Journal of the Brazilian Computer Society, 2026, 32:1, doi: 10.5753/jbcs.2026.5745

Breaking Communication Barriers: Key Requirements for User Interaction with Oral-Sign Language Translation Systems

Natália Sales Santos [**Universidade Federal de Minas Gerais** | ***nattes@gmail.com***]
Lucas Almeida S. de Souza [**Universidade Federal de Minas Gerais** | ***lucasalmeidas.souza@gmail.com***]
Julia Manuela G. Soares [**Universidade Federal de Minas Gerais** | ***juliamanu@ufmg.br***]
Raniere Alislan A. Cordeiro [**Universidade Federal de Minas Gerais** | ***ranierealislan@ufmg.br***]
Elidéa Lúcia Almeida Bernardino [**Universidade Federal de Minas Gerais** | ***elidea@ufmg.br***]
Mario Fernando Montenegro Campos [**Universidade Federal de Minas Gerais** | ***mario@dcc.ufmg.br***]
Raquel Oliveira Prates [**Universidade Federal de Minas Gerais** | ***rprates@dcc.ufmg.br***]

Computer Science Department, Universidade Federal de Minas Gerais, Av. Antônio Carlos, 6627 – Prédio do ICEx – Pampulha, Belo Horizonte - MG, 31270-901, Brazil.

Received: 25 March 2025 • **Accepted:** 26 March 2026 • **Published:** 26 May 2026

Abstract Communication between deaf and hearing people is still a major challenge, especially in critical moments such as medical emergencies, where information needs to be transmitted clearly and quickly. In this context, for the development of efficient machine translation systems, it is essential that they meet the real needs of both deaf people and health professionals. This article identifies and defines fundamental requirements for the development of such systems, taking into account the experience of both groups. In addition to mapping these requirements, an evaluation was carried out to validate their relevance and ensure that they correspond to the real demands of users. The work is part of the project "Captar-Libras: Video communication system for the deaf applied to pre-medical care" (In Portuguese: *"Captar-Libras: Sistema de Comunicação por vídeos para surdos aplicado ao pré-atendimento médico")*, which develops a bidirectional translation solution between Portuguese and Brazilian Sign Language (Libras) using photorealistic avatars, with a focus on healthcare. The results of this article offer guidelines for the development of more accessible and efficient systems, promoting more inclusive communication between deaf and hearing people.

Keywords: Automatic translation, Brazilian sign language, Libras, Portuguese, deaf user, accessibility, healthcare

1 Introduction

According to the World Health Organization [WHO, 2021], approximately 900 million people worldwide may develop deafness by 2050. In Brazil, the Health Ministry estimates that this condition affects more than 2.4 million people [Ministério da Saúde, 2019], with varying degrees of severity. This increase in hearing loss is partly explained by the aging of the population, a global trend. This scenario underscores the urgent need to develop effective measures to support and include deaf people. Although technologies such as hearing aids and lip reading training are available to orally-educated deaf, they are not sufficient to transpose the communication barriers involved.

Among the numerous barriers faced by the deaf community, access to healthcare has been reported as a major challenge in many countries (e.g. [Rogers *et al*., 2025; Izquierdo-Condoy *et al*., 2024; Beaver and Carty, 2021; Barnet, 2021; Wheatley, 2021]), including in Brazil [Bond, 2019; Rezende *et al*., 2021; Francisco and Mesquita, 2023]. Effective communication between healthcare professionals and patients is essential to ensure adequate patient-centered healthcare services. However, deaf individuals often face significant barriers when accessing these services, due to the scarcity of interpreters fluent in sign language and the lack of adequate translation resources [Rogers *et al*., 2025; Rezende *et al*., 2021; de Souza *et al*., 2020; Steinberg *et al*., 2006; Soares *et al*., 2024]. These communication gaps may result in misdiagnosis, inadequate treatment, and even iatrogenesis, jeopardizing the health and well-being of deaf people.

Automatic translation between oral and sign languages emerges as an important research topic, aiming at facilitating communication between deaf and hearing individuals. In this direction, Naranjo-Zeledón *et al*. [2019] have conducted a systematic study on sign language translation technologies in which they analyzed the investigated topics, publication frequency, and specific approaches, highlighting the importance of addressing knowledge gaps in those technologies. Their results indicate that automatic translation systems are a promising solution to enhance communication between deaf patients and healthcare professionals.

Currently, within the Brazilian context, there are systems available that automatically translate Portuguese into Brazilian Sign Language (Libras) through virtual avatars (e.g. Handtalk[1], Rybená[2], VLibras[3], etc). However, these systems do not fully fulfill the requirements to allow for adequate communication between deaf patients and healthcare professionals [Soares *et al*., 2024]. The first limitation of such systems is that they only offer a unidirectional translation

[1] https://www.handtalk.me/br/
[2] https://rybena.com.br/
[3] https://www.vlibras.com.br/

from Portuguese into Libras. Second, their representation of facial expressions is very limited which in many cases compromises the translation. Thirdly, they are not well suited for healthcare settings, as they present various translation errors in this context, many of which are critical.

The work presented in this article was developed within the scope of the *"Captar-Libras Project: A video-based communication system for deaf individuals applied to pre-medical care"*, which aims to create a bidirectional, automatic translation system between Brazilian Portuguese and Libras. In one direction, the system translates the video of a person signing in Libras into Brazilian Portuguese, and from a health care professional speaking in Brazilian Portuguese into Libras by a signing photorealistic avatar. As such, the system aims to facilitate communication in healthcare settings. It is important to emphasize that this article is an extension of the study presented by Santos *et al*. [2024] at the XXII Brazilian Symposium on Human Factors in Computational Systems (IHC 2024). The previous article discussed the study conducted to define requirements for designing the interface and the interaction of the proposed system. Based on these requirements, a horizontal prototype was developed. Those requirements were useful in guiding the decisions regarding the design of the interaction of the prototype. The prototype represents a proof-of-concept of how the requirements can be applied, and allows for an initial formative evaluation of the interface, and, consequently, an initial assessment of the set of requirements.

This extended article includes the evaluation of the prototype conducted with the participation of real end-users, including deaf individuals and healthcare professionals. The goals were to identify opportunities for prototype improvement, and to analyze which indicators they generate about the set of requirements. The results of the user evaluation conducted provided valuable insights for improving the system's interface, aiming to better support the communication between patients and doctors. They also generated relevant indicators towards consolidating an initial set of requirements.

It is worth pointing out that other works share the objective of developing automatic oral-to-sign language translation systems using photorealistic avatars [Guo *et al*., 2023; Sosa-Jiménez *et al*., 2022]. However, technical solutions developed in other countries cannot be directly imported for use in the Brazilian context due to differences in respective oral and sign languages. Even in other countries that have Portuguese as their official language, such as Portugal, their sign languages are different [David *et al*., 2024]. The few works that investigate automatic translation in the healthcare domain for the Brazilian context focus on system architecture and translation issues [da Silva *et al*., 2020, 2024], and do not address their interface or interaction.

This article is structured into ten sections. Section 2 discusses the related work, providing an overview of the literature in the field related to (i) sign-oral language translations and (ii) deaf user interaction. Section 3 presents an overview of the Captar-Libras Project within which this work was developed. Section 4 brings an overview of the Methodology adopted in our study, and the ethical considerations taken throughout the process. Section 5 describes the workshops conducted which produced a set of requirements and considerations for the system's interface and interaction design, which are presented in Section 6. Section 7 describes the horizontal prototype developed based on the requirements, which was then evaluated with representatives of the intended end-users, as described in Section 8. Finally, Sections 9 and 10 provide an in-depth discussion of the results,the learned lessons,and future work directions.

2 Related Work

Communication between deaf and hearing individuals in healthcare environments is a challenge that has been addressed from various perspectives in research. On the one hand, there are studies aimed at developing technologies capable of automatically translating sign language into oral language, and vice versa, using machine learning, gesture recognition and virtual avatars. The goal of these studies is to facilitate communication between deaf patients and healthcare professionals, effectively overcoming linguistic barriers.

On the other hand, there is also significant research focused on investigating how deaf individuals interact with interactive systems. These studies explore usability aspects, preferences of the deaf community, and the necessary adaptations to ensure that technologies are accessible and functional for everyone. It is not enough to create translation tools. They must provide users with an intuitive and positive experience, and meet the specific needs of the deaf community.

In this section, the main studies related to the topic are presented, divided into two categories: (1) studies on automatic translation between sign and oral languages, exploring methods for recognizing and synthesizing sign languages in healthcare contexts, and (2) research on the interaction of deaf individuals with interactive systems, analyzing users' perceptions of these technologies and making suggestions to improve their interface and user experience.

2.1 Sign and Oral Language Translation

Automatic sign language translation has been increasingly attracting researchers' attention, as highlighted by Farooq *et al*. [2021]. Several different contexts have been considered (e.g. electronic commerce [Rocha *et al*., 2020], serious games in education [Escudeiro *et al*., 2015], teaching and learning [Oliveira *et al*., 2019]), but this section presents studies aimed at developing automatic translation systems in healthcare contexts. These studies approach unidirectional and/or bidirectional sign language translation [da Silva *et al*., 2020; Areeb and Nadeem, 2021; Xia *et al*., 2022; Sosa-Jiménez *et al*., 2022; da Silva *et al*., 2024; Das *et al*., 2024]. However, some of these solutions require additional sensors or hardware, such as gloves or armlets, complicating their application in real-world scenarios, as seen in Hisham and Hamouda [2019]; Dere *et al*. [2022]; Guo *et al*. [2023]. Although many of these studies achieved significant advancements in translation, most focus solely on translation, not addressing requirements related to interaction or interface design.

The work of Areeb and Nadeem [2021] proposes an innovative model combining a pre-trained convolutional neural network (VGG-16) with a Long Short-Term Memory (LSTM) network to recognize manual gestures in Indian Sign Lan-

guage (ISL). This model identifies both static and dynamic gestures in videos, leveraging the capacity of convolutional networks for spatial learning and the LSTM networks for temporal learning. The combination of these technologies results in a system suitable for emergency situations, when deaf individuals need assistance or medical consultation. However, the model presents some limitations, including the need for robust and diverse datasets for broader applicability and system accuracy. Additionally, certain gestures are challenging to detect and differentiate, such as the signs for "pain" and for "ask for help" in ISL, due to similarities in hand movements.

A significant advancement is the Heart-Speaker system, proposed by Xia *et al*. [2022], representing an innovation in sign language recognition applied to medical consultations with deaf patients. This system integrates sign and spoken language recognition, enabling bidirectional conversion. Developed with a focus on portability and efficacy in medical treatment for deaf patients, the device still faces challenges, such as ensuring accurate real-time sign language recognition during medical consultations, as well as handling the complexity of a healthcare environment, which may introduce varying levels of noise. Although developed specifically for healthcare environments, this study was conducted from a Chinese perspective and would require adaptations for Brazilian contexts both linguistically and regarding healthcare environments, which may differ significantly.

