electronics

MDPI

Article

Sign Language Representation by TEO Humanoid Robot: End-User Interest, Comprehension and Satisfaction

Jennifer J. Gago [*,†,‡], **Juan G. Victores** [†,‡] **and Carlos Balaguer** [†]

Robotics Lab, University Carlos III de Madrid, 28911 Madrid, Spain; jcgvicto@ing.uc3m.es (J.G.V.); balaguer@ing.uc3m.es (C.B.)

* Correspondence: jenniferjoana.gago@uc3m.es or jennifer.gmunoz@gmail.com; Tel.: +34-91-624-6241

† Current address: Av. de la Universidad, 30, 28911 Leganés, Madrid, Spain.

‡ These authors contributed equally to this work.

Received: 27 November 2018; Accepted: 2 January 2019; Published: 4 January 2019

Abstract: In this paper, we illustrate our work on improving the accessibility of Cyber–Physical Systems (CPS), presenting a study on human–robot interaction where the end-users are either deaf or hearing-impaired people. Current trends in robotic designs include devices with robotic arms and hands capable of performing manipulation and grasping tasks. This paper focuses on how these devices can be used for a different purpose, which is that of enabling robotic communication via sign language. For the study, several tests and questionnaires are run to check and measure how end-users feel about interpreting sign language represented by a humanoid robotic assistant as opposed to subtitles on a screen. Stemming from this dichotomy, dactylology, basic vocabulary representation and end-user satisfaction are the main topics covered by a delivered form, in which additional commentaries are valued and taken into consideration for further decision taking regarding robot-human interaction. The experiments were performed using TEO, a household companion humanoid robot developed at the University Carlos III de Madrid (UC3M), via representations in Spanish Sign Language (LSE), and a total of 16 deaf and hearing-impaired participants.

Keywords: accessibility; anthropomorphic robotic hands; assistive robotics; Cyber–Physical Systems; dactylology; household companion; humanoid; human–robot interaction; robotics; sign language; statistics; survey; vocabulary

1. Introduction

User accessibility and Universal Design (UD, also known as Design For All), are currently getting a growing consideration worldwide to reduce the physical and attitudinal barriers among people of all ages and abilities [1]. Regarding deaf and hearing-impaired people accessibility, Spanish and Catalan Sign Languages were recognised to be official languages in Spain by the national Parliament (BOE 27/2007) in 2007 [2].

Several measures regarding the learning of this language from an early age and empowering deaf people to request interpreters in public and private services and areas have been taken. Following this approach, the use of resources that enhance and enable oral communication, such as lipreading, hearing aids, subtitling and other technological advances, has been declared a fundamental right. These measures aim to overcome any kind of discrimination of people with hearing disabilities in their access to information and communication, keeping in mind their heterogeneity and the specific needs of each group.

Regarding UD, there is a need to focus on the development of products that are easily accessible to as many people as possible, without the need to adapt or redesign them in a special way. In order to meet these objectives in the field of Cyber–Physical Systems (CPS), human–robot interaction must be not only accessible, but also usable. This guarantees easy access attributes and the possibility of understanding and learning how to communicate with the robot in a natural and intuitive way, without the need to investigate or get additional assistance.

Finding a way to make sure deaf or hearing-impaired individuals feel comfortable about interacting with technology is a step forward towards achieving the accessibility goal. The most widely used resource is to display subtitles on a screen, since sign language interpretation is not always an available option and it represents numerous challenges regarding its correct use and implementation. For that reason, there are many open questions whether or not it is likely that sign language users feel comfortable interacting with a robot in their everyday language.

1.1. Challenges of Representing Sign Language

Representing sign language is a complex task which needs from advanced software and hardware to be done properly. It is not only a matter of precision, speed and movement fluidity, it is important to consider that signing is commonly complemented with facial expressions, shoulder raising, mouth morphemes, head tilt/nod/shake among other non-verbal communication signals that affect the meaning of the message, those are part of a set of behaviours called "non-manual markers" [3].

