Development of a Touchless Control System for a Clinical Robot with Multimodal User Interface

Julio Alegre Luna, Anthony Vasquez Rivera, Alejandra Loayza Mendoza,
Jesús Talavera S., Andres Montoya A.
Universidad Nacional de San Agustín de Arequipa, Perú

***Abstract*—This article introduces the development of a multimodal user interface for touchless control of a clinical robot. This system seamlessly integrates distinct control modalities: voice commands, an accelerometer-embedded gauntlet, and a virtual reality (VR) headset to display real-time robot video and system alerts. By synergizing these control approaches, a more versatile and intuitive means of commanding the robot has been established. This assertion finds support through comprehensive assessments conducted with both seasoned professionals and novices in the domain of clinical robotics, all within a controlled experimental setting. The diverse array of test results unequivocally demonstrate the system's efficacy. They substantiate the system's ability to proficiently govern a robotic arm in the clinical environment. The user interface's usability is measured at an impressive 90.2 on the system usability scale, affirming its suitability for robotic control. Notably, the interface not only offers comfort but also intuitiveness for operators of varying levels of expertise.**

***Keywords*—Multimodal user interface; human–robot interaction; clinical robot**

I. Introduction

Robotic systems are progressively assuming greater significance in the development of conventional human activities, especially within the realm of healthcare, owing to their array of merits. These encompass heightened efficiency and precision, risk mitigation, and enhanced patient comfort [1]. These systems find versatile utility across a spectrum of clinical applications, spanning from surgical procedures and rehabilitation to technologies tailored for the aid of the elderly or disabled [2], [1]. Yet, the management of these systems within a clinical milieu poses intricate challenges. Conventional interfaces like buttons and joysticks, while conventionally employed, engender potential infection hazards and present difficulties for patients possessing restricted mobility or dexterity [3].

In response to these challenges, researchers have embarked upon the exploration of novel paradigms for robotic control that are imbued with heightened intuition, naturalness, and touchlessness [4], [5]. An auspicious avenue in this endeavor is the adoption of multimodal user interfaces, which amalgamate an assortment of control methodologies, thus engendering a more adaptable system [6]. Multimodal user interfaces empower users to seamlessly transition between diverse control modes contingent upon their proclivities or the specific task at hand [6]. As an illustration, a user might seamlessly oscillate between voice commands to oversee the robot's locomotion, while seamlessly transitioning to gesture-based control for tasks demanding precision in manipulation [7].

This article introduces a multimodal user interface developed for touchless control of a clinical robot. This system seamlessly integrates two distinct control methodologies: voice commands, and an accelerometer-embedded gauntlet. The voice command system empowers users to steer the robot through spoken directives, while the accelerometer-equipped gauntlet detects user-initiated gestures. Finally, the virtual reality headset allows the operator to visualize in real time the video captured by the webcam, and also allows the system to display on screen different alerts triggered by the two methods mentioned above. This system boasts a spectrum of prospective applications within varied clinical contexts. For instance, it could find utility in surgical settings, enabling surgeons to orchestrate robot movement while maintaining a sterile environment. Similarly, within rehabilitation realms, patients might exercise dominion over robotic devices using their voice or gestures. Furthermore, the system's utility is magnified for individuals with restricted mobility or dexterity, as it allows them to exert control over the robot devoid of physical interaction. To the best of current knowledge, this represents the maiden multimodal user interface tailored for touchless control of a clinical robot, concomitantly amalgamating voice commands, accelerometer-equipped gauntlet, and a VR headset that displays the developed user interface. The system's architectural blueprint prioritizes user-friendliness, safety, and reliability, buttressed by a series of meticulously devised experiments aimed at scrutinizing its efficacy in robot manipulation.

The subsequent sections of this article are structured as follows: In Section II, Related Works, prior research concerning voice-controlled systems, gesture recognition systems, and multimodal user interfaces for robot and robotic arm control is comprehensively surveyed. Section III, Experimental Development, an exhaustive account is provided regarding all employed electronic components, the clinical robotic platform, the implementation of the voice control system, the accelerometer-embedded gauntlet, and the associated software architecture conceived for this integrated system. The subsequent segment, Section IV, Results and Discussions, unveils the empirical outcomes derived from the conducted experiments, accompanied by the ensuing discussion arising from their interpretation. Ultimately, in Section V, this document culminates as conclusions are drawn and prospective avenues of research are deliberated upon.

II. Related Works

A. Voice-Control Systems

Sagar's article [8] reviewed the current status of speech recognition systems. In addition, the potential industrial applications of speech recognition technology, such as public safety solutions, were discussed. Furthermore, the article delves into the future scope of voice recognition, with the potential for artificial intelligence to reshape how we interact with devices [9]. In another study [10], a voice recognition control system for a robot is delineated, designed to operate effectively in noisy environments. This system employs generalized sidelobe canceller techniques resilient to outliers, noise suppression in the feature space, and reverberation mitigation. The article also delves into obstacle detection, local map design, as well as target search and avoidance behaviors using fuzzy decision-making. The system's efficacy is evaluated on a communication robot deployed within a real noisy environment. The article also contemplates the integration of robust voice recognition and navigation systems for autonomous navigation within unfamiliar surroundings.

In another study [11], a system is proposed that provides a mobile robot with the ability to separate simultaneous sound sources; an array of microphones is used along with a dedicated real-time implementation of geometric source separation and a post-filter that provides us with further reduction of interference from other sources. The work of [12] discusses the creation of target-seeking and avoidance behaviors employing fuzzy decision-making. The author in [13] introduces a method for selecting an appropriate behavior from numerous primitive behaviors using a fuzzy decision-maker. author in [14] describes an obstacle detection method and local map design utilizing an array of ultrasonic sensors. Finally, [15] introduces a novel approach to voice recognition in noisy environments, grounded in multi-condition training techniques, maximum likelihood linear regression, and missing feature theory.