The work of Sosa-Jiménez *et al*. [2022] proposes a real-time Mexican Sign Language (LSM) recognition and synthesis system, destined to facilitate communication between deaf and hearing individuals during medical consultations. The system utilizes Hidden Markov models for sign recognition and an animated avatar for dynamic synthesis of LSM. This 3D avatar, representing a virtual human interpreter, is capable of bidirectional translation, enhancing communication with deaf individuals. The system recognizes healthcare-related signs, letters and numbers, demonstrating positive results regarding translation accuracy and efficiency. Nevertheless, there remains a lack of practical and detailed guidelines for developing comprehensive systems in this area.

The work of Das *et al*. [2024] proposes a gesture recognition system in Indian Sign Language (ISL) to detect medical-related signs in real-time. The system employs an integrated model of convolutional neural networks and Long Short-Term Memory (LSTM) to recognize words, along with an attention network to form phrases from recognized words. The goal is to facilitate communication with deaf and hard of hearing patients in medical consultations. The system achieved an 100% accuracy in detecting three different medical-related signs in ISL. However, some limitations include scarcity in available ISL datasets and difficulties in recognizing dynamic gestures.

In the Brazilian context, there are some initiatives for automatic translation between Portuguese and Brazilian Sign Language (Libras) in healthcare settings. da Silva *et al*. [2020] proposed a Libras recognition model for use in healthcare contexts. They developed a deep learning architecture that does not require sensor or external equipment, relying exclusively on images and videos. Despite achieving significant accuracy in their internal tests, they still face challenges such as the scarcity of data specifically designed for Libras signing recognition, which may affect the translation accuracy. In a later work, da Silva *et al*. [2024] present advancements by refining the model through the implementation of a multiple-flow architecture for Libras sign recognition. In this approach, 3D convolutional neural (I3D) networks and Long Short-Term Memory (LSTM) networks were employed to process spatiotemporal characteristics. The model incorporates distinct flows, each designed to capture specific visual attributes of health-related Libras signs. The results obtained were highly accurate, with an average of 99.80% accuracy, precision, recall, and F1-score in Libras sign classification.

2.2 Deaf Users' Interaction

In addition to research on automatic translation, some studies analyzed deaf users' interaction with these systems, investigating aspects such as preferences within the deaf community, interface usability and challenges in adopting assistive technologies. However, most studies do not address systems specifically designed for healthcare.

The study by Tran *et al*. [2023] investigated preferences from the perspective of the deaf community in the United States regarding automatic sign language translation systems. For this, a survey was conducted on application domains, performance, interface aspects, and concerns, receiving 32 responses from individuals representing the deaf community. Among the results, the domain in which the participants missed a translator the most was healthcare. They also indicated a preference for realistic avatars over animated models.

The study by Kipp *et al*. [2011] addresses the perceptions of the deaf community about sign language virtual avatars, highlighting their potential to improve the accessibility of written content. The research combined two primary methods: focus groups and accessible online study. Eight participants engaged in in-depth discussions in the focus groups, while the online study collected quantitative data from 317 respondents. The overall results indicated a relatively positive acceptance of avatars, but also reported significant concerns about their effectiveness and the risk of replacing human professionals. However, the study concentrates exclusively on general perceptions of using sign language avatars.

On the other hand, the study by Süzgün *et al*. [2015] analyzes deaf users' interaction with a healthcare-oriented system. The study designed an interactive interface to facilitate communication of deaf individuals in hospital environments. The system employs real-time Turkish sign language recognition and guides users through questions and response options, aiming to reduce the need for human interpreters. The system's interface was tested with five participants across five distinct scenarios, evaluating usability and functionalities. Participants positively rated their experience, highlighting the ease of use and utility for hospital communications. Despite promising results, the study was limited by the small number of participants and did not explore additional variables that might influence broader system adoption.

Therefore, more research is needed to provide specific guidelines for developing bidirectional translation systems including users' perspectives in healthcare contexts. This article intends to fill this gap by presenting a set of requirements for interface and interaction development in Portuguese-Libras

bidirectional translation systems using photorealistic avatars, aiming to enhance communication in healthcare services.

3 Project Captar-Libras - Overview

This section briefly presents the project within which the research described in this article was conducted. The *Captar-Libras Project: A video-based communication system for deaf individuals applied to medical pre-consultation* is a project that involves research, technological development, and innovation, with a special emphasis on assistive technology aimed at the deaf community in the healthcare context. This project is being developed by the Universidade Federal de Minas Gerais (UFMG), in partnership with the startup company SignumWeb Comunicação Inclusiva Ltda[4]. The project is inherently interdisciplinary and is coordinated by the Computer Science Department in collaboration with the Faculty of Medicine and the Faculty of Languages, Literature, and Linguistics. The team is composed of over 30 people, including researchers, graduate and undergraduate students, and lab technicians in the fields of computer science, linguistics (with expertise in Libras), medicine, and staff of SignumWeb, the partner company. The team includes deaf, hearing, and bilingual (Portuguese and Libras) individuals. This diversity of knowledge, expertise, and experiences has proven to be essential for the project's development, fostering a comprehensive and integrated understanding of the challenges and solutions required.

The goal of the project is to investigate new techniques and approaches to the problem and to develop a prototype of a computer system designed to assist in-person medical care. Communication in such settings represent a major challenge for deaf patients and hearing doctors, as described in the literature and as experienced and reported by several deaf members of our team. The proposed solution is an automatic, real-time, bidirectional Libras-Portuguese translation system. It captures the signing in Libras and translates it into Brazilian Portuguese, and in the other direction, translates Brazilian Portuguese to Libras through the use of a photorealistic avatar, including facial expressions, which are essential in Libras communication [de Menezes and Feitosa, 2015], thus simulating a human interpreter. The translation is performed in real time, eliminating the need for human interpreters in many situations.

The project is funded by Finep[5] and the expected outcome is a functional prototype that can be evaluated with final users. The prototype and the know-how produced will then be transferred to SignumWeb (the partner company), who will then be able to further develop it and turn it into a commercial product ready to be deployed in real contexts.

In addition to the development of the prototype, the project required the creation of a video dataset in Libras, with each video's respective translations to Brazilian Portuguese, containing terms and sentences commonly used in healthcare contexts. These sentences were signed by a large number of individuals to account for differences in anatomy, gender, skin color, and age. This dataset is being used to train machine learning models that have been developed to perform translation and to synthesize the photorealistic avatar. The project also plans to make this dataset available to other researchers as part of its contributions.

Within the field of computer science, the project primarily involves research in Computer Vision, Artificial Intelligence (AI) and Human-Computer Interaction (HCI). In HCI, the project faces the challenge of creating an interface that can be easy to understand and use by both deaf patients and hearing doctors. The focus of this paper is on the HCI aspects of the research being conducted, and more specifically in eliciting the requirements for the interface for Libras-Portuguese translation systems, and consolidating them through the application in the prototype developed and its evaluation with potential real users of the system.

4 Methodology

To identify the requirements for the prototype, the HCI team[6] decided to conduct the first part of the study with members of the project team themselves rather than with end users. This decision was based on several factors: (1) the photorealistic avatar solution, displayed through computers in a healthcare setting, had already been defined based on the studies and discussions prior to the project's submission; (2) the team included members who represented the profile of potential users – healthcare professionals and deaf patients; and (3) the project's technical team had already identified certain technical limitations that needed to be addressed at the interface level.

Therefore, for the purpose of identifying the requirements, the HCI team decided to conduct workshops with team members that represented our intended users (healthcare professionals and deaf patients). The workshop included an activity to generate paper-based interface proposals, which allowed participants to think about the elements that would be needed and their use in the pre-medical care context, and discuss their rationale for their proposal. Next, we describe the methodology adopted in this study, which was structured into four steps, as shown in Figure 1.

In step 1, workshops were held with project members, who acted as representatives of the end users. These workshops aimed to understand the needs and expectations of users and to collaboratively create proposals for the system. In section 5 we explain in detail how the workshops were conducted.

In step 2, the HCI team performed a detailed analysis of the material generated during the workshops, including the interface designs, the participants' explanation for their design decisions registered in video, and notes taken during the discussions. The analysis process was iterative, involving the review and contrast of the different proposals created. During this step, potential requirements for the system were identified, taking into account the decisions and justifications

[4]Website: `https://signumweb.com.br/`

[5]FINEP is a public company of the Ministry of Science, Technology and Innovation (MCTI) that finances Science, Technology and Innovation studies and projects. See `http://www.finep.gov.br/` .

[6]The HCI team in the Captar-Libras project consists of a senior researcher in HCI, a specialist with master's degree in HCI, and two undergraduate Computer Science students who have taken the course and have been working in the field.

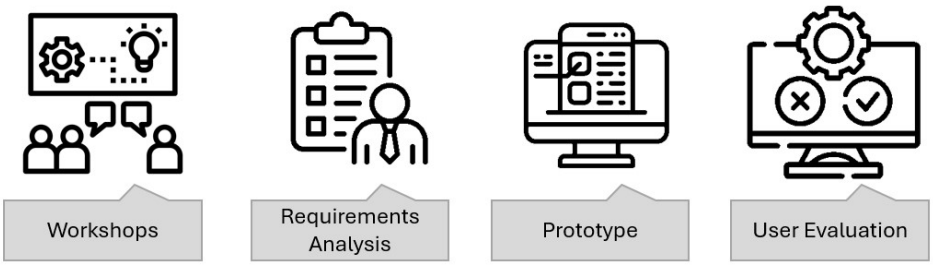


Figure 1. Methodology steps.

of the participants.

In step 3, an interactive horizontal prototype was developed based on the identified requirements. During the design of the prototype we explored different alternatives for the most critical points of the interface. The prototype was designed to simulate user interaction to allow for a formative evaluation of the interface with potential users.

In step 4, the prototype was evaluated with real users. This evaluation aimed to identify inconsistencies, assess the clarity of the interface elements, and verify whether the prototype adequately reflected the requirements defined in step 2.

This methodology allowed for two main results to be achieved: (1) the identification of the requirements, through the workshop, and their consolidation based on users' perspectives during the evaluation; (2) the design of the interface for the translating system based initially on the requirements, and adjusted according to users experiences and feedback.

4.1 Ethical Considerations

From the very beginning of the project, the research team discussed the relevance of the ethical issues of the project and how to address them. In this section, we present an overview of the ethical considerations and precautions taken during the project, but focus on those related with the steps of the methodology described.

The first issue considered was the need for the direct participation of deaf people in the team, as has been pointed out in the literature [De Meulder, 2021], as well as expressed in the "*Nothing about Us, Without Us*" motto[7], as well as for an interdisciplinary team that could encompass the different knowledge fields necessary for the project. Thus, as described in the previous section, an interdisciplinary team consisting of both deaf and hearing researchers, students and company members, with expertise in computer science, deaf culture and translation, and medicine was brought together for the project. Additionally, Libras-Portuguese interpreters were included to ensure efficient communication among all team members.