The complexity of sign language is the main reason why it is still a quite incipient developing area in human–robot interaction, in comparison to other topics. There are relatively few projects related to robot reproduction of sign language. The assistant android developed in 2014 by Toshiba Corporation in collaboration with other Japanese technological institutes can mimic some simple movements, such as greetings and signing in Japanese [4]. In addition, humanoids Robovie R3 (five-fingered robot) and Nao robot (three-fingered robot) were tested by the Istanbul Technical University for tutoring sign language in adults and children with typical hearing [5,6]. This work proved the relevance of the hand anthropomorphism in sign language vocabulary comprehension. There are other studies regarding the design and development of robotic hands which have covered this topic independently from a humanoid robot, as it is the case of Project Aslan, from the University of Antwerp, which consists in a text dactylology translator arm [7].

Participatory Design (PD) has been considered, since involving users, designers and technology in a process of development and obtaining a distinct and diverse set of perspectives is highly valuable when developing a universal user oriented product [8]. It is important to take into consideration that the representation of sign language in CPS may be controversial without the feedback and participation of deaf and hearing-impaired people in the signing learning and implementation process. It is important to meet the expectations and needs of the target audience of this work before investing time and resources in specialising robots in certain areas. That is the main principle underlying this project.

1.2. TEO as a Household Companion

TEO, also known as RH-2, is a full-size humanoid robot developed by researchers at the Robotics Lab research group, from UC3M. It features 28 Degrees of Freedom (DOF), two actuated hands and several sensors to provide it with information about its environment.

Regarding manipulation, TEO features two 6 DOF arms, each with a five-finger dexterous hand, which can be seen in Figure 1. Thanks to their anthropomorphic characteristics, humanoid robots can perform human tasks such as greetings, waiter functions, folding and unfolding clothes, ironing and painting [9,10]. Task performance is achieved by perception-manipulation loops through a variety of machine learning techniques.

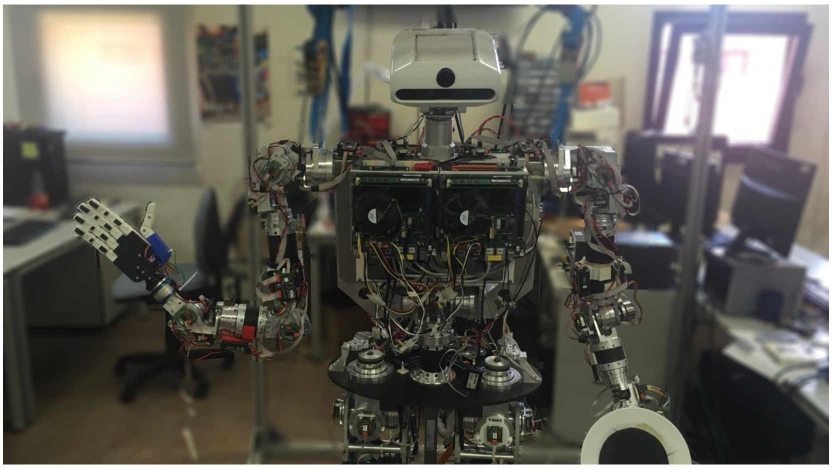

Figure 1. Humanoid robot TEO performing informal greeting with Dextra TPMG90-2 robotic hands.

As shown in the previous subsection, developing a robotic interpreter is an ambitious long-term project, since grammar, dialects, idioms and facial expression analyses would be needed. Currently, human–robot interaction with TEO relies on short command sentences delivered in both directions, so it is an affordable start point to test the user acceptance. To illustrate the interaction mentioned before, performing a greeting would consist in TEO using its voice to ask for the user name and, right after receiving that information, a short welcome sentence would be sent through the speakers. Therefore, the point of this work is to ensure this kind of communication can be established via sign language.

1.3. TEO Robotic Hands and Sign Language

The development and adaption of new anthropomorphic robotic hands for TEO started in September 2017. Dextra TPMG90-2 is the version name of the current undergraduated hand prototypes operative and available in the robot [11]. They each have 15 DOF (14 for flexion/extension and 1 for abduction/adduction) and 6 actuators. The motion transmission system is based on a tendon-driven mechanism.