B. Gesture Recognition Systems

The article [16] introduces a human-computer interaction (HCI) model based on somatosensory interaction for robotic arm manipulation. The model utilizes a 3D SSD architecture for gesture and arm movement localization and identification, coupled with the Dynamic Time Warping (DTW) template matching algorithm for dynamic gesture recognition. Interactive scenarios and modes are designed for experimentation and implementation, with virtual experimental results demonstrating the method's efficacy. In [17], a real-time hand gesture recognition system is presented for controlling mobile robots using vision sensors. The system employs image processing techniques to extract the center of mass and features of a red glove worn by the user. These features are then used to control the robot's movements. The design of the mobile robot is uncomplicated and tailored for the system, consisting of three layers with a 4 cm separation to accommodate circuit placement. The system employs a motor control circuit and a PIC18F452 microcontroller control circuit. Additionally, the system incorporates XBee wireless transmitter and receiver modules for data transmission. The system employs color filtering to extract the red glove's shape and spot size filtering to eliminate objects below a certain size [18].

The article [19] presents a gesture recognition system for interacting with computers in dynamic environments. The system employs image processing techniques for hand gesture detection, segmentation, tracking, and recognition, transforming them into meaningful commands. The proposed interface finds applicability across diverse domains like image navigation and gaming. Real-world scenario testing exhibited effective performance in low-noise environments and balanced lighting conditions. The designed gesture vocabulary can be expanded to control different applications, enhancing adaptability in human-computer interaction. This work is aligned with research in the field of human-computer interaction and gesture recognition. The article [20] pertains to the realm of human-robot interaction, focusing on real-time hand gesture recognition to enhance human-robot interaction within dynamic environments. Enhanced classifiers are employed for hand detection and static gesture recognition, while a Bayesian classifier is utilized for dynamic gesture recognition. Additionally, the system incorporates contextual information, such as human face detection and tracking, to enhance robustness and speed. Relevant works utilizing contextual information to improve the accuracy and speed of gesture recognition systems are also referenced. The proposed system's validation is conducted on actual video sequences, achieving a recognition rate of 70% for static gestures and 75% for dynamic gestures, operating at varying speeds of 5-10 frames per second.

C. Multimodal User Interfaces

The document [21] presents work related to human-robot collaboration (HRC) in manufacturing, specifically in assembly tasks. The challenges and limitations of existing HRC systems are discussed, including issues such as lack of adaptability and flexibility in task programming, along with the need to ensure human safety in the working environment. The article proposes a solution based on the utilization of function blocks and intuitive multimodal control to enhance flexibility and adaptability in complex assembly tasks. Concepts of multimodal control, function blocks, and their application in human-robot collaboration within manufacturing are thoroughly examined. Conversely, the article [22] delves into a usability study of three interfaces designed to present search engine results on the Internet. The study compared a text-only interface with two others that combined text, visual metaphors, and voice messages. Results indicated that the multimodal interfaces were more usable than the text-only interface. In a third work, Lunghi's article [23] details the design and software engineering process behind the development of a multimodal Human-Robot Interface (HRI) for intervention with a cooperative team of robots. The operator gains the capability to enter the control loop between the HRI and the robot, customizing control commands in accordance with the operation.

D. Robots in the Clinical Environment

Poirier's paper [24] presents the design and preliminary evaluation of a voice command system prototype for the control of assistive robotic arms' movements; the prototype of the voice command interface developed is first presented, followed by two experiments with five able-bodied subjects in order to assess the system's performance and guide future development. In the work of Morgan et al. [25], a comprehensive literature review is presented, focusing on the utilization of robots within the realm of healthcare. The study identifies ten primary roles that robots can undertake in clinical settings, encompassing surgery, rehabilitation and mobility, radiotherapy, social assistance, telepresence, pharmacy, disinfection, delivery and transportation, image intervention, and assistance. Furthermore, the article underscores robots' potential to adapt to the dynamic demands of healthcare, including those that arise during pandemics.

In Peter's study [26], a novel multimodal human-machine interface system is developed using combinations of electrooculography (EOG), electroencephalography (EEG), and electromyogram (EMG) to generate numerous control instructions; the results indicate that the number of system control instructions is significantly greater than achievable with any individual mode. In other paper [27], an interface centered on the deployment of the Leap Motion (LM) controller is examined. This interface facilitates the real-time tracking of a surgeon's hand position and orientation, capturing nuanced finger gestures and movements, which are subsequently relayed to a computer. Subsequently, a surgical robotic arm is manipulated using data gleaned from the LM controller, data that is systematically classified through programming. Beyond the capabilities attributed to the LM controller, attributes like its cost-effectiveness, acceptable precision, and high-speed data processing have rendered it a feasible and efficient tool for application.

III. Methodology

The proposed system enables the user to control a contactless robotic arm through a multimodal user interface. two control methods are integrated: voice commands, and hand gestures. The voice command system empowers users to steer the robot through spoken directives, while the accelerometer-embedded gauntlet detects user-initiated gestures. A VR headset that displays the developed user interface. Fig. 1 illustrates the block diagram of the proposed system, which is subsequently elaborated upon in each stage.

