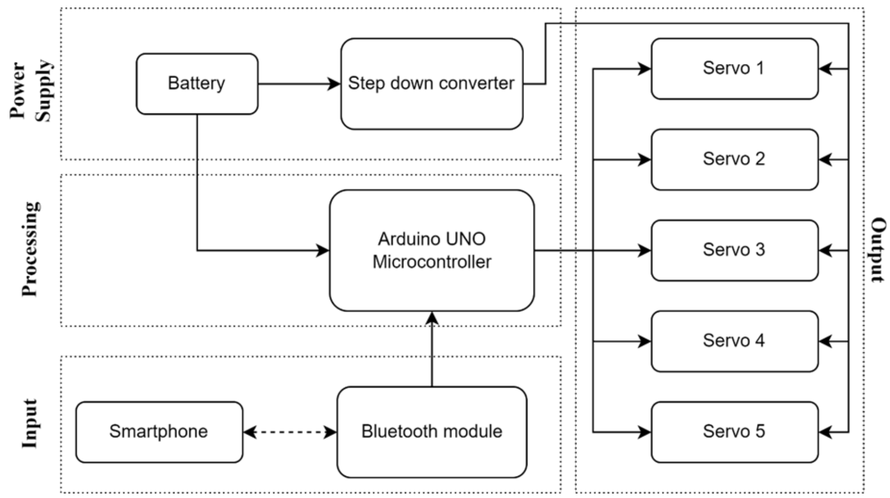
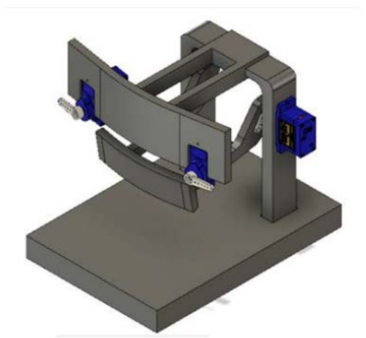


Figure 2. System block diagram

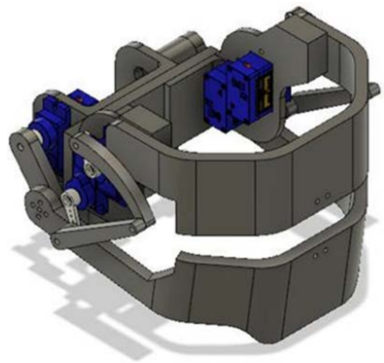
Table 1 Hardware and software specifications

Component	Specification	Function
Microcontroller	Arduino Uno	Main processing and servo control unit
Bluetooth module	HC-05	Wireless voice communication
Servo motor	SG90 Micro Servo	Lip and jaw actuation
Power supply	7.4V LiPo Battery	Main power source
Voltage regulator	LM2596 Buck Converter	Voltage regulation to 5V
Mechanical material	PLA (Polylactic Acid)	3D-printed robotic structure
3D design software	Autodesk Fusion 360	Mechanical modelling and design
Programming software	Arduino IDE	Embedded programming development
Voice recognition platform	AMR_Voice Android application	Speech-to-text voice command input

The mechanical development was conducted progressively through multiple prototype stages to evaluate and improve the facial expression mechanism. The second prototype was redesigned using Autodesk Fusion 360 and fabricated using 3D printing with PLA material. This version utilised 3-DOF, consisting of one jaw servo and two corner lip servos, as shown in Figure 3(a). However, the resulting lip movement appeared limited and less natural due to the restricted actuator movement range. Therefore, an improved final model was developed using five servo motors and additional linkage mechanisms to better imitate the movement of human facial muscles, particularly the zygomaticus muscle for smiling motion and the depressor "anguli oris" muscle for sad expressions. The final 5-DOF robotic mouth structure is shown in Figure 3(b).



(a) 3-DOF prototype



(b) Final 5-DOF prototype

Figure 3. The developed prototype

The enhanced mechanism was intended to improve the natural movement of the robotic mouth while also allowing future expansion such as integration with robotic eyes or a complete humanoid head system.

2.1 Servo Motor Configuration for Facial Expression Simulation

Figure 4 illustrates the final robotic mouth structure together with the servo-driven linkage mechanism used to simulate facial expressions. Five SG90 servo motors were employed in the system, consisting of one jaw servo (J), two upper lip servos (LU and RU) and two lower lip servos (LD and RD). Table 2 summarises the functional assignment of each servo motor used in the robotic mouth system.

