IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 14, No. 2, April 2025, pp. 917~924
ISSN: 2252-8938, DOI: 10.11591/ijai.v14.i2.pp917-924

Interactive communication human-robot interface for reduced mobility people assistance

Robinson Jimenez Moreno[1], Anny Astrid Espitia Cubillos[2], Esperanza Rodríguez Carmona[2]
[1]Mechatronic Engineering Program, Faculty of Engineering, University Militar Nueva Granada, Bogotá, Colombia
[2]Industrial Engineering Program, Faculty of Engineering, University Militar Nueva Granada, Bogotá, Colombia

Article Info

Article history:

Received Mar 12, 2024
Revised Oct 22, 2024
Accepted Nov 14, 2024

Keywords:

Convolutional networks
Interactive communication
Interface
Man robot communication
Usability

ABSTRACT

Communication between a robot and its user is essential for the execution of tasks, even more so in a scenario where the robot is designed to assist people with reduced mobility. This document presents the evaluation of a conversation script between a human user and a robot for assistance using pre-recorded responses, for this a methodology with three phases was proposed and applied: establishment of the training scheme of a convolutional network that allows recognize user's words for execution of tasks by the robot, generation of dialogue between the user and possible interactions with the assistive robot and finally, the measurement of perception of interface users. Results show a high level of accuracy with words selected to command the robot, using a convolutional neural network, with an audio input discriminated in its components mel frequency cepstral coefficients (MFCCs) and command sets of male and female voices. It was possible to establish a dialogue model with three scenes to recognize the residential environment, rename spaces and execute action commands to move elements. It is concluded the designed instrument is reliable and the perception of proposed interactive communication interface is good in terms of usability (effectiveness, efficiency, and user satisfaction).

Corresponding Author:

Robinson Jimenez Moreno
Mechatronic Engineering Program, Faculty of Engineering, Universidad Militar Nueva Granada
Carrera 11 #101-80, Bogotá, Colombia
Email: robinson.jimenez@unimilitar.edu.co

1. INTRODUCTION

Both artificial intelligence and robotics have undergone exponential development in recent years, with various applications in several fields of knowledge. In consideration of [1], the usage of robots is quite common in the productive sector, but progress has been made recently in the service sector, within [2], [3] shows the impact produced for society and culture. Inside [3] the changes in the processes of communicative mediation between man and machines are perceived, considering hybrid options and the need for transdisciplinary work to find more friendly interfaces for man-machine communication.

Now multiple apps are available for services based in robots. Almeida [1] confirm that robots can benefit for the tourism. Actually digital agents of travels from organizations such as Expedia and Booking, are capable of using a pseudo natural language in its communication. García [2] shown applications in journalism, for example, inside north American media as The Washington Post, The Wall Street Journal, Forbes and convolutional neural network, and authors propose collaborative work man-machine similar to the one proposed within the present document. Misses *et al.* [4] exposes the construction to a robot que qualifies some parameters present in the human body for process of teaching-learning for a nonnative language. Inside [5] the use of robots in the health sector, oriented to repetitive or high-risk activities, which increased after the pandemic caused by COVID-19, where applications can be found in: telemedicine, healthcare, telehealth, online counseling, e-Health, home care, logistics, planning, coordination and communication, virus protection and sanitation, ongoing immunological assessment, job redesign and reorganization, psychological support, and continuing medical education. In the same sense, Santos *et al.* [6] indicates that robotics in this sector can be useful for both employees and customers, the first ones, for example, to avoid the execution of respective tasks, when carrying out surgeries precisely (with or without specialist supervision), as well as for controlling medication supply. The latter, for example, with the control of the administration of medications and by providing emotional support in psychological treatments.