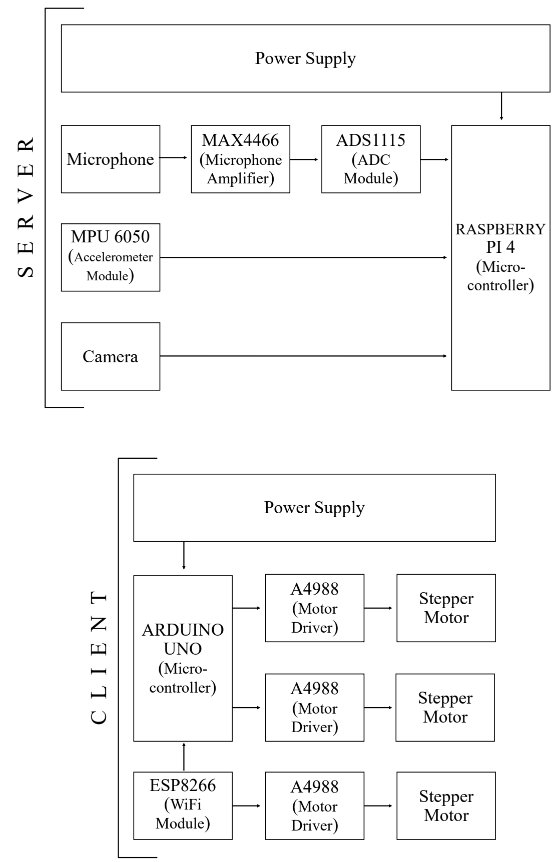

Fig. 1. Block diagram of the complete system.

A. Hardware Components

The hardware components utilized in this study encompass a Raspberry Pi, an Arduino Nano with WiFi module, a microphone, an accelerometer-embedded gauntlet, and a webcam. The Raspberry Pi 4 serves as the central processing unit of the system. The Arduino is employed to manage the motor drivers of the robotic arm. The microphone, along with its associated circuitry, is employed for voice command recognition and transmission to the Raspberry Pi. The accelerometer-embedded gauntlet captures hand gestures executed by the user. Lastly, the webcam connected to the Raspberry Pi detects the user's facial features.

1) Microphone Circuit: The selected transducer type is a microphone, which is connected to an amplification stage (MAX4455 amplifier) to condition the signal to the desired voltage level, ranging between 0 and 5V. The microphone captures sound waves and converts them into an electrical signal, which is then transmitted to the Raspberry Pi microprocessor. Positioned between the amplification stage and the microprocessor is an analog-to-digital conversion stage (ADS1115 converter). This conversion stage is crucial as it enables the collection of analog signals and their subsequent processing in a digital-origin microprocessor.

In Fig. 2, the two pins of the microphone are connected, one to the amplification stage and the other to GND. The amplifier is configured in a non-inverting setup. The amplifier's output is connected to pin 1 of the ADC. Pin 7 of the ADC is connected to GPIO9 on the Raspberry Pi, serving as the data transmission pin. Lastly, pin 8 is connected to GPIO10 on the Raspberry Pi, serving as the clock signal pin.

2) Glove Circuit with Accelerometer: Accelerometers are devices that measure acceleration force in units of gravity (g) and can measure in one, two, or three planes (X, Y, and Z). The chosen module for this stage is the MPU-6050, which integrates a MEMS accelerometer and a MEMS gyroscope on a single chip. This module is installed in a glove worn by the operator of the robotic arm, capturing hand movements as well as any rotations they perform.

In Fig. 3, the GND pin of the MPU-6050 module is connected to the circuit's ground, while the VDD pin is linked to the voltage output of the Raspberry Pi 4. The SDA pin transmits accelerometer module data to the Raspberry Pi and is connected to GPIO3. The SCL pin of the MPU-6050 module transfers the clock signal to the module and is linked to GPIO5.

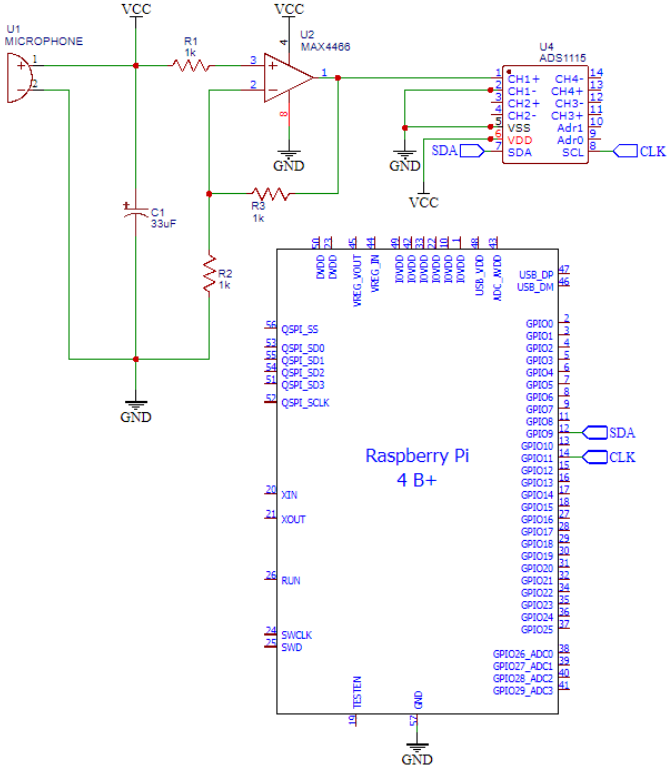

Fig. 2. Connection circuit between the microphone and the Raspberry Pi microprocessor.