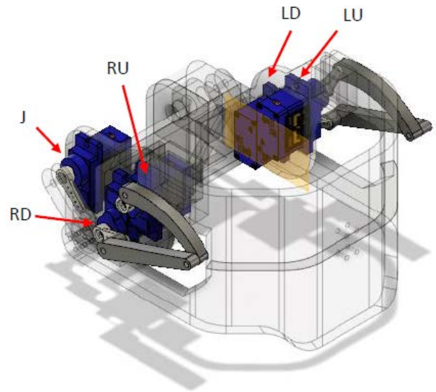


Figure 4. Labelled servomotor

Table 2 Servo motor functional assignment

Servo Label	Function
J	Jaw opening movement
LU	Left upper lip movement
LD	Left lower lip movement
RU	Right upper lip movement
RD	Right lower lip movement

The servo arrangement was designed to coordinate the movement of the upper lip, lower lip and jaw to simulate different emotional states. Various servo angle combinations were experimentally calibrated to produce the required facial expressions while maintaining stable mechanical operation and smoother lip movement. The calibrated servo angle configurations for each emotional expression are presented in Table 3.

Table 3 Servo angle configuration for facial expression simulation

Emotion	Jaw (J)	Left Upper (LU)	Left Lower (LD)	Right Upper (RU)	Right Lower (RD)
Relax	180°	90°	90°	90°	90°
Smile	180°	45°	90°	135°	90°
Sad	180°	100°	170°	80°	10°
Surprise	160°	20°	170°	160°	10°

The programming methodology focused on simple embedded control using C/C++ language in Arduino IDE. The software algorithm continuously reads incoming Bluetooth data and stores the received command string. Conditional statements were then used to compare the received command with predefined emotion commands such as relax, smile, sad, surprise and normal. Once matched, the program sends angle commands to the five servo motors so that each motor moves to a specific position. Servo angle combinations were experimentally determined to generate accurate emotional expressions. After execution, the command buffer is cleared and the system returns to standby mode for the next instruction.

Finally, the complete system was assembled by combining the mechanical structure, electrical components and programmed controller into one functional prototype. The software was uploaded to the Arduino Uno, electronic components were mounted onto the robot structure, and performance testing was conducted to verify that spoken commands successfully generated the intended mouth expressions. This integrated methodology ensured the successful development of a voice-controlled humanoid robotic mouth system.

3. RESULTS AND DISCUSSION

The findings demonstrate that the proposed prototype successfully generated multiple facial expressions through voice-controlled actuation while improving realism compared with earlier prototypes. The fabrication results confirmed that the final robotic mouth model was successfully assembled using mechanical, electronic and control components. The electrical connection consisted of five servo motors powered by a regulated 5V supply obtained by stepping down 7.4V battery voltage through an LM2596 buck converter. Servo control signals were connected to Arduino Uno PWM pins 3, 5, 6, 9 and 10. The HC-05 Bluetooth module was powered from the Arduino 5V output, while TX and RX pins were cross-connected for serial communication. A separate 9V battery supplied the Arduino controller, ensuring stable operation during servo actuation.

The final fabricated model was produced using 3D printing with PLA material. The structure measured approximately 130 mm × 131 mm × 83 mm and was designed with 5-DOF. Compared with the earlier prototype versions, the final design incorporated improved corner lip mechanisms using two movable links on each side connected through common joints. Four servo motors were dedicated to lip corner motion, while one additional servo motor controlled jaw movement. This configuration enabled smoother and more human-like lip motion.

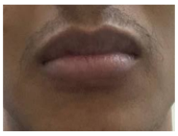
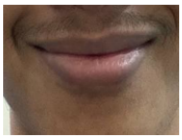
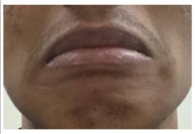
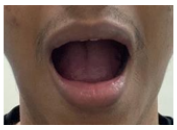
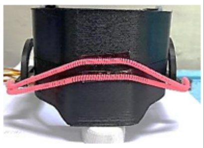
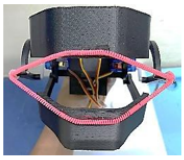
Functional testing was conducted using the AMR_Voice Android application. Before operation, the smartphone had to be connected to the Bluetooth module and internet service for speech recognition processing. Users activated the microphone icon and issued voice commands to control the robotic mouth. Four primary command modes were supported, namely relax, smile, sad and surprise. Once the command was recognized and transmitted, the robotic mouth changed its facial expression accordingly. However, clear pronunciation was necessary because unclear speech could cause incorrect recognition by the smartphone application.

Comparative observation between a human mouth, the initial prototype and the final model showed significant improvement in realism. The first and second prototypes were limited to 3-DOF and mainly served as proof-of-concept models. Although they could reproduce four basic expressions, the movement appeared stiff and less attractive due to restricted servo rotation and limited lip motion radius. In contrast, the final model was closer to human mouth dimensions and used a mechanism inspired by natural facial muscles. Upper and lower linkage motions resembled the zygomaticus and depressor “anguli oris” muscles, allowing more natural expression changes. This made the robot visually more engaging during interaction.