Underactuation could have been an issue regarding adaptability and precision, since each single actuator is in charge of flexing and extending all the phalanges of a single finger, with the exception of the thumb which is governed by two actuators. Contrary to this assumption, due to the phalange inner design depicted in Figure 2, the finger shows a natural gradual joint rotation that starts from the proximal phalange and allows the hand to develop movements similar to the one of the human hand.

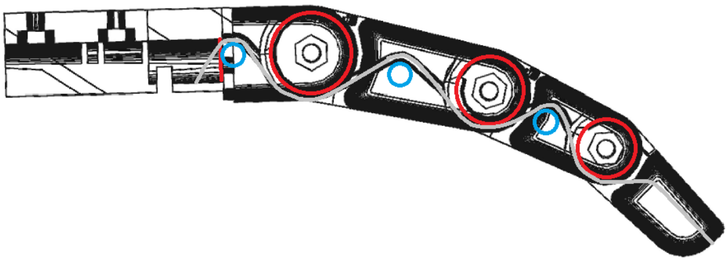

Figure 2. The phalange inner design allows the finger to show a natural gradual joint rotation starting from the proximal phalange.

Dactylology or fingerspelling requires a certain degree of position accuracy. Figure 3 shows how Dextra TPMG90-2 is able to represent the complete Spanish Sign Language (LSE) dactylology. This dactylology and its outcome demonstrate how reasonable is to expect a positive performance in robotic hand signing. Since the hand is able to reproduce the complete alphabet, the following step is to test it with deaf and hearing-impaired users not related to the project to obtain and evaluate feedback.

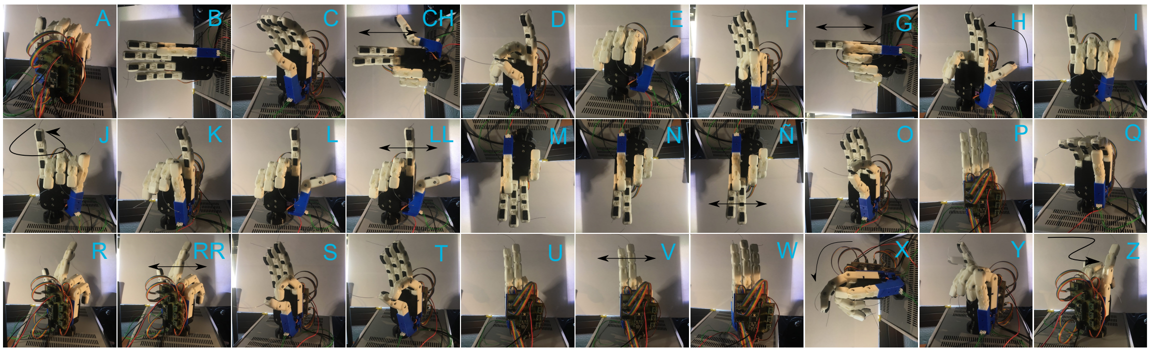

Figure 3. Spanish dactylology developed by Dextra TPMG90-2 robotic hand for joint position configuration.

2. A Preliminary Study: Subtitles or Sign Language

A general solution to procure deaf and hearing-impaired accessible communication in media and technology is to display subtitles. This settlement presents some advantages, such as ease of understanding, speed, or simplicity of implementation; and some disadvantages, such as the need of users' literacy, or the requirement of a sufficiently big readable screen.

To measure the target audience preferences regarding human–robot interaction, a preliminary study is performed in this section to obtain the rate of users that prefer sign language over subtitles in this assistive robotics context, before and after watching a TEO humanoid robot demonstration. These preferences are asked and shown as it is important to measure the user interest regarding the use of sign language within the context of humanoid robotics, before engaging in deeper studies.

2.1. Preliminary Study Experimental Setup

A group of 16 anonymous deaf and hearing-impaired users were recruited in collaboration with CILSEM (Spanish Sign Language Interpreters of Madrid Association) and Signapuntes Lengua de Signos (an LSE forum) and asked to choose between using sign language or subtitles with a humanoid robot, before and after watching a demonstration in which the robot asks "how are you?" in LSE. The sampling group consists of 16 Spanish men and women between 22 and 56 years old. The only characteristic taken into consideration for this study is the users' age, as the generational factor is considered to be the determining factor to measure users' predisposition to interact with or use technology.