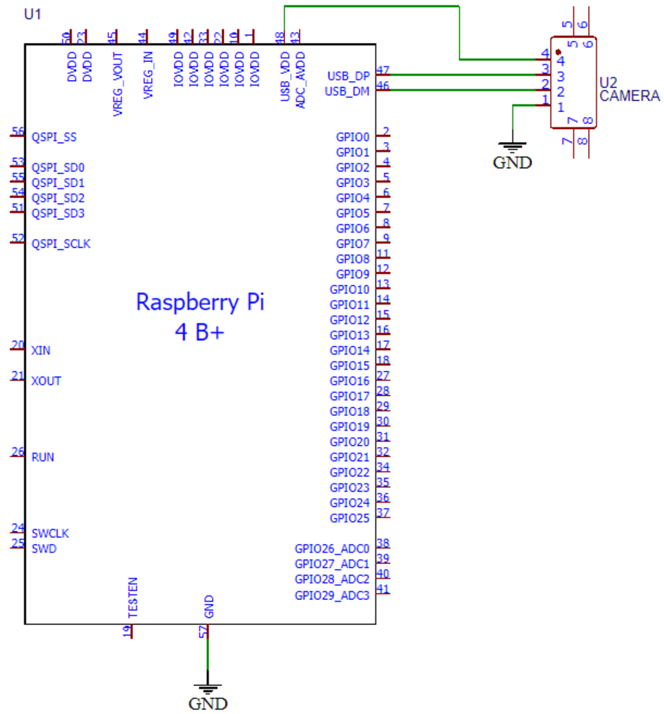

Fig. 4. Connection circuit between the camera and the Raspberry Pi microprocessor.

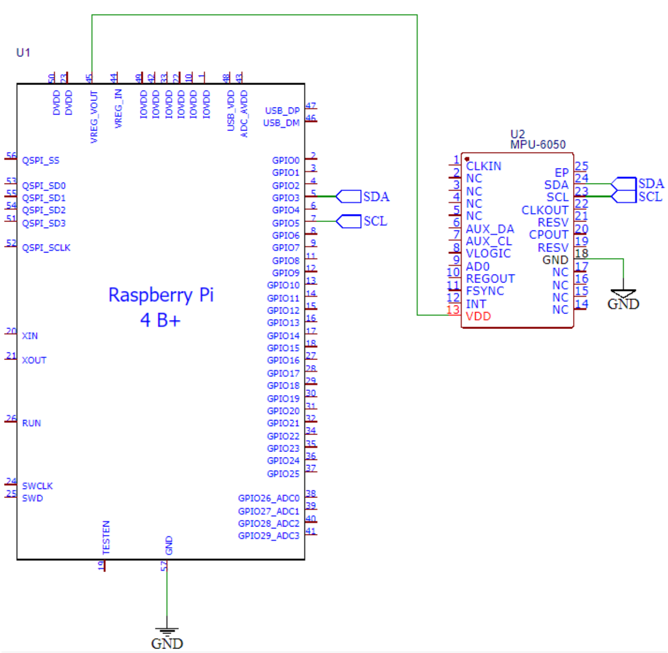

Fig. 3. Connection circuit between the accelerometer of the glove and the Raspberry Pi microprocessor.

3) Webcam Circuit: The operation of this stage is straightforward. A webcam is used to transmit video of the robot, which will serve as feedback to the multimodal system. In Fig. 4, the Logitech C922 camera is connected to the Raspberry Pi 4 via its USB port.

4) Wireless Communication: In this stage, there are two components. On one side, there's the Raspberry Pi 4, which comes equipped with integrated Wi-Fi. This Wi-Fi functionality is utilized to create a server, enabling the robotic arm to connect to it as a client. As for the robotic arm segment, an Arduino Uno is employed for control. However, since the Arduino Uno doesn't have a built-in Wi-Fi module, an external Wi-Fi module, specifically the ESP8266, is utilized for this purpose.

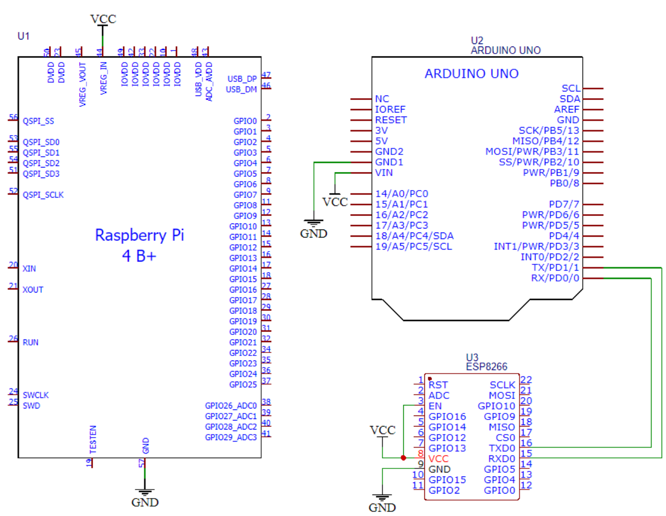

Fig. 5. Wireless communication connection circuit.

In Fig. 5, the Raspberry Pi 4, Arduino, and the ESP8266 module share common VCC and GND connections. The

Arduino and ESP8266 are linked through the TX and RX transmission pins.

5) Connection of the Robotic arm: The testing robotic arm has 3 degrees of freedom (3-DOF), which is why 3 stepper motors and 3 drivers are employed to control its movements. These components are connected to the Arduino to issue commands for their respective functions. The A4988 drivers are chosen due to their high reliability in tasks of this nature.

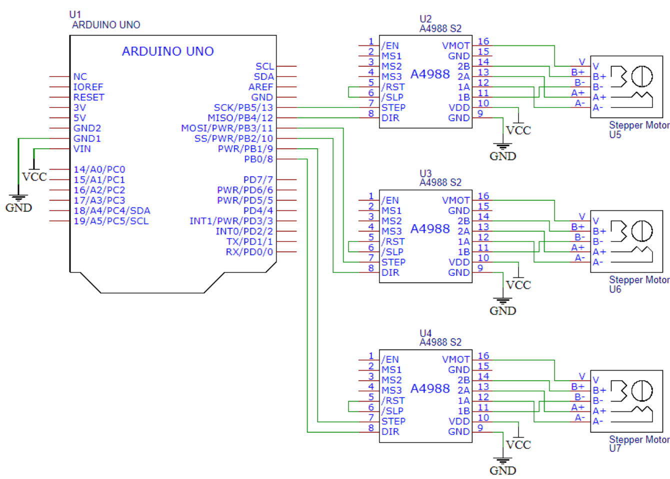

Fig. 6. Robotic arm connection circuit.