Research by Castillo *et al.* [5], a service robot is defined as an entity with basic behaviors used for complex tasks in real environments capable of interacting through interfaces that can be graphical or natural (language, gesticulation, eye tracking, facial expression, emotional behavior, and language body). According to Pineda-de-Alcázar [3], human-machine interactions are analyzed at a communicative level. The ludibot created by [4] integrates verbal and non-verbal communication in human-robot interaction. The authors affirm that the implementation of robots must be supported by the considerations of the design of biomedical products and general criteria, such as users, usability, manufacturing, cycle product life, and safety guidelines for robots for personal use.

The World Health Organization (WHO) [7] reports 1.3 billion people have some major disability, that corresponds to 16% of the population. Emerging opportunities for the application of robotics to support healthcare may be reflected in reducing medical costs and increasing convenience for patient users. Many of them need assistance, according to [6], where they report that the most frequent use of robotics in human care is therapy and entertainment, and the most used means of interaction between humans and robots is verbal communication, which is useful in helping the elderly, children and people with a mental health disorder. Human-robot interaction is an important aspect in the introduction of robots in daily human activities.

According to Schneider *et al.* [8], the need to break the barrier of the use of robots from academia to the company is established under the use of social robotics. This field presents developments such as teaching robots for science explanations in museums [9], robots as receptionists [10], as assistants for manipulating an elevator [11], or for purchasing items [12]. In all these cases, the use of human-robot interaction in a practical way is evident, as in [13]. The assistance of robots to people in a support situation, such as that resulting from the COVID 19 pandemic, which requires them to be isolated, is vital if there are mobility limitations or the inability to pick up an object. Derived from the above, this work presents the development of a dialogue model for a care robot in a residential environment that interacts socially with a user. So, the aim is to assess the usability of the interactive man-robot communication interface designed to assist people with reduced mobility. Convolutional neural networks have recently been used to improve the inference of interrogative sentences in dialogues with robots [14], to maintain or follow a dialogue [15], [16] and in general to allow human-robot interaction by voice commands [17] or conversations [18]. This is why the dialogue model presented in this work uses convolutional networks to recognize words of affirmation or denial and action by the user, in actions of carrying and/or bringing specific objects.

The structure of this document consists of a state of the art in the present introductory section, focusing on the use of robots and human-robot interaction, in section 2 the methodology used is presented through a flow chart. Section 3 presents the results obtained in the design of the convolutional networks for the recognition of user orders and the dialogue model for the robotic command, which subsequently allows us to evaluate and present the usability model of the dialogue. Finally, the conclusions are presented.

2. METHOD

To achieve the project objective of evaluating a proposed model of conversation script between man and robot under prerecorded responses, a three-phase methodology was used. The first phase corresponds to the training scheme of its own convolutional network that allows the recognition of the user's base words for the execution of tasks by the robot. For first phase, the aim is to establish the training scheme for two convolutional neural networks, for which the starting point is to determine the basic actions that the robot can execute, and in this case, it will be to assist a person in conditions of reduced mobility or quarantine-type isolation in a residential environment. The robot's actions will be to carry and bring basic objects such as a towel, a glass or a medicine, being activated by the word “Robot” and confirming the commands desired by the user and interpreted by basic affirmation “Yes” or denial “No”. Considering what was pointed out by [2], artificial intelligence involves the interaction of a user with a robot that, since it is not completely human, requires mediation software, which transforms it into a more rational than emotional communicative process, with limited options. For second phase, a script model is established to guide how a user will ask the robot to carry or bring an object. For this purpose, a minimum interaction is sought in the form of a short dialogue guided by the robot. For third phase, human-robot interaction is evaluated through the dialogue model applied by a group of users. This corresponds to usability phase without the real robotic action, only conversational one. Which allows users to answer a survey about interaction where they indicate their perception on a Likert scale. Likert scales are a set of options in which individuals must choose their level of agreement with a series of statements. They are metrics commonly used in human-robot interaction to measure perceptions and attitudes [19]. Then, the Cronbach's Alpha value is calculated to verify the reliability of the survey, which according to [19] it measures the internal consistency of a scale; if its value is greater than 0.7, it is usually considered acceptable. Finally, an analysis of the answers obtained in each question is carried out using descriptive statistics.