The communication system relied on Bluetooth technology between the smartphone and HC-05 module. The Android device converted spoken words into text strings, which were transmitted wirelessly to the Arduino microcontroller. The controller then interpreted the command and executed predefined servo angle movements. In addition to direct commands, several alternative keywords were also accepted. For example, smile could also be activated using happy or cheese, while relax could be triggered using normal or chill. This feature increased system usability and flexibility. Nevertheless, speech recognition errors sometimes occurred, such as interpreting “sad” as “set” or “normal” as “Norman.”

Table 4 presents the comparison of human lip expressions and the developed humanoid robotic mouth for four emotional expressions: relax, smile, sad and surprise. The figure also summarizes the corresponding servo response and estimated recognition success rate for each expression. Recognition percentages were estimated based on 20 visual observation trials conducted in indoor laboratory conditions. The generated facial expressions were evaluated through comparative visual observation and preliminary observer assessment to determine the recognizability of each emotional expression under indoor laboratory conditions.

Table 4 Simulation results

EMOTION	Relax	Smile	Sad	Surprise
Human				
Humanoid Robot Face				
Servo Response	Successful	Successful	Successful	Successful
Recognition Success	90–95%	80–88%	65–75%	90–95%

Bluetooth signal coverage was evaluated by gradually increasing the distance between the smartphone and the robotic system. Stable communication was successfully maintained at distances of 2 m, 4 m, 6 m and 8 m, while signal loss occurred at 10 m. The results indicate that the practical operating range of the HC-05 Bluetooth module under the conducted test conditions was approximately 8 m. This observation is consistent with the typical short-range characteristics of Bluetooth communication operating within the 2.4 GHz frequency band.

The wireless signal attenuation was analysed using the Free Space Path Loss (FSPL) model which describes the reduction in radio signal strength during wireless transmission [11]. The increase in communication loss as the transmission distance increases can be represented by Equation (1):

$$FSPL(dB) = 20log_{10}(d) + 20log_{10}(f) + 32.44 \quad (1)$$

where (d) represents the transmission distance in kilometres and (f) represents the operating frequency in MHz. Figure 5 shows that Bluetooth communication remained highly stable at short distances, achieving 100% signal reliability at 2 m and 4 m. Since the HC-05 module operates at approximately 2.4 GHz (2400 MHz), the signal experiences increasing free space path loss (FSPL) as the communication distance increases, resulting in weaker received signal strength. As the distance increased to 6 m and 8 m, the signal reliability gradually decreased to approximately 95% and 90%, respectively, although the robotic mouth still responded correctly to most voice commands. Complete communication failure occurred at 10 m when the received signal strength likely dropped below the minimum threshold required for reliable Bluetooth transmission. Environmental factors such as surrounding obstacles, electromagnetic interference and device orientation may also contribute to signal degradation during wireless communication. Overall, the results demonstrate that the HC-05 module provides reliable short-range wireless communication for indoor human-robot interaction applications within an effective operating range of approximately 8m.

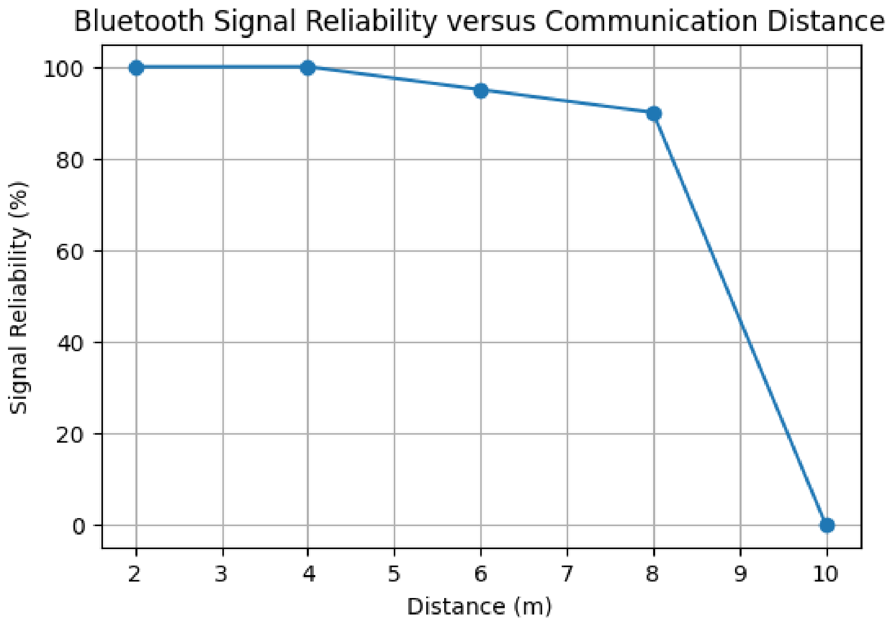


Figure 5. Bluetooth communication result

Overall, the results indicate that the developed robotic mouth system successfully achieved its objective of producing multiple facial expressions using voice commands. The final design demonstrated clear improvement over earlier prototypes in terms of movement realism, interaction quality and functional performance.