In Fig. 6, the Arduino and the motor drivers share the VCC and GND power supply. From the motor drivers, two pins are used to connect to the Arduino: the STEP and DIR pins. These pins determine the number of steps and the direction of rotation, respectively. The pins of the first driver are connected to Arduino pins 12 and 13, the pins of the next driver are connected to pins 10 and 11, and finally, the pins of the third driver are connected to pins 8 and 9 of the Arduino. Each driver is linked to a 4-wire stepper motor, with the 4 wires connected to A+, A-, B+, and B- pins.

B. Software Components

The system was developed using the Python programming language due to its extensive library support and versatility for programming innovative systems. Python was utilized to integrate the various software components of the system and to control the robot based on user inputs. Additionally, several software components were employed to make the system function effectively. On the other hand, the Arduino was programmed using its own platform and libraries for motor drivers. The following are the most significant details for this purpose.

1) Voice Interaction: For voice recognition, Google Speech-to-Text was employed to transcribe the voice commands issued by the user. It was implemented using the Google Cloud API and integrated into the Python code executed on the Raspberry Pi 4. Voice command language was preferred as the input to the system because it allows a more intuitive interaction to the human's natural being [28]. The commands used and recognized by the system are displayed in Table I. An indicator provides visual information to the operator of the action being executed. During active navigation mode, the indicator is illuminated according to the Table I. This lets the operator know which command is currently being executed, as well as whether the spoken command was successfully acknowledged.

TABLE I. DESCRIPTION OF INTEGRATED VOICE CONTROL COMMANDS

Voice Command	Indicator	Description
Start	🟢 🟢	Activates glove gesture detection
Translation	○ 🟢	Initiates translational navigation mode
Rotation	🟢 ○	Initiates rotational navigation mode
Move to	🟢 🟡	Opens the options assignment display
Cancel	🟡 ○	Closes the options assignment display
Stop	🟡 🟡	Deactivates the selected navigation mode
End	🔴 🔴	Concludes the current interaction task

2) Glove Gesture Interaction: Control through the glove is achieved by integrating an accelerometer which captures the degrees of inclination of the hand in its different axes (X, Y and Z). The multimodal system allows the operator to execute different commands to the robotic arm thanks to the integration of voice commands, some examples are the function of pausing the sending of accelerometer data to the Raspberry Pi, this allows the operator to rest momentarily or move the hand without worrying that the robot will recreate this movement; Other examples are the rotation and translation commands that allow the robotic arm to move according to the indication executed and the hand movement performed.

3) User Interface Display: The user interface collects the webcam video sent by the Raspberry Pi 4 from the server, this video is processed and the spoken command indicator is added so that the operator can realize that it worked correctly; to process the video, the OpenCv library belonging to the Python programming language was used. OpenCV is a powerful computer vision library that was used to detect hand gestures performed by the user. In Fig. 7, you can see an image of the processed video.

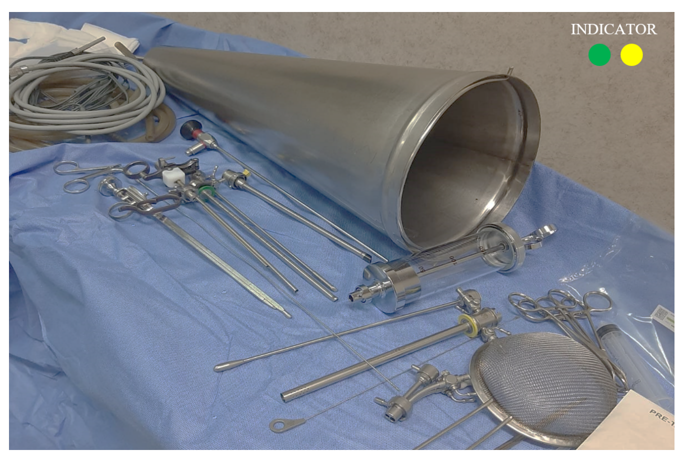

Fig. 7. User interface develops visualized in the virtual reality headset.

C. Testing

To assess the performance of the proposed multimodal system, a series of experiments were conducted involving 12 participants. Each participant was assigned a set of tasks to perform with the robot, including moving the robotic arm to a desired position, touching specific elements in the environment with the end effector, among others. Fig. 8 depicts the testing scenario used to evaluate the system's effectiveness.

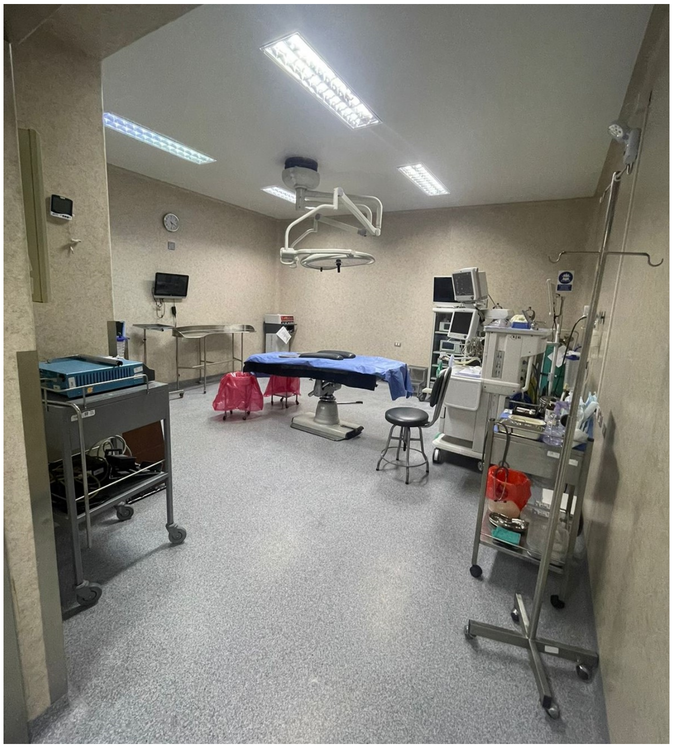

Fig. 8. Testing environment for the robotic arm.