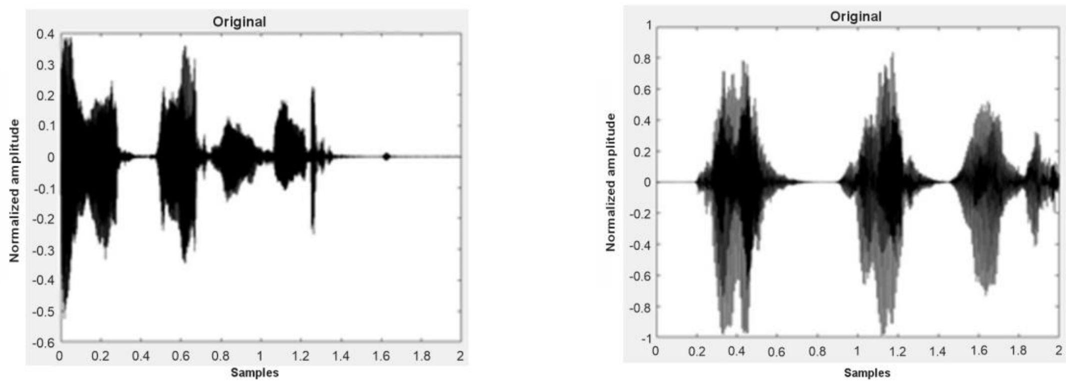
3. RESULTS AND DISCUSSION

3.1. First phase: Establishment of the training scheme of its own convolutional network that allows recognizing the user's base words for the execution of tasks by the robot

To recognize the group of selected words that allow the robot to be commanded, which are tabulated in Table 1, the design of a personalized architecture of a convolutional neural network presented in [20] is used, which has six convolution layers. For this network, an audio input discriminated in its mel frequency cepstral coefficients (MFCCs) components is used at a frequency of 16000 Hz. Sets of male and female voice commands are used in equal proportion; Figure 1 illustrates spectrograms of the “robot carry medicine” command, in Figure 1(a) the male voice and in Figure 1(b) the female voice.

Table 1. Group of words selected to command the robot

Action	Object
Robot	--
Stop	--
Traer	Papel/Toalla/ Vaso / Medicina
Llevar	Papel/Toalla/ Vaso / Medicina



(a) (b)

Figure 1. Command robot carry medicine for (a) female voice and (b) male voice [20]

For the recognition of the yes and no response commands, the same convolutional architecture of six convolution layers is used, but with a database based on these commands, where the response is validated only if the confidence level obtained at the output of the network is greater than 90%. The use of this separate network is due to the reduced time and parameter variation presented by yes/no responses, which are confused with phonemes of longer words and/or phrases when using the same network. Figure 2 illustrates the confusion matrices of both networks, in Figure 2(a) the results are presented for the commands and in Figure 2(b) for the responses, in both cases the high level of precision achieved can be observed.

3.2. Second phase: Generation of the dialogue model that will be carried out between the user and the possible interactions with the assistance robot

The second phase of the methodology corresponds to the dialogue model to be carried out between the user and the robot. To do this, the robot activation and deactivation conditions are established using the “robot” and “stop” commands respectively. The initial part corresponds to the robot's ability to navigate the environment to recognize the spaces in the residential environment in which it will interact, as described in

[21]. This stage is called “no inicializado”, that is, the robot must explore the residential environment to identify the spaces where it will move and assign each area a label, for example, bathroom 1, bedroom 2, this interaction corresponds to the dialogue in scene 1, as follows:

SCENE 1

ESTADO NO INICIALIZADO
USUARIO: Robot
ROBOT: Hola, ¿vamos a reconocer el ambiente? Contesta Si o No
USUARIO: Si
ROBOT: Reconocimiento finalizado, ¿genero un reporte? Contesta Si o No
USUARIO: No
Estado Reposo