From the project proposal and throughout all of its development, the team was concerned and paid attention to ethical issues involved in each and every step taken. Furthermore, the project was submitted to the Research Ethics Committee of the university and approved[8]. The project described all of the steps planned and described the initial data collection and methodology. The subsequent steps that required data collection were described in terms of their goals and identified as future steps. It was indicated that the detailed description of the methodology and specific instruments for these future steps would be submitted as amendments once the project evolved and they were fully defined. It was explicitly stated that all materials presented to participants – such as consent forms, terms of use, invitations, explanations, and data collection instruments – would be presented in both Portuguese and Libras to ensure full comprehension by all participants.

One aspect that is worth pointing out is that for some of the activities proposed, the project planned for them to be conducted both by team members, as well as volunteer research participants [9]. Thus, before submitting the project, the team discussed with the Chair of the Research Ethics Committee how to handle these activities. It was clarified that only activities that involved volunteer participants should be classified as research with human subjects. Thus, in relation to the steps presented in this article, the workshop was conducted with team members who represented end-users, and the user evaluation was conducted with volunteer participants in the research.

In step 1, 3 workshops were conducted, and 6 members from the project team, besides the HCI team, participated in them, 2 in each workshop. They were asked to design an interface (as will be detailed in the next section), and present to the HCI team members their proposal and rationale for them. With their consent, the video of their presentation was recorded to support the analysis step.

In step 4, the user evaluation was conducted with eight deaf people and two medical students, all as volunteer participants.

Participants were recruited through convenience sampling, leveraging team member's personal networks and by use of the snowball sampling[10]. Deaf participants should not have

[7]The motto relies on the principle of the full participation and equalization of opportunities for, by and with persons with disabilities, especially in policies, services, projects, etc aimed at them. It has been echoed globally, as can be seen in: UN (`https://www.un.org/esa/socdev/enable/iddp2004.htm`),Wikipedia (`https://en.wikipedia.org/wiki/Nothing_about_us_without_us`) or in Inclusive News in Brazil (`https://inclusivenews.com.br/?p=33637`).

[8]The project was approved in October 2023 (CAAE: 71623323.8.0000.5149), and the amendment for the user evaluation of the prototype was approved in July 2024.

[9]The Brazilian law (Resolução 466/2012 available at: `https://www.gov.br/conselho-nacional-de-saude/pt-br/acesso-a-informacao/legislacao/resolucoes/2012/resolucao-no-466.pdf/view`) requires all research participants to be volunteers – they may be refunded for any expenses related to their participation, but not paid for their participation.

[10]The snowball sampling technique is a method for recruiting eligible participants to take part in a research study through the recommendation of

any previous knowledge of the prototype, therefore, team members were not considered for this evaluation.

Before the evaluation, it was explained to all participants its goals, what data would be collected, and how they would be used to inform the research and the development of the system. A consent form was presented to them, in Libras for deaf participants and in Portuguese for hearing participants. They all agreed freely to participate. They were told that they could interrupt the activities and change their minds about participating at any given moment, without any prejudice to them.

5 Workshops

As mentioned in the previous section, in **Step 1**, three workshops were conducted, each with two team members: two workshops included team members who were deaf, and one included members from the medical team. The first workshop had two members from the partner company, one of them with a degree in Information Systems, and the other in the Humanities; Workshop 2, had two team members from the Faculty of Languages, Literature, and Linguistics (a professor and an undergraduate student); and Workshop 3, had two team members who are from the Faculty of Medicine (a professor and an undergraduate student). In each workshop, one participant was male and one was female. Workshops 1 and 2 included a Libras interpreter. Furthermore, some other team members attended as observers: in Workshop 1, an undergraduate student from the technical team, and in Workshop 2, another professor from the Faculty of Languages, Literature, and Linguistics, who could hear and was fluent in both Portuguese and Libras.

During the workshops, participants were invited to design an interface on paper that they considered appropriate for the system from the perspective of the user they represented (either the deaf patient or the hearing doctor). In order to create the interface they were handed printed images of blank desktop screens to represent the interaction device, as well as geometric figures – such as rectangles of different sizes – to represent elements on the screen (e.g., buttons, windows, dialog boxes, etc.) (see Figure 2a). Note that no labels were assigned to these elements, allowing participants to associate them with the sizes and shapes they considered most suitable. In addition, colored post-its, pencils, pens, markers, scissors, glue, and tape were also made available (see Figure 2b), enabling participants to draw, write, or include whatever they thought was necessary for the interaction. In each workshop the participants were given the choice to design an individual interface proposal, or work together and generate a single proposal.

In Workshop 1, participants generated a single interface, and in the following workshops, each participant designed their own interface, thus resulting in a total of 5 distinct proposals. After designing their interface, each author was invited to present their solution.

During these presentations, the other team members present were able to clarify their understanding of the proposed interface, their rationale for the decisions made, and discuss issues related to technical demands or constraints, and how these were or could be addressed.

The workshops were held on distinct days during November 2023, each lasting approximately 1 hour and 30 minutes. In the next section, we briefly present the proposals and discussion outcomes from each workshop.

5.1 Workshop 1 - Perspective of Deaf Patients

In this workshop, participants worked together and presented a single proposal containing four different screens, shown in Figure 3: a welcome screen, a translation screen with elements displayed on demand (e.g., a chat) - represented in two screens at different moments (see Figures 3b and 3c), and a closing screen in which patients can provide their feedback. During the screen presentations, participants mentioned points and elements they considered important for the interaction.

The first point addressed by the participants was the possibility of including and customizing settings in the interface, such as adjusting the chat font size, subtitles, and the doctor's audio volume, for better accessibility, as illustrated in Figure 3b. Additionally, in terms of communication, they highlighted the importance of having a fallback mechanism in case of failures. Thus, the inclusion of a chat was proposed (see Figures 3b and 3c). The chat could be opened on demand, allowing both the deaf patient and the doctor to type messages. By default, this window would remain minimized and could be expanded or closed as needed by the user. Participants also indicated that the inclusion of control buttons could improve the quality of interaction, allowing users to start and stop movement and audio capture. According to them, this would prevent the avatar from translating unintentional sounds and movements that could interfere with communication. Furthermore, they mentioned the possibility of adjusting the volume of the captured audio in case the doctor spoke too softly.

Finally, one point they emphasized as very important was the presence of the Portuguese translation being displayed on the patient's screen. According to them, for deaf individuals who understand Portuguese, it is essential to verify the translation to ensure its accuracy, allowing for corrections if they notice anything unexpected. Similarly, they highlighted the usefulness of transcribing the doctor's speech in Portuguese while the avatar performs the translation into Libras, enabling them to verify what the doctor said and avoid possible translation errors that could alter the meaning of the sentence.

As additional observations, participants expressed concern about background sounds and how these could affect the doctor's sound capture, potentially impacting the translation. Additionally, they positioned the avatar's in the center of the screen, emphasizing its importance to deaf users.

5.2 Workshop 2 - Perspective of Deaf Patients

In Workshop 2, participants chose to work individually on their proposal, and each one generated one screen, as shown in Figure 4. At the end, they each presented their screen and discussed issues they considered relevant.

As in workshop 1, participants emphasized that the presence of a window displaying the Portuguese translation of the

an initial set of participants [Parker *et al*., 2019].

(a) Printed material **(b)** Additional material

Figure 2. Material provided during the workshops.

Libras signs is extremely important, as it allows deaf individuals to verify whether the signs were translated correctly and take action if they need to correct any information. Moreover, they mentioned that, for the deaf community, the verification process is crucial to ensure they feel more confident and trust that the translation – and consequently, the medical consultation – is taking place properly.

Finally, the participants suggested the inclusion of auxiliary buttons in the interface. They proposed recovery buttons in case of communication failure or breakdown, with options such as *"I didn't understand"* and *"Please repeat"*. In addition to these, they suggested buttons for objective responses like *"Yes"*, *"No"*, *"Sometimes"*, and *"More or less"*, so that the patient would not need to activate motion capture to indicate confirmation, for example.

Additionally, going beyond the scope of the prototype, the participants mentioned that it would be interesting to have a window for displaying images presented by the system itself as a support to what the doctor was saying. One of them emphasized that deaf people better learn and understand the information presented visually. Thus, if the system could display images that could support the understanding of symptoms, diseases, or specific words it would facilitate communication. Another participant suggested that the images could serve as support during communication, enabling interactions. One example given was when the doctor asks *"Where is the pain?"* , the patient could respond by click on the corresponding part of the body.

5.3 Workshop 3 - Perspective of Doctors

In Workshop 3, each participant generated a proposal, and by the end of the session, a total of two screens were produced, as shown in Figure 5. It is important to highlight that, in this workshop, the participants mentioned that they did not feel confident in drawing screens. However, it was explained to them that the activity was merely a way to materialize their ideas for discussion, and they did not need to worry about the quality of the drawing or whether it was a well-designed interface.

In terms of communication, the participants from the medical team indicated that the ideal translation method for doctors should be audio-based rather than text-based. They mentioned that if the doctor had to focus on typing, it could have a negative impact on the consultation.

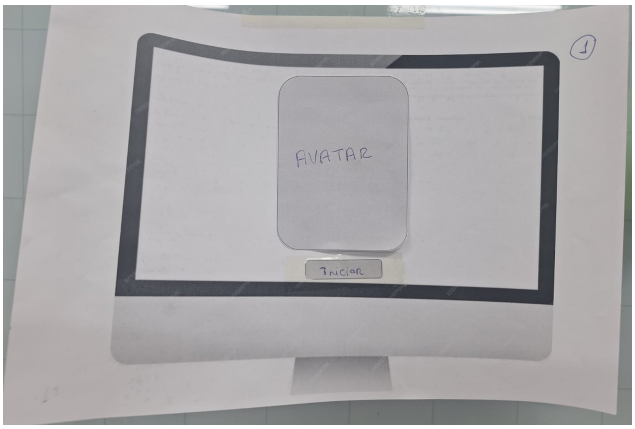
To include the audio, they considered it would be viable to have interface elements to indicate when the system should start or stop the recording for the translation (see Figure 5a).

Participants also discussed concerns about potential communication breakdowns during the consultation and how to recover from them. In this direction, they thought a chat would serve as a resource to assist in case the translation failed (and the patient understands Portuguese). Furthermore, they pointed out the need for quick ways to indicate a communication breakdown, such as saying, *"I didn't understand"* or *"Can you repeat that?"*. To address this, they proposed a set of predefined phrases that could be easily selected. Considering interaction aspects, they also suggested that it could be useful to reference a previous message in the chat to indicate which part of the conversation the failure refers to or even to provide additional clarification.