Participants were instructed to use each of the three control methods (voice commands, glove gestures, and computer vision) individually and in combination to control the robot. The sequence of method usage was randomized to mitigate order effects. The system's performance was assessed based on task completion time, the accuracy of robot movements, and participants' subjective feedback on the ease of interface use.

D. Data Analysis

Task execution times and robot movement accuracy were recorded for each participant and analyzed using descriptive statistics. Subjective opinions were collected using the System Usability Scale (SUS) method and analyzed through qualitative approaches. SUS provides a "quick and dirty", reliable tool for measuring the usability, it consists of a 10 item questionnaire with five response options for respondents; from Strongly agree to Strongly disagree [29]. The multimodal user interface was implemented and tested on a clinical robot within a simulated laboratory setting. System performance was assessed in terms of accuracy, speed, and user-friendliness.

IV. Results and Discussions

The multimodal user interface was implemented and tested on a test clinical robot in a laboratory environment. The system's performance was evaluated in terms of accuracy, speed, and user-friendliness. The tests were carried out with a prototype of a robotic arm manufactured with three stepper motors and an Arduino Uno microcontroller. Fig. 9 shows the system server made up of a Raspberry Pi 4, glove with accelerometer, electronic components and the Virtual Reality headset.

Fig. 9. Electronic components of the developed system.

A. Voice Command Control

Voice control proved effective in maneuvering the robot and executing various commands. The accuracy of the voice recognition system was evaluated using a speech recognition rate metric, which measures the percentage of correctly recognized commands out of the total number of given commands. The speech recognition rate was 92%, indicating a high level of accuracy in recognizing voice commands.

B. Glove Control with Accelerometer

The accelerometer-equipped glove proved to be effective in capturing hand gestures and providing a natural way to control the robot. Fig. 10, 11, 12 displays the graph obtained by comparing accelerometer values along its 3 axes (X, Y, and Z) with the angles of rotation of the robot arm's corresponding 3 axes.

Fig. 10, 11, and 12 depict each of the three accelerometer axes positioned within the operator's gauntlet. The operator executed hand movements for a duration of one minute, yielding a total of 500 samples collected per accelerometer axis. These measurements correspond to the angular velocity (°/s) of motion recorded during the trials. Fig. 10 showcases the data acquired from the X-axis accelerometers, encompassing both the gauntlet and the robotic arm, while Y-axis data is presented in Fig. 11, and Z-axis data is delineated in Fig. 12. On average, a variation of 3.87% was observed, attributed to the motor configurations driven by the actuators.

C. Multimodal Control

The three control methods were amalgamated to forge a comprehensive multimodal user interface. Users were empowered to seamlessly switch between diverse control modes, tailoring their choice based on personal preferences and the specific task at hand. Empirical evidence substantiated the

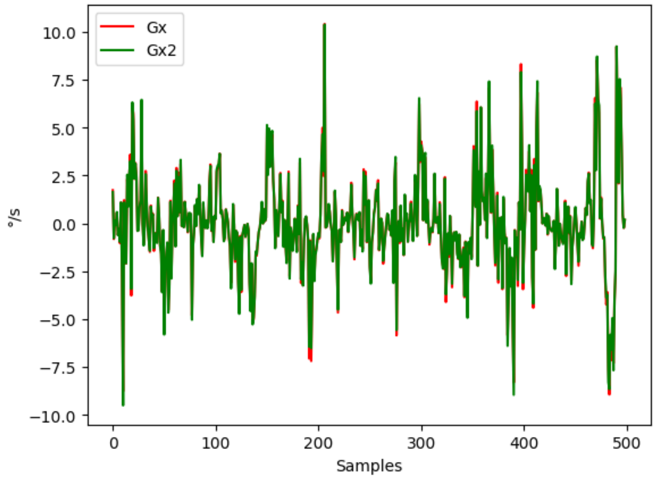

Fig. 10. Plot of accelerometer measurements and rotation of the robotic arm axes in the X-axis.

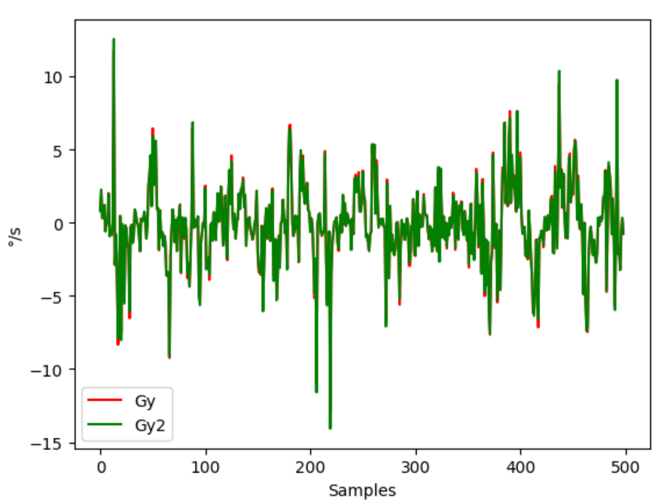

Fig. 11. Plot of accelerometer measurements and rotation of the robotic arm axes in the Y-axis.