SCENE 1.1

USUARIO: Robot
ROBOT: Reconocimiento finalizado, ¿genero un reporte? Contesta Si o No
USUARIO: Si
ROBOT: Encontré en el ambiente # espacios, espacio 1, …, Etc.
ROBOT: ¿Desea cambiar de nombre algún espacio? Contesta Si o No
USUARIO: No (go to scen 3)

(a) (b)

Figure 2. Confusion matrix of (a) commands and (b) answers

After that, the robot will allow you to reassign names to the spaces to facilitate interaction with the user in their residential environment. That is, the user will be able to reassign labels to the spaces found, for example, instead of bathroom 1, main bathroom, instead of bedroom 1, it would be Pedro's bedroom. This interaction corresponds to the dialogue in scene 2, as follows:

SCENE 2

USUARIO: Robot
ROBOT: Encontré en el ambiente # espacios, espacio 1, …, Etc.
ROBOT: ¿Desea cambiar de nombre algún espacio? Contesta Si o No
USUARIO: No (go to scen 3)

SCENE 2.1

USUARIO: Si
ROBOT: ¿Cuál es el nombre del espacio que vamos a cambiar? Escoge solo un espacio 1…... Etc.
USUARIO: Espacio #
ROBOT: ¿Cuál es el nuevo nombre?
USUARIO: Espacio Nuevo
ROBOT: Cambiaré el espacio # por el Espacio Nuevo, ¿seguro quieres cambiarlo? Contesta Si o No
USUARIO: No
ROBOT: ¿Desea cambiar otro nombre de espacio? Contesta Si o No
USUARIO: No
ESTADO NO INICIALIZADO
USUARIO: SI

The final and most recurring scene will be the usual interaction to execute carry and bring action commands, under the following script:

SCENE 3 (for any command)
ESTADO INICIALIZADO
*** USUARIO: Robot traer toalla… medicamento, papel…
ROBOT: Con gusto, indique ubicación
** USUARIO: Baño
ROBOT: Ubicación no reconocida, repita por favor **
ROBOT: ¿Usted desea traer toalla de baño… medicamento, papel…?
ROBOT: ¿es este comando válido? Si o no
USUARIO: Si
USUARIO: No
ROBOT: Por favor repita comando ***

3.3. Third phase: Measurement of the perception of various users regarding the usability of the interface

Taking into account that according to [4] within the considerations of the design of biomedical products, usability is counted as one of the evaluation criteria for robots for personal use. This criterion will be applied in this particular case, to evaluate the robot in assisting users in the established scenes. About, Ferreira *et al.* [22] affirms that usability is crucial as it is a characteristic of product quality by achieving its purpose with effectiveness, efficiency and user satisfaction; Its concept was popularized thanks to the standard ISO 9241-11:2018 [23]. Also, Albornoz *et al.* [24] points out that usability has become paramount for software quality. To evaluate the usability of the user's interaction with the robot, under the established dialogue model, the model in Figure 3 is proposed, which relates the three usability criteria defined by the standard ISO 9241-11:2018 [23] with the variables proposed.

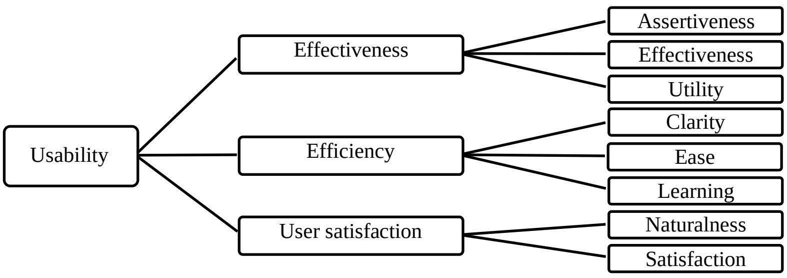


Figure 3. Robot user conversation usability evaluation model

For this, a preliminary evaluation instrument is generated that corresponds to a user survey that is applied after using the designed interface in a controlled environment. The survey seeks to know the perception of users on a Likert scale from 1 to 7, regarding efficiency, effectiveness and finally user satisfaction regarding the conversation with the robot. Additionally, the voice preference for the robot is investigated. Table 2 lists the questions, variables and criteria of the survey. This preliminary version of the instrument was applied to an initial sample of 10 people, obtaining the general results listed in Table 3.