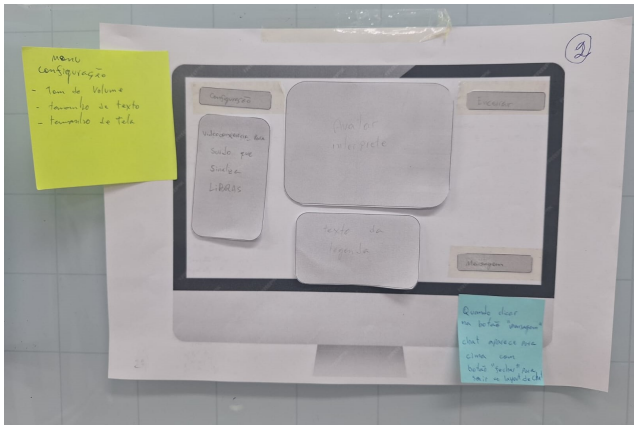
Participants also addressed issues related to the use of the system in real-world contexts and how to facilitate its use, even if some of them were beyond the scope of the prototype. For example, they indicated that ideally, the translator should be integrated into the system used in the medical context for Electronic Health Records (EHR). Furthermore, they highlighted several ethical questions related to the use of the system, such as what would happen if a translation error caused harm to the patient's health and how to identify or prevent it. And if prevention were not possible, who would be responsible? They also discussed whether it would be possible to store conversations and translations – in which case, it would be necessary to consider who would have access to them and under which circumstances.

6 Requirements

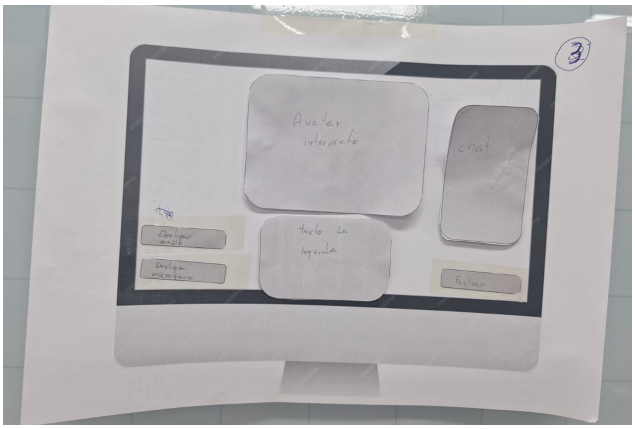
The analysis of the workshops described in the Methodology section, allowed us to identify a set of requirements for the system being proposed. In this section, we organize the resulting requirements into two subsets: those that refer to the interaction design of the system and those specific to the final system (product) and its use in a real context. For each subset of the requirements, we state the resulting requirement and add an explanation for it.



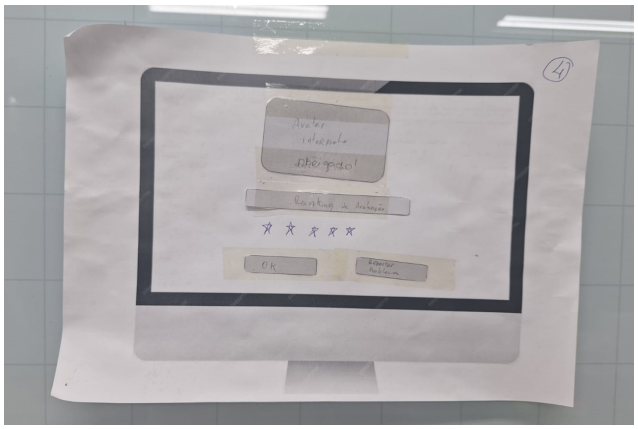
(a) Welcome screen for the user



(b) Translation screen with configuration options and buttons

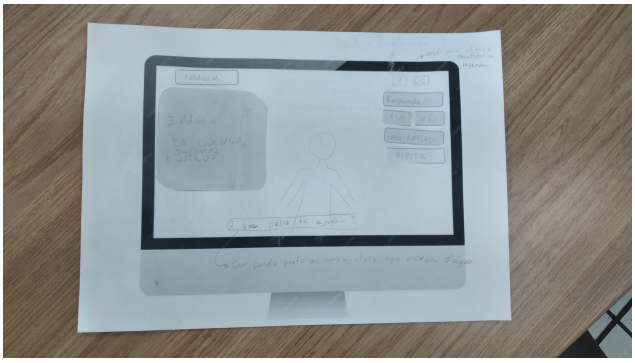


(c) Translation screen with capture control buttons

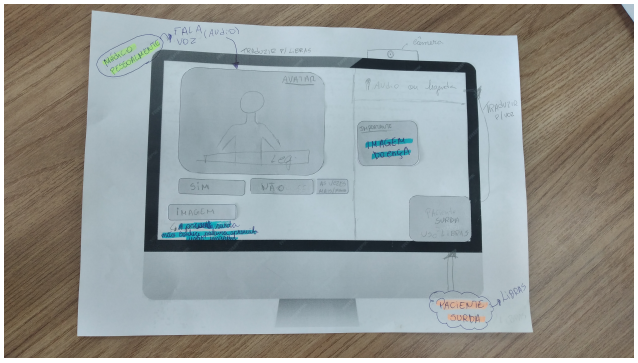


(d) Feedback screen

Figure 3. Screens produced in Workshop 1

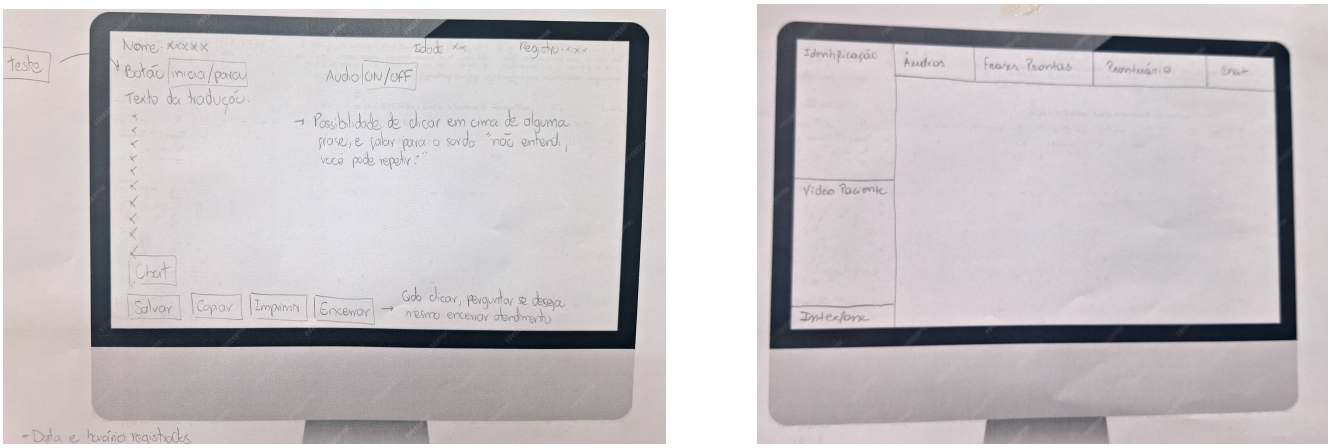


(a) Screen proposed by participant 1



(b) Screen proposed by participant 2

Figure 4. Screens produced in Workshop 2



(a) Screen proposed by participant 1 **(b)** Screen proposed by participant 2

Figure 5. Screens produced in Workshop 3

6.1 Interaction Design Requirements

A set of nine identified requirements for the interaction design of the system are presented below:

- **R1.** The translator avatar, representing the doctor, must be centralized on the system screen presented to the patient, ensuring it is clearly visible and larger than the other elements of the interface, automatically adjusting its size to accommodate the device's resolution and type (mobile or desktop).
  - Explanation: The avatar performing the translation is crucial for effective communication and constitutes the core component of the system from the patient's perspective. Therefore, the avatar's size and central positioning are essential elements for enhancing the patient experience, as they direct the patient's attention to the translation, enabling them to observe not only the avatar's hand movements but also its facial expressions. In both workshops focused on the perspective of deaf patients, participants highlighted the relevance of this requirement (see Figures 3 and 4).
- **R2.** The system must support audio to Libras translation and vice-versa.
  - Explanation: The prototype takes as input a text in Portuguese or a video in Libras and outputs, respectively, a video in Libras or a text in Portuguese. However, it is crucial for the doctor to be able to communicate using spoken language. Thus, it is important to capture audio as input and produce a voice as output, as part of the translation process. As discussed in Workshop 3, if the doctor needs to type to communicate with a patient who is face-to-face, it may negatively impact the experience and the doctor-patient rapport. Therefore, the inclusion of audio input and output is a necessary requirement for the prototype.
- **R3.** The system must provide the Portuguese translation in text format related to both patient and doctor discourse, ensuring an adequate font size.
  - Explanation: Providing a textual translation in Portuguese in both the doctor's and patient's interfaces was identified as necessary in all of the workshops. From the deaf patient's perspective, if they are bilingual in Portuguese and Libras, the text enables them to double-check both the translation of their message to the doctor and the translation in Libras the avatar provides of what the doctor says. Thus, the translation would be important to allow patients to trust the system. From the doctor's perspective, it allows them to verify whether their speech is being properly translated, which ensures their confidence in the system, while also providing access to the whole conversation during the appointment. Additionally, the translation font size must be clearly legible to ensure the double-checking will be smooth and efficient. These aspects are illustrated on the screens generated during the workshops depicted in Figures 3b, 3c, 4 and 5a.
- **R4.** Users must have control over when the system should start and end capturing their speech (video or audio), allowing interruptions as necessary.
  - Explanation: This requirement is associated with two aspects, one technical and the other from the user's perspective. From a technical perspective, there's a limitation on the length of the video or text the system can translate at a time, requiring longer inputs to be divided into segments. From the user's perspective, it is important that the system does not attempt to translate noises between communications – such as gestures from patients that do not constitute communication in Libras, or background conversation or other sounds unrelated to the doctor's communication to the patient. Therefore, the user, whether the deaf patient or the doctor, must indicate the start and end of their speech, so the system can effectively translate only the relevant communications. It is interesting to note that although this requirement came initially from a technical limitation, participants indicated that there were other reasons – translation of noise,

or even privacy (when communicating with a companion or another member of the staff) – for this requirement to be taken into account.