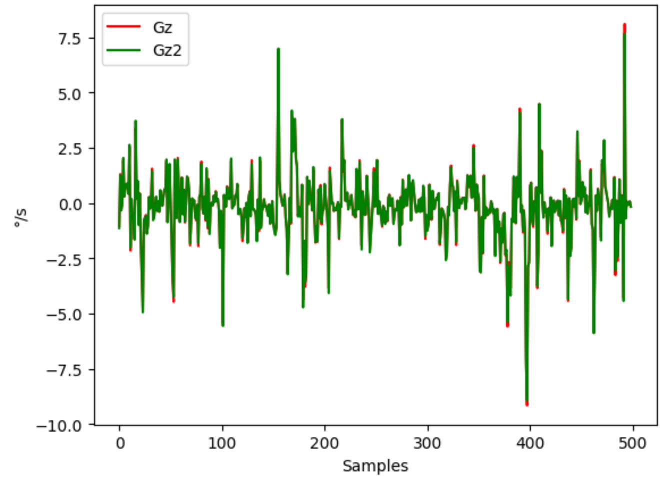

Fig. 12. Plot of accelerometer measurements and rotation of the robotic arm axes in the Z-axis.

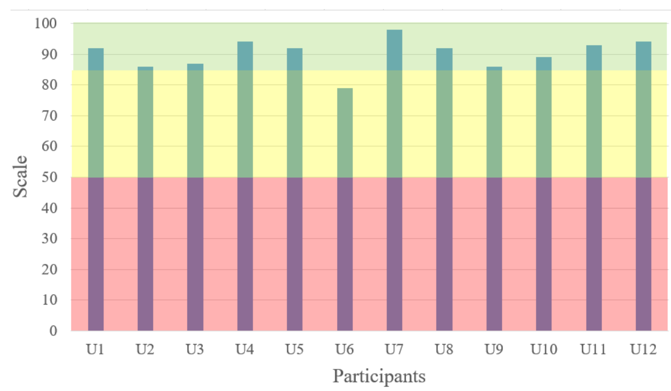

Fig. 13. Graph of results obtained from SUS measurement of the system.

superiority of the multimodal interface over singular control approaches, demonstrating that users adeptly transitioned between voice commands, hand gestures, and artificial vision control. This versatility imbued human-robot interaction with enhanced flexibility and intuitive fluidity. Moreover, the amalgamation of distinct modalities endowed a heightened precision of control, particularly advantageous for tasks necessitating meticulous accuracy, such as surgical procedures. Refer to Fig. 13 for a graphical representation of the values obtained through the evaluation of the user interface and the proposed system, utilizing the System Usability Scale method (SUS), a widely adopted metric for gauging the effectiveness of an interface for a given task. The color background of this graph shows three different scoring areas: light red for poor usability ($SUS_{score} < 50$), light yellow for good usability ($85 > SUS_{score} \geq 50$), and light green for excellent usability ($SUS_{score} \geq 85$).

The obtained average value for the proposed interface was $SUS_{score} = 90.2$ points, falling within the range indicative of commendable interfaces. These findings strongly indicate that the suggested user interface is exceptionally well-suited for orchestrating robotic arms within clinical scenarios. On the whole, the outcomes of this study strongly propose that the developed multimodal user interface holds substantial potential for enhancing the efficiency and efficacy of clinical robots within healthcare settings. The capability to govern the robot through voice commands, hand gestures, and artificial vision confers a heightened level of flexibility and intuitive interaction with the robotic system. This, in turn, stands to enhance patient outcomes and foster a higher adoption rate of the technology amongst healthcare professionals.

V. Conclusions

In summary, a multimodal user interface has been introduced for touchless control of a clinical robot, seamlessly integrating voice commands, an accelerometer-equipped gauntlet, and display of the user interface on the virtual reality headset in real time. The outcomes derived from the conducted trials robustly suggest that the utilization of a multimodal interface holds the potential to enhance the efficiency and efficacy of clinical robots within healthcare environments, as evidenced by the notable 90.2 point outcome on the SUS scale. The capacity to manipulate a robotic arm through the fusion of voice commands, hand gestures, and artificial vision engenders a more adaptable and intuitive means of interacting with the arm, a facet that has the potential to enhance patient outcomes and bolster the technology's embrace amongst healthcare professionals. Future endeavors will be concentrated on refining the interface and appraising its effectiveness within clinical environments, involving real patients. Additionally, the incorporation of other modalities, such as haptic feedback and augmented reality, could be explored to further heighten user experience and system performance.

ACKNOWLEDGMENT

The authors would like to thanks Universidad Nacional de San Agustín de Arequipa.

REFERENCES

[1] Giovanni Morone, Ilaria Cocchi, Stefano Paolucci and Marco Iosa. Robot-assisted therapy for arm recovery for stroke patients: state of the art and clinical implication. Expert Review of Medical Devices 2020, vol. 17, pp. 223-233.

[2] Mahdieh Babaiasl, Seyyed Hamed Mahdioun, Poorya Jaryani and Mojtaba Yazdani. A review of technological and clinical aspects of robot-aided rehabilitation of upper-extremity after stroke. Disability and Rehabilitation: Assistive Technology 2015, vol. 11, pp. 263-280.

[3] Rafael Verano M, Jose Caceres S, Abel Arenas H, Andres Montoya A, Joseph Guevara M, Jarelh Galdos B and Jesus Talavera S, "Development of a Low-Cost Teleoperated Explorer Robot (TXRob)" International Journal of Advanced Computer Science and Applications(IJACSA), 13(7), 2022.