The average execution and response time for the questionnaire was 10 minutes per person. The value of Cronbach's alpha is 0.7454, so it is considered that the internal consistency of the information collection instrument designed is adequate. The standard deviation indicates that the data are grouped and close to the mean value. It is observed that 75% of the questions tend to have a low concentration of values, that is, they have a platykurtic distribution because their kurtosis is negative and presents atypical data, while questions with positive kurtosis have a high concentration of values data showing a leptokurtic distribution. According to the coefficient of variation in the questions, the dispersion and variability is low because the value is less than 25%, the question that presents the greatest variability is question 4 and the one with the least variability is question 2, so its deduced that there is a relative homogeneity between the values obtained and the average is representative.

Based on these results, it was decided to adjust the evaluation instrument in accordance with the preliminary data, improving its wording, including in the informed consent, to be completed before its application, the possibility of withdrawing from the test at any time. The laboratory area test was conditioned in such a way that there was no interference in the sound, given the controlled training conditions of the convolutional network. On the other hand, the responses from the pilot test had ranges that did not exceed 3, so a scale of 1 to 7 is not useful, which is why the Likert scale is reduced from 1 to 5.

Table 2. List of questions, variables and usability criteria

Usability criteria	# ask	Variable
Efficiency	1	Clarity
Effectiveness	2	Assertiveness
User satisfaction	3	Naturalness
Effectiveness	4	Effectiveness
User satisfaction	5	Satisfaction
Efficiency	6	Ease
Effectiveness	7	Utility
Efficiency	8	Learning
Does not apply	9	Voice

Table 3. Descriptive statistics of the pilot test

	P1	P2	P3	P4	P5	P6	P7	P8
Half	6.3	6.7	6.2	5.3	5.9	6	6.5	5.9
Fashion	6	7	6	5	6	7	7	6
Variance	0.41	0.21	0.36	0.81	0.49	1	0.45	0.49
Deviation	0.6749	0.4830	0.6325	0.9487	0.7379	1.0541	0.7071	0.7379
Minimum	5	6	5	4	5	4	5	5
Maximum	7	7	7	7	7	7	7	7
Range	2	1	2	3	2	3	2	2
Kurtosis	-0.2830	-1.2245	0.1786	-0.3469	-0.7336	-0.4500	0.57143	-0.7336
Coefficient of variation	10.71%	7.21%	10.20%	17.90%	12.51%	17.57%	10.88%	12.51%

After applying the validated evaluation instrument with the adjustments described above, the average test time reduced to 5.7742 minutes per person. Table 4 presents the descriptive statistics of the first eight questions of the final survey. The overall average score obtained was 4.226 out of 5, so the degree of general satisfaction with respect to the usability of the proposed conversation script between a user and a robot for assistance, using pre-recorded responses, was 84.52%. Regarding the usability criteria, it is observed that the degree of satisfaction was 82.58%, 86.02%, and 84.09% in effectiveness, efficiency and user satisfaction, respectively. Only one of the eight questions obtained a satisfaction level of less than 80%, which was question 4, which corresponds to the effectiveness variable. This variable obtained the lowest score because the test does not show the robot executing the tasks. It is observed that the average values of all the questions are close to each other, evidencing that the data are close to said measure given the low deviation and therefore are relatively concentrated.