- **R5.** The system must provide an indication to users about the translation status, specifying the current stage - capturing, processing or presenting[11].
  - Explanation: In the context of a real-time translation system with a noticeable processing delay, it is essential to provide clear and continuous feedback to the user about the translation process. For instance, the deaf patient will rely on visual indicators to understand when their video input is being captured, processed and presented in Portuguese to the doctor, since audio feedback is unavailable to them. Similarly, the system must indicate to the doctor when the audio input is being captured, translated and presented in Libras to the patient.
- **R6.** A chat feature must be accessible as fallback option in case of potential translation failures.
  - Explanation: Although the goal is to create a translator focused on communication within the medical context, we must anticipate the possibility of translation errors. This may occur when communications include vocabulary outside the range of the system's knowledge base or that was not present in the translator's training data. In such cases, providing textual communication as an alternative is the best solution, although it is restricted to patients who are bilingual in Libras and Portuguese.
- **R7.** The chat must include a history of all messages' exchanged between patient and doctor.
  - Explanation: Once the chat is available on the interface, it is interesting that it keeps a record of the communication transcripts. The history of the messages exchanged between the patient and the doctor makes it easier to review the information discussed during the consultation, or even to reference a point in the communication more easily, if necessary. Furthemore, this resource could be used as a documentation of the consultation which could be useful for both the patient (e.g. remembering the explanation given by the doctor) and the doctor (e.g. making entries in the patient's medical file).
- **R8.** The system should anticipate and provide a quick way for patients or doctors to express some predefined phrases that represent either a way of recovering from some communication breakdown or even words or phrases that may be expressed frequently, such as *"Can you repeat that?"*, *"I didn't understand"*, *"Yes/No/Maybe"*.
  - Explanation: Providing predefined key sentences can enhance communication efficiency, as their associated translation would be immediately available. The main goal of these predefined sentences would be as fallback for communicative breakdowns or to facilitate frequent sentences expected in healthcare settings. These predefined sentences may reduce the time needed to formulate questions or responses as well as the system's translation processing time. Nonetheless, as a fallback resource, its use would be optional. Participants from Workshops 2 (perspective of deaf patients) and 3 (perspective of doctors) proposed similar suggestions, as illustrated in Figures 4 and 5.
- **R9.** The system will be new to users and must make it clear to them how to use it.
  - Explanation: Upon introducing the system in healthcare settings, it is crucial to provide clear guidance on its use, allowing both patients and doctors to effectively use it during consultations. From the perspective of the deaf patient, one key aspect is ensuring they understand how to position themselves relative to the camera so that the video can be properly captured for translation. Clear, visually oriented instructions will help patients adjust their position or even the camera to ensure the video capture is adequate, so that it can be properly translated.

6.2 Considerations for the Final System

As mentioned in section 3 (Project Captar-Libras - Overview), this projects' goal is to develop a prototype of a photorealistic Libras-Portuguese bidirectiontal translation system for the pre-medical context. The know-how generated will be transferred to the partner company to create a final system (product) that can be used in real-world contexts. Although in the workshops our focus was on the the system's interaction design some of the resulting requirements were related to the use of the system in real contexts. They are relevant to decisions regarding the final system and its introduction in the "real world". This subsection presents these requirements and issues discussed regarding each one of them.

- **C1. Ethical issues:** During the workshops, some important ethical issues were raised and require a thorough analysis before the automatic translation system can be effectively applied in real-world medical contexts. Among these questions are:
  - Storage of previous conversations: this issue applies to both the patient's and the doctor's perspective. Will the system store the doctor-patient conversations? If so, will it store only textual information or all captured data (e.g., video input from patients and audio input from doctors)?
  - Access to previous conversations: if any data is stored, who will have access to it, and under what circumstances? Will this data be treated as part of the medical examination, allowing the patient to

[11] Notice that this requirement is aligned to well known HCI usability recommendations, such as Nielsen's heuristic "Visibility of system status" [Nielsen, 1994]. Thus, although one could argue that it is not specific to this context, it was a result of the workshops. During the workshops the translation status was indicated as essential for users' interaction with the system, as described in the explanation. The same association with Nielsen's heuristics can be made regarding R9.

access and copy it? Could it be associated to the patient's record for doctor's access?
  - Accountability for translation correctness and errors: how will the system ensure translation accuracy? In the event of an error, who will be held accountable? This is especially critical if the error results in harm to the patient's health.

It is worth noting that the resulting prototype of the project focuses on the technical development of the translator and will not directly address these ethical concerns. Nonetheless, it is essential that they are considered and discussed before the system can become a product and be introduced into the context of medical care.

- **C2. Integrating the translation system with electronic health record systems:** Ideally, during a medical consultation, it should be possible to integrate the translation system with the EHR system used by the healthcare facility, allowing the professional to access all patient information in a single platform. This integration would allow the translation system to be utilized from the patient's arrival (e.g. during screening) throughout the entire consultation, including examinations. However, such integration requires compatibility not only from the translation system, but also from the EHR systems already in use by healthcare institutions. As full integration cannot always be ensured, the translation system should also consider allowing simultaneous use of both systems, including features to facilitate information exchange between applications – for instance, allowing easy copying of one or more of the patient's transcribed statements to the clipboard for pasting into the EHR system.
- **C3. Required infrastructure for using the system:** The prototype requires at least two computers[12] for its intended use: one for the patient, equipped with a camera, and one for the doctor, equipped with a microphone and audio output. Additionally, both devices must support typing messages and selecting options on the screen (e.g. via a keyboard and mouse or a touchscreen). To ensure real-time operation, access to the internet and a capable processor are required for the translation process. In Brazil, many healthcare facilities may not be able to afford the cost of this infrastructure. Therefore, it is crucial to explore more affordable alternatives that would allow the system to be adopted more widely. Moreover, it is important to consider that the computers would need to be physically positioned between the doctor and the patient, potentially impacting their interaction, the establishment of the doctor-patient rapport during the consultation, and their overall experience.
- **C4. Providing additional visual information along with the translation:** Including visual support in the form of images related to the doctor's speech to complement the translation could be an interesting and effective way to help the patients better understand the message, especially when describing symptoms, illnesses or complex medical terminology. However, while visual support can be beneficial for comprehension, how to include this visual addition would need to be carefully considered. For instance, some images associated with symptoms or diseases may be overwhelming to a patient receiving a diagnosis and might have a negative impact on the patient, at an already difficult and sensitive moment. Thus, how to curate which images to show or not, and in which situations, would need to be thorougly investigated.

[12] A single computer with two screens may also be considered, but then control of the cursor would need to be shared, which could increase the difficulty to interact.

7 Building the Prototype

The prototype presented in this section was built using the Figma[13] tool due to its support for an interactive and collaborative design experience. This tool enables the exploration of various ideas and the assessment of different scenarios before making important decisions. Moreover, it allows for simulating interactions and exploring issues related to the use of the system. The development of the prototype allowed the HCI team to consider how to meet the requirements raised and discuss different design alternatives. The choice of the final design was based on the discussion of the team of the costs and benefits for the user of the different possibilities considered. The prototype was also used to support the discussion of relevant issues of the interaction and the solutions being proposed with the whole project team.

We next present the prototype (Figures 6, 7, 8 and 9) and an overview explanation on how it addressed the requirements presented in section 6.

The patient's main interface, shown in Figure 6, is divided into three vertical sections:

- The left section displays the patient's camera feedback (R9) along with the button used to capture the user's video of the message in Libras for translation (R4). Just below, quick-phrase buttons (R8) describe common symptoms ("pain", "fever" and "shortness of breath") and an intensity scale ("weak", "moderate" and "strong"). Each button is labeled with a caption in Portuguese and contains a video in Libras that plays on hover. The user can see behind the button label a shadowed image of the video. When the cursor hovers over the button the video presenting its meaning in Libras is shown, allowing patients with limited proficiency in Portuguese to understand the meaning of the button. At the bottom, the Portuguese transcription of the patient's last recorded message in Libras (R4) is displayed.
- The central section contains the photorealistic avatar representing the doctor's interpreter (R1) and a text box displaying the Portuguese transcription of the doctor's last utterance (R3). In the right corner of this text box, a quick phrase button (R8) labeled *"I didn´t understand"* is available for easy access.
- The right section presents the chat box displaying the conversation history between the patient and the doctor (R7), including a text box for the patient to type a message in cases of translation system fails (R6), and on the

[13] See: `https://www.figma.com`

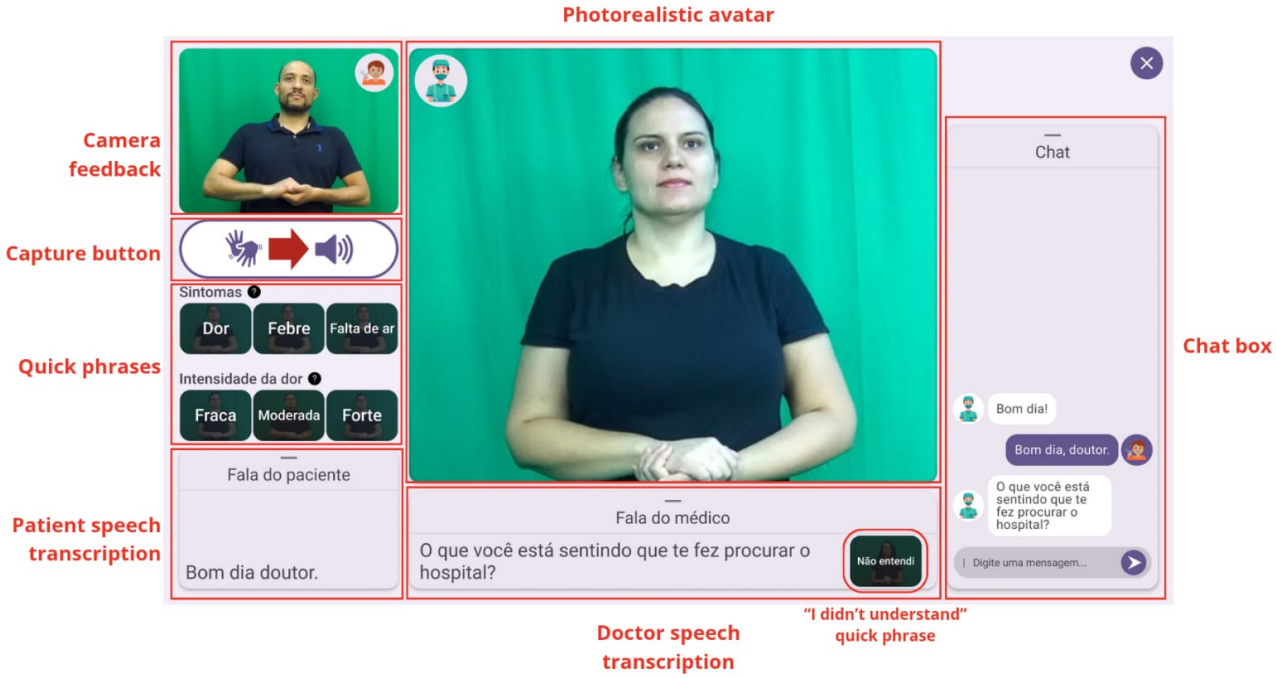


Figure 6. Patient's main interface

top right hand corner, the button to close the system and end the consultation. The chat box, as well as the text boxes displaying the doctor's and patient's speech, can be hidden by clicking the trace-shaped icon at the top of these elements.

The interaction depicted in the sequence of images in Figure 7 begins when the user presses the record button. The patient's camera feed moves to the center of the screen and a countdown is provided for the patient to prepare for recording the message. Once the countdown ends, the recording begins and a stop button is displayed, allowing the patient to stop recording at any time (R4). When the patient finishes recording, the system displays the message *"Traduzindo..."* (*"Translating..."* in English) in the center of the screen, indicating that the video is being translated into Portuguese (R5). In the interactive prototype, the ellipsis in this message is animated, conveying the idea that a process is ongoing, helping patients who are not proficient in Portuguese to understand the message. When the system finishes processing the translation, the new sentence appears in the patient's caption box (R3) and in the chat, confirming that the message was successfully transmitted to the doctor.