4. CONCLUSION

As a conclusion, a fully functional 5-DOF robotic mouth was fabricated and tested. The robotic mouth consists of mechanical, electrical and programming elements which result in emotional expression. The robot is capable of recognizing and responding to different voice commands and producing the corresponding emotional expressions, which are smile, sadness, relax / calm and surprised / shocked. The robotic structure was fabricated using PLA material and has 130 x 131 x 83 mm^3 in dimensions. The developed prototype was successfully tested and was capable of generating the intended facial expressions. A humanoid robot's head portion can be fitted with the final version of the robotic mouth. As technology and human intelligence have advanced, emotional expressive robots have become crucial for improving human-robot interactions (HRI). Its application in live advertising, home companions, nursing care, storytelling and interactive robotic machines is rapidly increasing. Robots that resemble humans are evidence that people have learned enough about the human body to spread positive human-robot interactions.

For future work, the system can be further enhanced by implementing real-time lip synchronization with speech, AI-based emotion recognition, additional facial degrees of freedom for more natural expressions and wireless IoT monitoring capabilities to support advanced human-robot interaction applications.

ACKNOWLEDGEMENTS

The authors would like to express their sincere appreciation to the Centre of Excellence for Intelligent Robotics and Autonomous Systems (CIRAS) and the Faculty of Electrical Engineering

& Technology (FKTE), Universiti Malaysia Perlis (UniMAP) for the financial support provided for this research work.

REFERENCES

[1] R. Stock-Homburg, 2022. Survey of Emotions in Human–Robot Interactions: Perspectives from Robotic Psychology on 20 Years of Research. Int J of Soc Robotics 14, 389–411, doi: https://doi.org/10.1007/s12369-021-00778-6.

[2] T. Kishi *et al*., 2013. Impression survey of the emotion expression humanoid robot with mental model based dynamic emotions, in 2013 IEEE International Conference on Robotics and Automation, pp. 1663-1668, doi: 10.1109/ICRA.2013.6630793.

[3] X. Ke, B. Qiu, J. Xin, Y. Yun, 2015. Vision development of humanoid head robot SHFR-III, in 2015 IEEE International Conference on Robotics and Biomimetics (ROBIO), pp. 1590-1595, doi: 10.1109/ROBIO.2015.7418998.

[4] G. Trovato, T. Kishi, N. Endo, K. Hashimoto, A. Takanishi, 2012. Development of facial expressions generator for emotion expressive humanoid robot, in 2012 12th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2012), pp. 303-308, doi: 10.1109/HUMANOIDS.2012.6651536.

[5] T. Kishi *et al*., 2012. Development of expressive robotic head for bipedal humanoid robot, in 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 4584-4589, doi: 10.1109/IROS.2012.6386050.

[6] M.E. Reyes, I.V. Meza, L.A. Pineda, 2019. Robotics facial expression of anger in collaborative human–robot interaction. International Journal of Advanced Robotic Systems, January 2019, doi:10.1177/1729881418817972

[7] V. Seib *et al*., 2013. Enhancing human-robot interaction by a robot face with facial expressions and synchronized lip movements, in 21st International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG), pp. 70–77.

[8] H.-J. Hyung, B.-K. Ahn, D. Choi, D. Lee, D.-W. Lee, 2016. Evaluation of a Korean Lip-sync system for an android robot, in 2016 13th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), pp. 78-82, doi: 10.1109/URAI.

[9] Y. Usui, R. Niiyama, Y. Kuniyoshi, 2019. Anthropomorphic Face Robot having Soft Mouth Mechanism with Embedded Artificial Facial Muscles, in 2019 International Symposium on Micro-NanoMechatronics and Human Science (MHS), pp. 1-6, doi: 10.1109/MHS48134.2019.9249338.

[10] L.K.M.D. Harshani, W.M.A.S.B. Weerasooriya, H.M.C.S. Herath, P.M.K. Alahakoon, W.G.C.W. Kumara, M.N.A. Hinas, 2021. Development of a humanoid robot mouth with text-to-speech ability, in 2021 The Second Faculty Annual Research Session (FARS 2021), pp. 81-87.

[11] T. S. Rappaport, Wireless Communications: Principles and Practice, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall, 2002.