Table 4. Descriptive statistics of the final survey

	P1	P2	P3	P4	P5	P6	P7	P8
Half	4.4839	4.4516	4.4516	3.7419	4.0645	4.3226	4.1935	4.0968
Fashion	5	5	5	4	4	5	5	4
Variance	0.3143	0.4412	0.3122	0.7721	0.6410	0.6701	0.6077	0.6035
Deviation	0.5699	0.6752	0.5680	0.8932	0.8139	0.8321	0.7924	0.7897
Minimum	3	3	3	2	3	2	3	3
Maximum	5	5	5	5	5	5	5	5
Range	2	2	2	3	2	3	2	2
Kurtosis	-0.7180	-0.3246	-0.8124	-0.4375	-1.4704	0.4808	-1.2887	-1.3465
Coefficient of variation	12.71%	15.17%	12.76%	23.87%	20.02%	19.25%	18.90%	19.28%

With the adjustment of the instrument, a platykurtic flat distribution is obtained because in 87.5% of the questions the kurtosis is negative, having values that are far from the mean and do not allow a normal distribution. However, the coefficient of variation indicates that data are relatively grouped because the values returned are less than 24%, the question 1 has the greatest grouping of the data and the question 4 has the least grouping. The value of Cronbach's Alpha of the adjusted survey corresponds to 0.8031, this means there is good internal consistency in the information collection instrument designed, so it is reliable and the conclusions obtained from its results are valid. Additionally, with respect to question 9, which refers to the preference of the type of voice for the robot, it is observed that none of the respondents chose the option male, adolescent or child, 10% of the participants prefer the voice of an adult man, 29% of an adult woman and 61% of an adolescent woman, so this last result will be used for the final recordings of the dialogues of the proposed scenes.

Although according to [25], there is no unified model of human cognition and decision making, cognitive science can be very useful for the design of autonomous systems that interact with humans with tools such as: black box methods and theory-based methods, resource-rational decision making, cooperative cognition and rational speech act (RSA) theory, which could be explored in the future. As future research, it is suggested to carry out validations with other network architecture models, considering other feature extraction techniques such as a Fourier or Wavelet spectrogram that can give results in improving voice identification. It is also proposed to apply the final instrument to a large sample that allows the application of factor analysis models and the consideration of new evaluation variables and work on human-machine communication that includes the interpretation of gestures by the robot to assist people with reduced mobility. As well complementing the usability test with a satisfaction analysis, which turns out to be a typically subjective task.

4. CONCLUSION

It was possible to efficiently recognize, under a controlled environment, the group of words that allow the robot to be commanded, using a personalized architecture of a convolutional neural network with six convolution layers, which gives the possibility of being embedded in a large number of microcontrolled systems for devices portable, thus projecting future work in this aspect. A dialogue model was established between the user and the robot with conditions for activating and deactivating the robot, based on three scenes: the first seeks for the robot to explore and recognize the residential environment, the second allows the spaces to be renamed, and the third is intended to executing action commands to move elements. It was evident that this distribution allows the total interaction of the robot to be covered, from its initial use to general assistance. The tests show the data collection instrument is reliable and therefore the conclusions obtained from the analysis of the information collected are also reliable. A general satisfaction level of 84.52% was achieved, and greater than 82% for the usability criteria (effectiveness, efficiency, and user satisfaction). Concluding that the proposed usability measurement model can be replicated under similar application conditions. Based on the voice preference of the participants, the use of an adolescent woman's voice is defined for the final recordings of the dialogues used in the three proposed scenes. Which shows a degree of familiarity and friendliness of treatment sought by users.

ACKNOWLEDGEMENTS

Product derived from the research project titled "Design of a human-robot interaction model using deep learning algorithms" INV-ING-3971 financed by the vice-rector for research of the Universidad Militar Nueva Granada, year 2024.

REFERENCES

[1] M. M. A. Almeida, "Robots, inteligencia artificial y realidad virtual: una aproximación en el sector del turismo," *Cuadernos de Turismo*, no. 44, pp. 13–26, Nov. 2019, doi: 10.6018/turismo.44.404711.