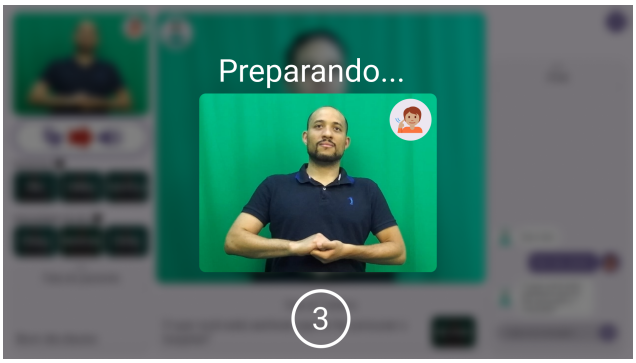
The doctor's interface, shown in Figure 8, differs from the patient's interface in several aspects. We expect that doctors will primarily use audio as their main channel to convey their utterances (R2). Therefore, the main interaction available to doctors is audio recording, activated through the record button located on the right, as shown in Figure 8. When the doctor presses the record button, a new highlighted element appears on the screen, visually indicating that the doctor's message is being recorded and allowing them to stop recording at any moment by pressing the stop button (R4) (Figure 9). On the left side of the screen, the doctor can view the avatar representing their interpreter and the patient's video feed. The interpreter's avatar allows the doctor to follow the translation process, knowing when the message is being transmitted and when it has concluded (R5). The patient's video feed helps the doctor to guide the patient in positioning themselves in front of the camera or ensuring proper video capture for translation.

Similar to the patient's interface, the doctor's interface includes a chat box displaying the conversation history and a text box for typing and sending messages. Unlike the patient's interface, the chat is the central element of the doctor's interface. The doctor's quick phrases section does not include non-verbal elements, as in the patient's interface, since the text alone fulfills its purpose (R8).

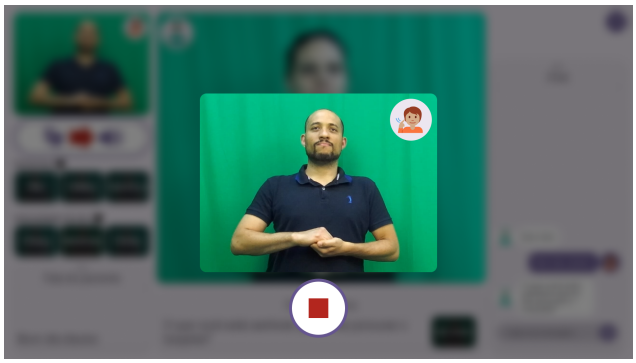
8 User Evaluation

The HCI team conducted a user evaluation of the interface prototype with ten participants. The goals of the evaluation were to assess the proposed interface with potential users and collect indicators about the elicited requirements. Thus, participants were invited to interact with the interface prototype, which simulated its use and, at the end, discuss their perceptions of the system and their experience with it.

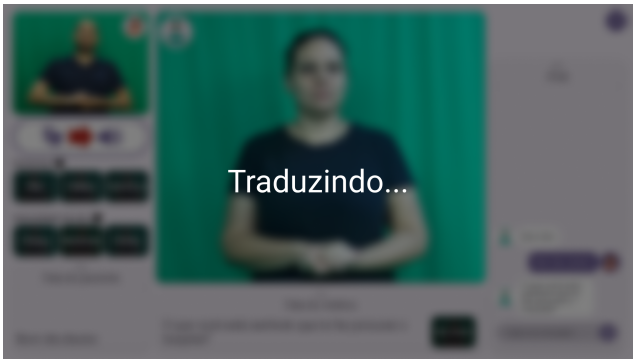
The user evaluation took place during October and November 2024, involving ten voluntary participants: eight bilingual deaf participants (four participants were male and four were female) and two medical students (both were female). Participants were recruited through the team members' network or snowball technique, as presented in section 4.1. Additionally, pilot tests were conducted beforehand with two other users to validate the evaluation procedure and material.



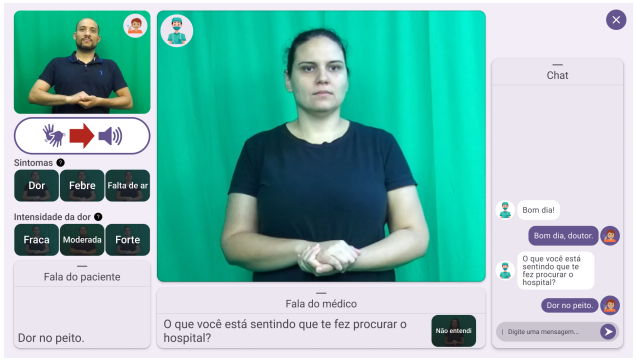
(a) Patient preparing to record



(b) Recording the message in Libras



(c) System processing the video and translating to Portuguese



(d) Main screen updated with the patient's new message

Figure 7. Interaction for patient recording a message.

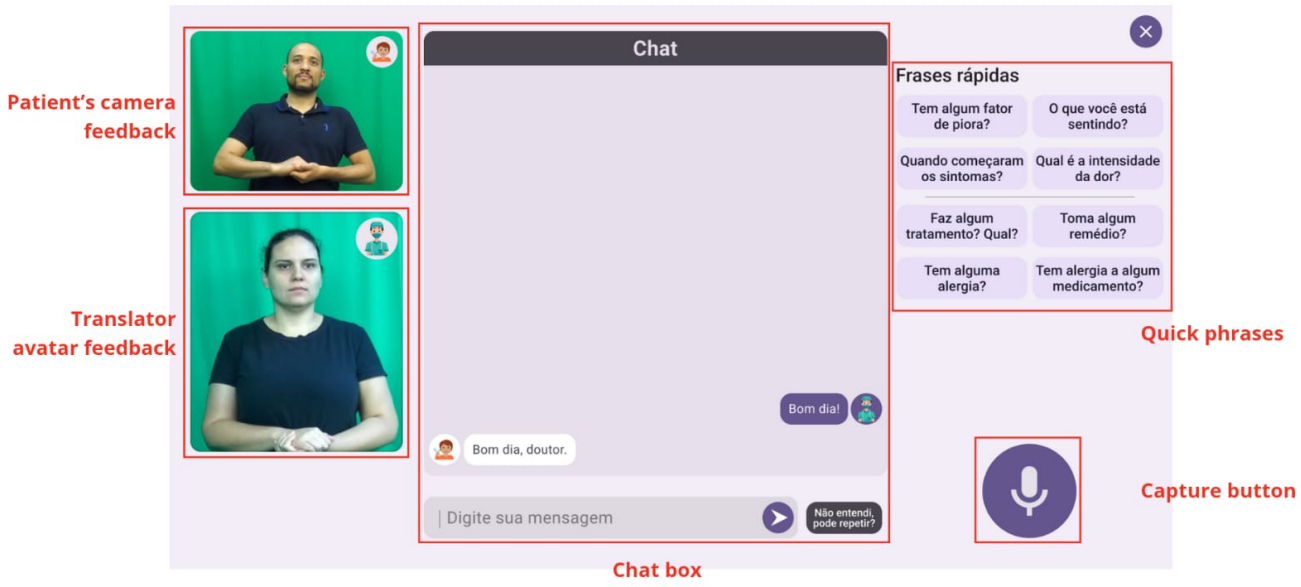


Figure 8. Doctor´s main interface

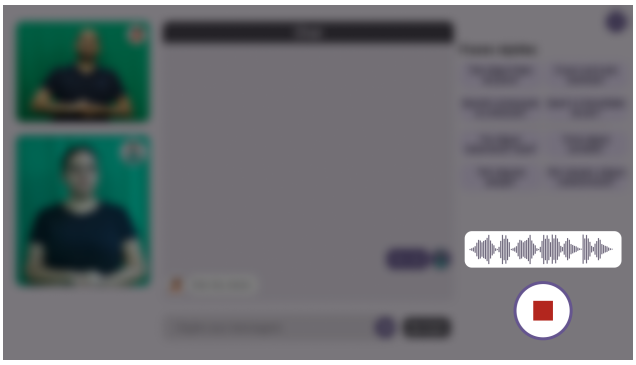


Figure 9. Interaction for the doctor recording a message.

8.1 Simulated Nature of the Evaluation

The interface prototype was a horizontal prototype that simulated the functioning of the system. Thus, participants were not entirely free to interact, in the sense that the messages being exchanged and the interface elements available at each moment were previously determined. For instance, messages being exchanged were not input by participants. They clicked on the video record button (deaf participant) or audio record (medical participant) and the system showed a predefined video of a Libras recording, or the feedback for audio recording, respectively, and its translation. If the participant clicked on the chat's text box to type a message, a predefined text was automatically displayed and could then be sent.

This methodological approach allowed the HCI team to observe users' interaction with the system and collect their perspectives on the proposed system and its interface, even before the translation system is completely functional. Moreover, the simulated nature of the tests ensured that all participants underwent the same communication exchanges through (at least) very similar interactive paths. This allowed the HCI team to conduct more specific comparisons between different users and facilitated the identification of behavioral patterns and recurring difficulties.

8.2 Conducting the Evaluation

The evaluation was carried out in a controlled environment, containing a table, a laptop, and a monitor — the laptop was used by the patient, while the monitor was used by the HCI team member to simulate the other user in the consultation. One participant was present per session and a member of the HCI team simulated the role of the other user (doctor or patient). Another researcher from the HCI team was present in the evaluation room and was responsible for conducting the evaluation and observing it from afar. Note that for the evaluation sessions of the patient role, in which the participants were deaf, an interpreter was also present to ensure clear communication between the participant and the research team.

At the beginning of each session, the researchers explained to the participants the goals of the project and the system. They also explained the goals of the evaluation being conducted, its simulated nature, and how it would be carried out. The consent form was presented to the participants and they expressed their agreement with its terms before the evaluation started.

An introductory tutorial was presented, outlining the structure and main functionalities of the system. Participants were encouraged to explore the interface and different elements to convey their message, including video or audio capture for translation, text-based chat, and quick-phrase buttons. They were also told that they could express comments or observations regarding the interface at any time during the session (and not only at the end).

At the end of each session, a semi-structured interview was conducted. The interview covered participants' first impressions of the system and the tutorial presented, ease of use, perception of the interface and its elements, overall satisfaction, potential social impact, if they would like to use such a system, and any other comments they might have of their experience or suggestions about the system. The participants' interaction with the system was recorded (through a screen recorder program), and the researcher conducting the evaluation took field notes during the session. Furthermore, the user session was recorded in video.