A statistical test is carried out to check the consistency in responses across the two options: sign language or subtitles. The same question is delivered on more than one occasion for each of the individuals included in the investigation, so the focus is on comparing whether the measurements made at two different times are the same or if, on the contrary, there is a significant change. McNemar's test [12] fits perfectly for this purpose, since the data has one nominal variable with two categories and one independent variable with two connected groups, the sample is random, and sign language and subtitles are mutually exclusive [13].

The importance of delivering this multiple choice test prior to the comprehension test needs to be highlighted. If most users refuse the idea of using LSE to interact with a robot in both cases, the utility of the project should be reconsidered. Otherwise, if any or both of the cases receive a positive feedback, there would be sound arguments to continue with the research.

2.2. Preliminary Study Results

The experimental outcome is shown in Figure 4. The user's predisposition to communicate with robots was over 80% positive, and more than 65% of reticent users changed their minds after their first experience with TEO. The experimental outcome predicts a positive response to human–robot interaction. However, a statistical analysis is needed to ensure this, which is performed in this section.

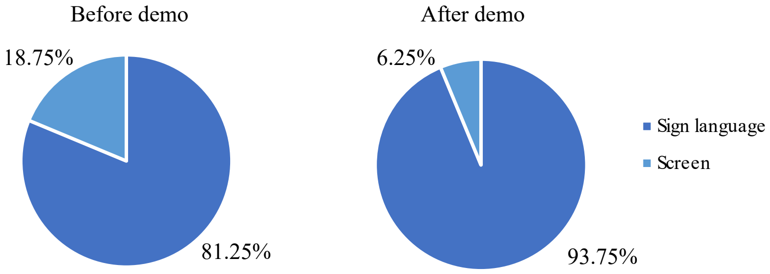

Figure 4. User preference rate regarding human–robot interaction before and after interacting with TEO.

Table 1 cluster the data, before and after demonstration, to analyse it via McNemar's test. If there were no association between the results before and after demonstration, it is reasonable to expect the number of pairs where users before demonstration preferred sign language but users after demonstration did not (top right), to equal the number of pairs where the users after demonstration preferred sign language but the users before demonstration did not (bottom left). In this study, there were two discordant pairs (results before demonstration and results after demonstration had different exposure to the demonstration factor). There were a 100% of pairs where users after demonstration preferred communicating via sign language but users before demonstration did not (bottom left), and no pairs where users before demonstration preferred communicating via sign language but users after demonstration did not (top right).

Table 1. McNemar's test table that shows user preferences regarding human–robot interaction before and after the demonstration.

		After demonstration		
		Sign language	**Subtitles**	**Total**
Before demonstration	**Sign language**	13	0	13
	Subtitles	2	1	3
	Total	15	1	16

Under the null hypothesis, with a sufficiently large number of discordants (elements of the antidiagonal), the chi-square (χ^2) test indicates that the distribution of the samples is chi-squared with 1 degree of freedom.

$$\chi^2 = \frac{(b-c)^2}{b+c} \tag{1}$$

When the elements of the antidiagonal sum less than 25, it is not well-approximated by the chi-squared distribution [14]. An alternative to the chi-squared distribution is the exact binomial test:

$$\text{exact-P-value} = 2\sum_{i=b}^{n}\binom{n}{i}0.5^i(1-0.5)^{n-i} \tag{2}$$

Edwards proposed a continuity corrected version of the McNemar test to approximate the binomial exact-P-value, which is the most widely used variant nowadays [15]:

$$\chi^2 = \frac{(|b-c|-1)^2}{b+c}. \tag{3}$$

From Equation (3), chi-squared equals 0.500 with 1 degrees of freedom. The P value is calculated with McNemar's test with the continuity correction and shows the probability of observing a large discrepancy between the number of the two kinds of discordant pairs. The two-tailed P value equals 0.4795. By conventional criteria, this difference is considered to be not statistically significant. Therefore, the percentage difference after and before watching TEO's demonstration can be attributed to chance and there is no consistent evidence of the effectiveness of TEO's performance in increasing the liking or interest rate. The odds ratio and its confidence interval cannot be calculated because one of the discordant values is zero.