[2] I. A. S. García, "Los robots y la inteligencia artificial. nuevos retos del periodismo," *Doxa Comunicación. R Revista Interdisciplinar De Estudios De Comunicación Y Ciencias Sociales*, no. 27, pp. 295–315, Dec. 2018, doi: 10.31921/doxacom.n27a15.

[3] M. Y. Pineda-de-Alcázar, "Inteligencia artificial y modelos de comunicación," *Razón y Palabra*, vol. 21, no. 4_99, pp. 332–346, 2017.

[4] M. A. O. Misses, H. S. Ochoa, and A. S. López, "Ludibot: Interfaz humano-robot móvil para el aprendizaje lúdico de idiomas," *Ingeniería Investigación y Tecnología*, vol. 22, no. 3, pp. 1–10, Jul. 2021, doi: 10.22201/fi.25940732e.2021.22.3.021.

[5] M. E. R. Castillo, A. J. F. Murillo, and C. R. Cr. Mendoza, "Consideraciones en el diseño de robots para la atención médica en el mundo post COVID-19," *Cultura Científica y Tecnológica*, vol. 19, no. 1, 2022, doi: 10.20983/culcyt.2022.1.3.3.

[6] N. B. Santos, R. S. Bavaresco, J. E. R. Tavares, G. D. O. Ramos, and J. L. V Barbosa, "A systematic mapping study of robotics in human care," *Robotics and Autonomous Systems*, vol. 144, Oct. 2021, doi: 10.1016/j.robot.2021.103833.

[7] WHO, "Disability," *World Health Organization*, 2023. Accessed: Sep. 07, 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/disability-and-health

[8] S. Schneider, W. Clas, and D. Brscic, "Human-robot interaction in public spaces," in *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Mar. 2022, pp. 1287–1289, doi: 10.1109/HRI53351.2022.9889525.

[9] J. Park *et al.*, "User perception on personalized explanation by science museum docent robot," in *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Mar. 2022, pp. 973–975, doi: 10.1109/HRI53351.2022.9889654.

[10] M. Moujahid, B. Wilson, H. Hastie, and O. Lemon, "Demonstration of a robot receptionist with multi-party situated interaction," in *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Mar. 2022, pp. 1202–1203, doi: 10.1109/HRI53351.2022.9889314.

[11] W. Law, K. Li, K. Fan, T. Mo, and C. Poon, "Friendly elevator co-rider: An HRI approach for robot-elevator interaction," in *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Mar. 2022, pp. 865–869, doi: 10.1109/HRI53351.2022.9889614.

[12] A. Yamazaki, A. Krummheuer, and M. Imai, "Interdisciplinary explorations of processes of mutual understanding in interaction with assistive shopping robots," in *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Mar. 2022, pp. 1293–1295, doi: 10.1109/HRI53351.2022.9889352.

[13] L. Takayama, "Putting human-robot interaction research into design practice," in *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Mar. 2022, p. 1, doi: 10.1109/HRI53351.2022.9889659.

[14] K. Kawai, R. Rzepka, and T. Nemoto, "Using convolutional neural network for improving inference of interrogative sentences in a dialogue system," in *2022 10th International Conference on Affective Computing and Intelligent Interaction Workshops and*

Demos (ACIIW), Oct. 2022, pp. 1–4, doi: 10.1109/ACIIW57231.2022.10086010.

[15] S. Liu, S. Liu, and W. Xu, "Gated attentive convolutional network dialogue state tracker," in *2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 6174–6178, doi: 10.1109/ICASSP40776.2020.9054225.

[16] H. Shi, T. Ushio, M. Endo, K. Yamagami, and N. Horii, "A multichannel convolutional neural network for cross-language dialog state tracking," in *2016 IEEE Spoken Language Technology Workshop (SLT)*, Dec. 2016, pp. 559–564, doi: 10.1109/SLT.2016.7846318.