8.3 Main Results

Based on the analysis of the material collected during the evaluation, the research team identified several challenges faced by participants and adjustments needed in the system's interface. Next we describe the main issues identified:

- **Confusion with video capture button**

  The graphical elements used to depict the capture (and translate) button caused confusion among participants. The button depicted the symbol for Libras followed by an arrow pointing to the symbol for audio (see 'Capture Button' indicated in Figure 6). Once the button was pressed, it would start the recording of the video of the patient signing, that would then be translated into Portuguese and presented both in the patient's and doctor's interfaces. However, some participants interpreted the button function as being to start a call with a human interpreter. Furthermore, during the evaluation, when they clicked on the button, a video of someone signing started. This video was supposed to indicate the simulation of the recording of the users themselves. However, many participants did not initially understand that it was a simulation of their own signing. As a result, some participants were unable to correctly interpret the button, which caused frustration and made it difficult for them to operate the system. As a result, some participants chose not to use the button – and, consequently, the main functionality of the system – and instead preferred the chat option as an alternative for communication. This issue indicated the need to redesign the button for the final system.

- **Quick-phrase video dimensions**

  Most participants reported that the videos displaying the meaning of the quick-phrases buttons in Libras were too small to clearly understand the signed message. This issue impacted the intended use of this feature, as many

participants found it difficult to use, especially in a medical context, where the user must fully comprehend the conveyed message. To improve visualization, the videos should be enlarged when shown to users.

- **Inconsistency in quick phrases**

  The symptom descriptions and intensity scale provided by the quick-phrase buttons were not well-structured (see 'Quick phrases' indicated in Figure 6) . There were 3 symptoms being shown – *"pain", "fever" and "shortness of breath´´* – in the first row of quick-phrase buttons. The set of quick-phrase buttons below them referred only to the intensity of the pain, and had no relation to the other symptoms. This was not clear to participants, and some of them tried to make sense of how they related to all symptoms. Despite their initial misunderstanding of the messages, once clarified, participants highlighted the importance of the quick-phrase buttons for efficient communication. Thus, the phrases must be reformulated to mitigate misunderstandings and better align with the clinical context.

Notice that the identified interface problems were all related to the patient's interface. Participants did not observe or report any problems considering the doctor's interface.

The issues identified regarding the patient's interface provide valuable insights to improve the interface and its usability, and will be considered for the future implementation of the final system. The difficulties faced by the participants in the role of patients were mainly associated with the presentation of the elements and not their intended use in the interface. Thus, none of the problems indicates the need to review the requirements elicited and used to generate the interface prototype.

Finally, it is worth mentioning that during the interviews all participants said that they would be interested in using the system. Half of them pointed out that the system would be very useful as it would give deaf patients autonomy to go alone to a doctor´s appointment.

9 Discussion and Limitations

The workshops conducted in this study played a crucial role in identifying interface and interaction requirements for bidirectional Libras-Portuguese translation systems using photorealistic avatars in healthcare contexts. Although the workshops did not include end-users, the project's team members represented the intended user groups. By considering the user perspective and focusing specifically on interaction, relevant insights were obtained for developing these systems. It is important to note that although many studies have explored translation between sign and oral languages, most concentrate primarily on translation algorithms and technical aspects of image recognition [Areeb and Nadeem, 2021; da Silva *et al*., 2024]. The few studies considering interface or user perspectives present other gaps: they either incorporate these perspectives only during design and evaluation phases without specifying development requirements, or they provide only general features based on user preferences [David *et al*., 2024; Tran *et al*., 2023].

This study addresses a gap in current literature by developing specific interaction requirements for a real-time, bidirectional photorealistic translation system, and can potentially be useful to the development of similar systems for other contexts. Although the requirements in this study were developed for the project Captar-Libras, they do not address specific aspects of the healthcare context. For instance, features such as audio to Libras translation and vice versa (R2), translation status indication (R5), and chat history (R7) are relevant in multiple contexts requiring inclusive and accessible communication, such as educational platforms and public services. Thus, it would be interesting to investigate the applicability of the proposed set of requirements in other translation contexts.

The interface prototype built is a proof of concept of the feasibility of designing interaction based on the identified requirements. It illustrates concretely how the requirements can be met, facilitating the development team's progress.

This concrete presentation is an essential step for ensuring the proposed solutions are effective in practice, benefiting not only the current project, but also future studies and development of similar systems that could potentially be influenced by this study.

It is important to note that the user interface shown in Figures 6 and 7 does not rely on written text for comprehension. Although Portuguese labels are shown to facilitate the interaction for bilingual deaf users' understanding, they are also available to users in Libras. Constructing and maintaining the interface to ensure comprehension independent of textual elements is a constant challenge throughout prototype development.

After building the interface prototype, it was evaluated with end-users. The user evaluation revealed both positive aspects and challenges in interacting with the system. Overall participants were able to understand the visual organization of the elements and to interact with it, simulating the patient-doctor conversation. However, they had difficulty interpreting some of the elements, indicating the need for adjustments to ensure that the phrases convey clear and understandable meanings. Additionally, the use of text chat as an alternative communication method was well-received, providing greater flexibility. In short, as presented in the previous section, the main problems identified are related to the presentation of some elements, and not their role in the intended interaction.

Although the study presented in this paper resulted in relevant contributions, it is also important to recognize its limitations. First, conducting workshops with team members instead of actual end-users may introduce bias. However, this limitation is mitigated since team members also represent the profile of the intended end-users – deaf patients and healthcare professionals, and understand the realities, needs, and expectations of end-users. Moreover, the goal was not to create requirements for a hypothetical solution but rather for a predefined one, considering its inherent technical constraints that influence the interaction. Consequently, it was not possible to explore discussions and reflections on other possible solutions for communication barriers between doctors and deaf patients.

A noteworthy challenge faced during the user evaluation

was the difficulty of recruiting deaf volunteer participants. The use of team member's network contacts and snowballing strategies to recruit participants may have impacted the diversity of the people who volunteered to participate. Furthermore, the only recruitment requirement was being a representative of the intended users. Other aspects such as their fluency in Portuguese, experience with technology, background and age were not considered to recruit participants, and were not collected in the evaluation. However, these are important aspects and we intend to take them into account in the future evaluations of the system.

Another limitation was the use of a horizontal prototype, simulating the interaction. Due to limitations in building the prototype, it was not possible to simulate the capture of the participants' video. Thus, the simulation of the video capture of the participants' signing was depicted through a video of a (deaf) team member signing the message. However, around half of our participants did not initially understand this simulation, and often led them to not use the main functionality of the system, but the alternative communicative means (i.e. chat and quick-phrase buttons). In these cases, usually during the interview, when participants explained their interpretation, the researchers clarified the intended meaning of the button and simulated video.

We argue that the misunderstanding caused by the simulation does not change the usefulness of the insights that resulted from the user evaluation. However, the simulation nature of the evaluation may have prevented the identification of other issues that may impact the interaction. At any rate, the evaluation was intended as a formative evaluation, and not as the only user evaluation to be conducted throughout the development process.

Regarding the requirements, the results of the evaluation provided an initial positive indicator of their usefulness, but not enough to indicate their consolidation. Future evaluations of the functional prototype will allow us to collect more indicators towards the consolidation or need to review the requirements within the context of the Captar-Libras project. Furthermore, some technical challenges within the project imposed interaction constraints that were reflected in the requirements (e.g. R8, the need for elements to represent quick-phrases), and thus, such requirements may evolve alongside technological advancements.

10 Conclusion and Next Steps

The Captar-Libras project addresses a highly relevant issue of accessibility and social inclusion in society: healthcare access for deaf people. The project's scope covers research and development of the technical solution and involves transferring the technology to the partner company, who will then make it into a product. The proposed solution involves generating a photorealistic avatar capable of bidirectional translation between Libras and Portuguese, specifically in healthcare contexts. From an HCI perspective, the project presents challenges [Prates *et al*., 2025], including how technical limitations directly impact the intended interface and interaction. Therefore, it is essential to consider how to design an interaction that enables using and evaluating the system being developed while also considering issues related to the interaction needs of the target audience, especially deaf users.

The results presented in this article provide four main contributions to the HCI community, particularly researchers and developers working in the area: the methodology used; the set of requirements and considerations; the horizontal interaction prototype; and the evaluation with end-users. Regarding methodology, the conducted workshops allowed for consideration of both technical constraints and the needs of deaf patients and healthcare professionals. Our report of how the workshops were organized and conducted can be useful for applying similar methodologies in other contexts beyond automatic translation systems.

The set of requirements and considerations is the article's main contribution. Although generated within the context of the project, the descriptions and justifications for each requirement support the development of automatic Libras-Portuguese translation systems applied to other contexts beyond the healthcare context. Moreover, considerations presented regarding the development of the final system as a product are relevant both to researchers and developers working on translation between sign and oral languages. The horizontal interface prototype, developed based on these requirements, materializes a proposal designed to fulfill them. Consequently, the prototype enables a clearer understanding of a possible solution, and fostering deeper discussions and reflections about the interaction design.

The user evaluation conducted contributed both to advance the research presented, as well as to provide insights about the methodology used. The results of the user evaluation indicated necessary improvements in the interface, but did not point to the need to review the set of requirements, contributing to its consolidation, albeit still very preliminary. The interaction simulation methodology adopted was able to generate useful results and identify interface improvements. However, in our specific context, the use of another's person video to simulate the participants' own video proved to be challenging. In future work, using interaction simulation, it would be important to make clear to participants beforehand any limitations of the simulation that may impact their understanding of the (part of) the prototype.

As a next step in our research, the interface will be reviewed to address the problems identified in our user evaluation. This new interface will be implemented in the system's functional prototype being developed. Once the functional prototype is robust, a new set of user evaluations will be conducted. The plan is to conduct an evaluation of the prototype with potential users representing the deaf patients and doctors. The first evaluation being planned will simulate the medical consultation, through the use of scenarios, and have one participant per session, the other role being taken by a research team member that represents the intended user (deaf patient or doctor). Depending on the results of the individual participant evaluation, we may consider as next step the same evaluation, but with two participants, each one representing a different user profile.

Finally, in the last stage of our research the goal is to evaluate the system in a real context. However, there are a number of ethical issues to be considered before such an evaluation can take place. These new evaluation rounds will be useful

not only to assess the quality and identify how to improve the final system prototype but also to provide more indicators towards the consolidation of our set of requirements.

The research aimed at developing a more accessible translation system not only enhances the medical consultation experience for deaf patients and doctors, but also benefits society as a whole. This initiative strengthens social bonds and contributes to building a fairer and more supportive society by facilitating communication among all its members. The contributions provided by this study have the potential to inspire and guide the development of inclusive technological solutions by system developers, researchers, and professionals. These solutions may then play a crucial role in promoting the inclusion and accessibility for deaf individuals in healthcare services, leveraging technology to address their specific communication needs.

Declarations

Authors' Contributions

All authors contributed to the planning and definition of the research, its methodology and ethical considerations. ROP, JMGS, ELAB and RAAC contributed to the organization and execution of the workshops. All authors contributed to the decisions and review of the prototype. NSS, LASS, RAAC, ELAB and ROP contributed to the planning and execution of the user evaluation. NSS, LASS, JMGS and ROP were responsible for the analysis of the data collected. NSS, LASS, JMGS and ROP wrote this manuscript. MFMC, ELAB and RAAC reviewed and edited the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

We thank the workshop and user evaluation participants who kindly agreed to contribute to this research.