3. Experimental Setup: Materials and Methods

The first decision-making regarding the comprehension test setup is to consider how this test would be distributed. In order to preserve coherence in this experimental test, it is decided to keep using an anonymous online form distributed by LSE institutions and simulation-based multimedia files. There are several reasons for using simulation. On the one hand, this study aims to present the experiments in simulation as a first step within long-term work, where further studies will be performed with the physical humanoid robot. Therefore, the simulation outcome allows us to anticipate the effects of the embodiment and the robot appearance on user satisfaction and comprehension. On the other hand, it is convenient to use a neutral background and a simplified representation of the humanoid robot that allows the respondents to focus on the gestures, since a non-neutral background could affect the quality of the gesture identification.

TEO's signing simulation is developed by using OpenRAVE and QtCoin viewer, which provides a suitable environment for testing and developing. For that purpose, XML files were created to store all robot and scene descriptions. An example of this simulation can be found in Figure 5.

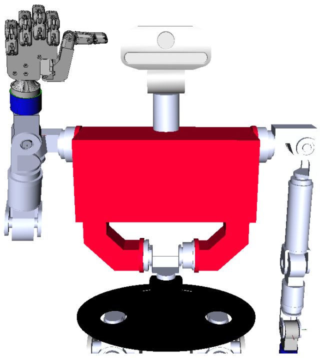

Figure 5. Frame of TEO's simulation signing letter E for the dactylology test. Vocabulary obtained from CNSE Foundation for the Suppression of Communication Barriers images and signs LSE database.

Usability testing is used to observe how easy to use sign language with TEO is by testing it with end-users. Participants are asked to complete these tests to detect problematic or confusing situations. Regarding the required number of participants to get acceptable results, Virzi [16], and more recently Lewis [17] and Turner [18], have published influential articles on the topic of sample size in usability testing. According to these authors, five is a proper number for usability testing, so counting with 16 samples would be enough to develop a precise and reliable study and reach a successful conclusion [19].

The subjects of the test are randomly selected deaf and hearing-impaired subjects, contacted by CILSEM and Signapuntes Lengua de Signos. There is no detailed information given before the beginning of the test, and they are kindly asked to complete a form to obtain feedback about a signing humanoid robot. As commented before, the test is completely anonymous. The only personal information collected from the participants is their age, in order to detect any tendency regarding preferences or understanding.

The developed test consists in two main parts: dactylology and vocabulary recognition. These two tests are selected to cover the study of the hand signing accuracy and the ability to communicate by using the upper part of the robot body. After the comprehension test with TEO, the user is ready to measure their satisfaction, so they will be asked to answer some questions about their experience.

Every test section is compulsory, which means that the responses cannot be submitted until the whole test is completed. There are just three additional optional questions about user preferences which can be completed at the end of each section.

3.1. Dactylology

Fingerspelling needs to be precise to be understood properly. There are some letters in LSE which share a quite similar hand configuration, so transitions, speed, and arm orientation must be treated carefully to obtain good results. It must be taken into account that TEO does not include anything similar to a human mouth, so it is not possible to aid the understanding of the signs with lip-speaking.

The confusion matrix of a class problem is a square matrix in which the columns are named according to the expected result, and the rows are named according to the experimental results. This kind of matrix is the tool selected for showing explicitly when one letter is confused with another letter. It is a powerful tool since it allows to work separately with different types of errors.

The selected tool needs the provided test to check each one of the 30 letters of the Spanish alphabet. In order to avoid predictability and check if transitions between letters may cause any kind of confusion, the letters are shown in groups of three, so the user is asked to fill 10 blank gaps with 3 letters each. The letters are represented in a loop, so the first frame of each loop is marked with a blue dot to help the user to identify the beginning of the letter signing.