[17] H.-A. Rusan and B. Mocanu, "Human-computer interaction through voice commands recognition," in *2022 International Symposium on Electronics and Telecommunications (ISETC)*, Nov. 2022, pp. 1–4, doi: 10.1109/ISETC56213.2022.10010253.

[18] M. Saloumi *et al.*, "Speech emotion recognition using one-dimensional convolutional neural networks," in *2023 46th International Conference on Telecommunications and Signal Processing (TSP)*, Jul. 2023, pp. 212–216, doi: 10.1109/TSP59544.2023.10197766.

[19] M. L. Schrum, M. Johnson, M. Ghuy, and M. C. Gombolay, "Four years in review," in *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, Mar. 2020, pp. 43–52, doi: 10.1145/3371382.3380739.

[20] R. Jiménez-Moreno and R. A. Castillo, "Deep learning speech recognition for residential assistant robot," *IAES International Journal of Artificial Intelligence*, vol. 12, no. 2, pp. 585–592, 2023, doi: 10.11591/ijai.v12.i2.pp585-592.

[21] J. J. Herrera, R. Jimenez-Moreno, and J. E. M. Baquero, "Virtual environment for assistant mobile robot," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 6, pp. 6174–6184, Dec. 2023, doi: 10.11591/ijece.v13i6.pp6174-6184.

[22] J. M. Ferreira *et al.*, "Impact of usability mechanisms: An experiment on efficiency, effectiveness and user satisfaction," *Information and Software Technology*, vol. 117, Jan. 2020, doi: 10.1016/j.infsof.2019.106195.

[23] ISO, "Ergonomics of human-system interaction - Part 11: Usability: Definitions and concepts," *International Standardization Organization*, 2018. [Online]. Available: https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en

[24] D. A. Albornoz, S. A. Moncayo, S. Ruano-Hoyos, G. E. Chanchí-Golondrino, and K. Márceles-Villalba, "Sistema software para la ejecución de pruebas de usabilidad bajo el enfoque de mouse tracking," *TecnoLógicas*, vol. 22, pp. 19–31, Dec. 2019, doi: 10.22430/22565337.1511.

[25] M. K. Ho and T. L. Griffiths, "Cognitive science as a source of forward and inverse models of human decisions for robotics and control," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 5, no. 1, pp. 33–53, May 2022, doi: 10.1146/annurev-control-042920-015547.

BIOGRAPHIES OF AUTHORS

Robinson Jiménez Moreno is an electronic engineer graduated from Universidad Distrital Francisco José de Caldas in 2002. He received a M.Sc. in Engineering from Universidad Nacional de Colombia in 2012 and Doctor in Engineering at Universidad Distrital Francisco José de Caldas in 2018. His current working as Associate Professor of Universidad Militar Nueva Granada and research focuses on the use of convolutional neural networks for object recognition and image processing for robotic applications such as human-machine interaction. He can be contacted at email: robinson.jimenez@unimilitar.edu.co.

Anny Astrid Espitia Cubillos performed her undergraduate studies in industrial engineering in the Universidad Militar Nueva Granada in 2002 and master in industrial engineering from the Universidad de Los Andes in 2006. She is an Associate Professor on Industrial Engineering Program at Universidad Militar Nueva Granada, Bogotá, Colombia. She can be contacted at email: anny.espitia@unimilitar.edu.co.

Esperanza Rodríguez Carmona was born in Pereira (Colombia). She performed her undergraduate studies in mechanical engineering (1997) from the Universidad Tecnológica de Pereira (UTP) and master's degree in university teaching from the Universidad de La Salle, Bogotá, Colombia. She is an Associate Professor on Industrial Engineering Program at Universidad Militar Nueva Granada, Bogotá, Colombia. She can be contacted at email: esperanza.rodriguez@unimilitar.edu.co.