[4] V. Duchaine and C. Gosselin. Safe, Stable and Intuitive Control for Physical Human-Robot Interaction. 2009 IEEE International Conference on Robotics and Automation, Kobe, Japan, 2009, pp. 3383-3388.

[5] William Percy Cuno Zuniga, Jordan Jonathan Quispe Navarro, Juan Daniel Pocohuanca Diaz, Jesus Talavera S., Andres Montoya A. Design of a Terrain Mapping System for Low-cost Exploration Robots based on Stereo Vision. Przeglad Elektrotechniczny 2023.

[6] Maurtua I, Fernández I, Tellaeche A, et al. Natural multimodal communication for human–robot collaboration. International Journal of Advanced Robotic Systems. 2017;14(4).

[7] Schreiter, J., Mielke, T., Schott, D. et al. A multimodal user interface for touchless control of robotic ultrasound. Int J CARS 18, 1429–1436 (2023).

[8] Sagar V. Fegade, Ashish Chaturvedi, Mukta Agarwal. Voice Recognition Technology : A Review. International Journal of Advanced Research in Science, Communication and Technology. 2021, vol. 8, no. 1.

[9] Cristina Bayón, Rafael Raya, Sergio Lerma Lara, Oscar Ramirez, Ignacio Serrano, Eduardo Rocon. Robotic Therapies for Children with Cerebral Palsy: A Systematic Review. Translational Biomedicine. 2016, vol. 7, no. 1.

[10] Jung, S.W., Park, M.Y., Park, I.M., Jung, Y.K., Shin, H.B. A Robust Control of Intelligent Mobile Robot Based on Voice Command. Intelligent Robotics and Applications. ICIRA 2013. Lecture Notes in Computer Science, vol 8102. Springer, Berlin, Heidelberg.

[11] I. Cohen, B. Berdugo. Microphone array post-filtering for non-stationary noise suppression. Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing. 2002, pp. 901–904.

[12] Jon Barker, Martin Cooke, Phil Green. Robust ASR based on clean speech models: An evaluation of missing data techniques for connected digit recognition in noise. 7th European Conference on Speech Communication and Technology. 2001, pp. 213-217.

[13] Philippe Renevey, Rolf Vetter, Jens Krauss. Robust speech recognition using missing feature theory and vector quantization. 7th European Conference on Speech Communication and Technology. 2001, pp. 1107-1110.

[14] S. Yamamoto, Kazuhiro Nakadai, Hiroshi Tsujino, Hiroshi Okuno. Assessment of general applicability of robot audition system by recognizing three simultaneous speeches. International Conference on Intelligent Robots and Systems. 2004, vol. 3, pp. 2111-2116.

[15] Jean-Marc Valin, Jean Rouat, François Michaud. Enhanced Robot Audition Based on Microphone Array Source Separation with Post-Filter. 2004, vol. 3. pp. 2123-2128.

[16] Xing Li. Human–robot interaction based on gesture and movement recognition. Signal Processing: Image Communication, vol. 81, 2020.

[17] Ahmad'Athif Mohd Faudzi, Muaammar Hadi Kuzman Ali, M. Asyraf Azman, Zool Hilmi Ismail. Real-time Hand Gestures System for Mobile Robots Control. Procedia Engineering, vol. 41, pp. 798-804, 2012.

[18] Malima, A., E. Ozgur, et al. A Fast Algorithm for Vision-Based Hand Gesture Recognition for Robot Control. Signal Processing and Communications Applications, 2006, pp. 1-4.

[19] Siddharth S. Rautaray, Anupam Agrawal. Real Time Gesture Recognition System for Interaction in Dynamic Environment. Procedia Technology. 2012, vol. 4, pp. 595-599.

[20] Correa, M., Ruiz-del-Solar, J., Verschae, R., Lee-Ferng, J., Castillo, N. Real-Time Hand Gesture Recognition for Human Robot Interaction. Lecture Notes in Computer Science 2009, vol 5949. Springer, Berlin, Heidelberg.

[21] Sichao Liu, Lihui Wang, Xi Vincent Wang. Symbiotic human-robot collaboration: multimodal control using function blocks. Procedia CIRP. 2020, vol. 93, pp. 1188-1193.

[22] Alejandro Jaimes, Nicu Sebe. Multimodal human–computer interaction: A survey. Computer Vision and Image Understanding. 2007, vol. 108, pp. 116-134.

[23] Lunghi, G.; Marin, R.; Di Castro, M.; Masi, A.; Sanz, P.J. Multimodal Human-Robot Interface for Accessible Remote Robotic Interventions in Hazardous Environments. IEEE Access. 2019, vol. 7, pp. 127290–127319.

[24] S. Poirier, F. Routhier, A. Campeau-Lecours. Voice Control Interface Prototype for Assistive Robots for People Living with Upper Limb Disabilities. International Conference on Rehabilitation Robotics, Toronto, ON, Canada. 2019, pp. 46-52.

[25] Morgan AA, Abdi J, Syed MAQ, Kohen GE, Barlow P, Vizcaychipi MP. Robots in Healthcare: a Scoping Review. Curr Robot Rep. 2022, 3(4):271-280.

[26] Zhang J, Wang B, Zhang C, Xiao Y and Wang MY (2019) An EEG/EMG/EOG-Based Multimodal Human-Machine Interface to Real-Time Control of a Soft Robot Hand. Front. Neurorobot. 13:7.

[27] M.H. Korayem, M.A. Madihi, V. Vahidifar. Controlling surgical robot arm using leap motion controller with Kalman filter. Measurement, vol. 178, pp. 109372, 2021.

[28] Ferre M, Macias-Guarasa J, Aracil R, Barrientos A. Voice command generation for teleoperated robot systems. 1998.

[29] Brooke J. SUS: A quick and dirty usability scale. Usability Evaluation in Industry. 1996, pp. 189-194.