Funding

We thank FINEP (nº 2797/20) and SignumWeb Comunicação Inclusiva Ltda, for their partial funding of the Project "Captar-Libras: Video Communication System for the Deaf Applied to Pre-Medical Care" (In Portuguese:"Captar-Libras: Sistema de Comunicação por vídeos para surdos aplicado ao pré-atendimento médico")

Availability of data and materials

The datasets generated and/or analysed during the current study will be made upon request.

References

Areeb, Q. M. and Nadeem, M. (2021). Deep learning based hand gesture recognition for emergency situation: A study on indian sign language. In *2021 International Conference on Data Analytics for Business and Industry (ICDABI)*, pages 33–36. IEEE. DOI: 10.1109/ICDABI53623.2021.9655842.

Barnet, H. (2021). Snapshot insight report: Deaf people's healthcare challenges. Available at:`https://nds.healthwatch.co.uk/reports-library/snapshot-insight-report-deaf-peoples-healthcare-challenges`. Last access: November, 2025.

Beaver, S. and Carty, B. (2021). Viewing the healthcare system through a deaf lens. *Public Health Research and Practice*, 31(5):1–4. DOI: 10.17061/phrp3152127.

Bond, L. (2019). Surdos enfrentam dificuldade para atendimento em saúde. Available at:`https://agenciabrasil.ebc.com.br/direitos-humanos/noticia/2019-10/surdos-enfrentam-dificuldade-para-atendimento-em-saude`. Last access: November, 2025.

da Silva, D. R., de Araújo, T. M. U., do Rêgo, T. G., Brandão, M. A. C., and Gonçalves, L. M. G. (2024). A multiple stream architecture for the recognition of signs in brazilian sign language in the context of health. *Multimedia Tools and Applications*, 83(7):19767–19785. DOI: 10.1007/s11042-023-16332-7.

da Silva, D. R. B., Araujo, T. M. U., Rêgo, T. G. d., and Brandão, M. A. C. (2020). A two-stream model based on 3d convolutional neural networks for the recognition of brazilian sign language in the health context. In *Proceedings of the Brazilian Symposium on Multimedia and the Web*, WebMedia '20, page 5–12, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3428658.3430980.

Das, H. V., Mohan, K., Paul, L., Kumaresan, S., and Nair, C. S. (2024). Transforming consulting atmosphere with indian sign language translation. *Multimedia Tools and Applications*, 83(5):13543–13555. DOI: 10.1007/s11042-023-15214-2.

David, D., Alamoodi, A., Albahri, O., Garfan, S., Albahri, A., Zaidan, B., and Chen, J. (2024). Sign language mobile apps: a systematic review of current app evaluation progress and solution framework. *Evolving Systems*, 15(2):669–686. DOI: 10.1007/s12530-023-09494-0.

de Menezes, J. E. S. A. and Feitosa, C. R. d. S. (2015). *Língua Brasileira de Sinais (LIBRAS)*. Editora da Universidade Estadual do Ceará (EdUECE). Available at:`https://educapes.capes.gov.br/bitstream/capes/176804/2/Livro_Linguagem%20Brasileira%20de%20Sinais_Libras.PDF`.

De Meulder, M. (2021). Is "good enough" good enough? ethical and responsible development of sign language technologies. In Shterionov, D., editor, *Proceedings of the 1st International Workshop on Automatic Translation for Signed and Spoken Languages (AT4SSL)*, pages 12–22, Virtual. Association for Machine Translation in the Americas. Available at:`https://aclanthology.org/2021.mtsummit-at4ssl.2`.

de Souza, V. D., Hoeckele, A. G., Borim, M. L. C., Christinelli, H. C. B., and Costa, M. A. R. (2020). Percepção de surdos sobre o atendimento nos serviços de saúde.

Brazilian Journal of Development, 6(8):55347–55356. DOI: 10.34117/bjdv6n8-091.

Dere, M., Dere, R. O., Adesina, A., and Yauri, A. R. (2022). Smartcall: A real-time, sign language medical emergency communicator. In *2022 5th Information Technology for Education and Development (ITED)*, pages 1–6. DOI: 10.1109/ITED56637.2022.10051420.

Escudeiro, P., Escudeiro, N., Reis, R., Lopes, J., Norberto, M., Baltasar, A. B., Barbosa, M., and Bidarra, J. (2015). Virtual sign–a real time bidirectional translator of portuguese sign language. *Procedia Computer Science*, 67:252–262. DOI: 10.1016/j.procs.2015.09.269.

Farooq, U., Rahim, M. S. M., Sabir, N., Hussain, A., and Abid, A. (2021). Advances in machine translation for sign language: approaches, limitations, and challenges. *Neural Computing and Applications*, 33(21):14357–14399. DOI: 10.1007/s00521-021-06079-3.

Francisco, G. d. S. A. M. and Mesquita, C. T. (2023). Access to health services among deaf people: An issue of inclusion and linguistic rights. *International Journal of Cardiovascular Sciences*, 36. DOI: 10.36660/ijcs.20220196.

Guo, Y., Zhao, J., Ding, B., Tan, C., Ling, W., Tan, Z., Miyaki, J., Du, H., and Lu, S. (2023). *Sign-to-911: Emergency Call Service for Sign Language Users with Assistive AR Glasses*. Association for Computing Machinery, New York, NY, USA. DOI: 10.1145/3570361.3613260.

Hisham, B. and Hamouda, A. (2019). Supervised learning classifiers for arabic gestures recognition using kinect v2. *SN Applied Sciences*, 1(7):768. DOI: 10.1007/s42452-019-0771-2.

Izquierdo-Condoy, J. S., Sánchez Abadiano, L. E., Sánchez, W., Rodríguez, I., De La Cruz Matías, K., Paz, C., and Ortiz-Prado, E. (2024). Exploring healthcare barriers and satisfaction levels among deaf individuals in ecuador: A video-based survey approach. *Disability and Health Journal*, 17(3):101622. DOI: 10.1016/j.dhjo.2024.101622.

Kipp, M., Nguyen, Q., Heloir, A., and Matthes, S. (2011). Assessing the deaf user perspective on sign language avatars. In *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility*, pages 107–114. DOI: 10.1145/2049536.2049557.

Ministério da Saúde (2019). Pesquisa nacional em saúde. Available at:`https://www.pns.icict.fiocruz.br/wp-content/uploads/2021/12/liv101846.pdf`. Last access: November, 2025.

Naranjo-Zeledón, L., Peral, J., Ferrández, A., and Chacón-Rivas, M. (2019). A systematic mapping of translation-enabling technologies for sign languages. *Electronics*, 8(9):1047. DOI: 10.3390/electronics8091047.

Nielsen, J. (1994). Heuristic evaluation. In Nielsen, J. and Mack, R. L., editors, *Usability Inspection Methods*, pages 25–62. John Wiley & Sons, Inc., New York, NY, USA. Book.

Oliveira, T., Escudeiro, P., Escudeiro, N., Rocha, E., and Barbosa, F. M. (2019). Automatic sign language translation to improve communication. In *2019 IEEE Global Engineering Education Conference (EDUCON)*, pages 937–942. IEEE. DOI: 10.1109/educon.2019.8725244.

Parker, C., Scott, S., and Geddes, A. (2019). Snowball sampling. *SAGE research methods foundations*. DOI: 10.4135/9781526421036831710.

Prates, R. O., Santos, N. S., de Souza, L. A., Marcolino, M. S., Bernardino, E. L., and Campos, M. F. (2025). Hci challenges in the development of captar-libras: Providing accessibility for deaf patients in health context. In *IFIP Conference on Human-Computer Interaction*, pages 313–322. Springer. DOI: 10.1007/978-3-032-05005-2_16.

Rezende, R. F., Guerra, L. B., and Carvalho, S. A. d. S. (2021). The perspective of deaf patients on health care. *Revista CEFAC*, 23:e0620. DOI: 10.1590/1982-0216/20212320620.

Rocha, J., Lensk, J., Ferreira, T., and Ferreira, M. (2020). Towards a tool to translate brazilian sign language (libras) to brazilian portuguese and improve communication with deaf. In *2020 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, pages 1–4. DOI: 10.1109/VL/HCC50065.2020.9127257.

Rogers, K. D., Lovell, K., Bower, P., Armitage, C. J., and Young, A. (2025). What are deaf sign language users' experiences as patients in healthcare services? a scoping review. *PLOS Global Public Health*, 5(2):e0003535. DOI: 10.37766/inplasy2022.1.0102.

Santos, N. S., Souza, L. A. S. d., Soares, J. M. G., and Prates, R. O. (2024). Connecting silent worlds: Requirements for automatic oral-sign language translation. In *Proceedings of the XXIII Brazilian Symposium on Human Factors in Computing Systems*, pages 1–14. DOI: 10.1145/3702038.3702066.

Soares, J. M. G., Carvalho, I. F. d., Bernardino, E. L. A., Marcolino, M. S., and Prates, R. O. (2024). An evaluation of portuguese to libras translator apps applied to the medical context. In Antona, M. and Stephanidis, C., editors, *Universal Access in Human-Computer Interaction*, pages 290–304, Cham. Springer Nature Switzerland. DOI: 10.1007/978-3-031-60884-1_20.

Sosa-Jiménez, C. O., Ríos-Figueroa, H. V., and Solís-González-Cosío, A. L. (2022). A prototype for mexican sign language recognition and synthesis in support of a primary care physician. *IEEE Access*, 10:127620–127635. DOI: 10.1109/access.2022.3226696.

Steinberg, A. G., barnett, S., Meador, H. E., Wiggins, E. A., and Zazove, P. (2006). Health care system accessibility: experiences and perceptions of deaf people. *Journal of general internal medicine*, 21:260–266. DOI: 10.1111/j.1525-1497.2006.00340.x.

Süzgün, M., Özdemir, H., Camgöz, N., Kındıroğlu, A., Başaran, D., Togay, C., and Akarun, L. (2015). Hospisign: an interactive sign language platform for hearing impaired. *Journal of Naval Sciences and Engineering*, 11(3):75–92. Available at:`https://izlik.org/JA46PX55YY`.

Tran, N., Ladner, R. E., and Bragg, D. (2023). U.s. deaf community perspectives on automatic sign language translation. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '23. Association for Computing Machinery. DOI: 10.1145/3597638.3614507.

Wheatley, M. (2021). Barriers faced by deaf people while accessing healthcare. Available at:`https://www.edf-feph.org/blog/barriers-faced-by-`

deaf-people-while-accessing-healthcare/. Last access: November, 2025.

WHO (2021). World report on hearing. Available at:https://www.who.int/publications/i/item/9789240020481. Licence: CC BY-NC-SA 3.0 IGO. Last access: November, 2025.

Xia, K., Lu, W., Fan, H., and Zhao, Q. (2022). A sign language recognition system applied to deaf-mute medical consultation. *Sensors*, 22(23):9107. DOI: 10.3390/s22239107.