3.2. Basic House Vocabulary

The representation of sign language vocabulary involves the action of the upper body, which includes hands, arms and head. This makes it specially important to coordinate all the simultaneous movements to make them seem human-like and, therefore, be more understandable by the end-user.

The tested vocabulary is selected according to the household companion context and considering some similar words to make it possible to apply the confusion matrix in this case, as well as in the dactylology test. There were nine related words and one unconnected word. "Iron" is the only unconnected word, selected due to its significance, since ironing is one of the most complex and relevant tasks that TEO can develop. "Machine" and "clothes"; "door", "kichen" and "closet"; "bedroom" and "table"; and "living room" and "telephone" are the related words that are expected to lead to confusion. Figure 6 shows an example of the kind of similarity tested, where the arms' movement is quite similar in both cases, and the position of the fingers is fundamental to understand the difference in meaning.

In this case, as house vocabulary comprises a much wider group of words than the Spanish alphabet and to avoid obtaining unexpected results that could affect the confusion matrix and the following study, the users have to select the word from a ten choices drop-down list. Each word is shown independently, so, in accordance with the dactylology test, each user submits a ten-time outcome.

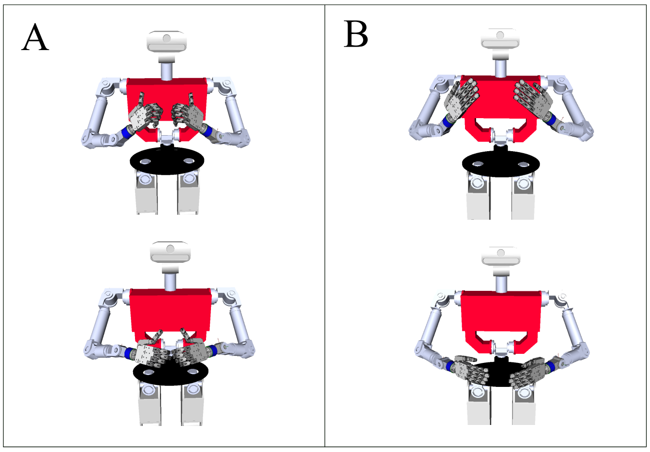

Figure 6. Initial and final frames of TEO's simulation signing (**A**) "machine" and (**B**) "clothes" for the vocabulary test. Vocabulary obtained from CNSE Foundation for the Suppression of Communication Barriers images and signs LSE database.

3.3. User Satisfaction

An important part of this experimental test is to measure satisfaction, since it is fundamental that the end-users are not only able to communicate with the humanoid robot by using sign language, but that they can also do it in the most comfortable way. Six topics have been considered in order to measure user satisfaction, inspired in the users' overall valuation test developed for the ASIBOT assistive robot [19], which are the following:

- **Aesthetics**: Although TEO is still in an experimental phase and the way it currently looks is temporary. The outcome shows the way this topic affects the interaction experience.
- **Anthropomorphism**: The degree of anthropomorphism or human resemblance of the humanoid robot may influence the emergence of the uncanny valley [20], so it must be taken into consideration.
- **Future prospects**: Since the technology shown in this test is under development, it is important to know if the user is willing or not to use it in the near future.
- **Comfort**: Uncomfortable experiences should not be present in assistive robotics, since these robots are made to work in close interaction with people; therefore, comfort must be handled properly.
- **Comprehension ease**: The user may find some difficulties to comprehend the way TEO reproduces LSE which sometimes cannot be completely detected by error-proofing tests.
- **Usefulness**: Although preferences regarding robot communication are asked at the beginning of the form, end-users might consider human–robot interaction pointless after the tests.

The Likert scale is a measurement tool that, unlike binary questions that can be answered affirmatively or negatively, allows to measure attitudes and know the degree of conformity of the respondent with any proposed statement [21]. It is especially appropriate in this context in which we want our end-users to provide their opinion quantitatively. In this sense, the response categories will serve to capture the intensity of the respondent's feelings toward each affirmation.

The most important requirement in this scale is that the distance between each possible answer choice is the same. It allows quantitative studies across different covered topics that have more than two outcome values [22]. There is no clear consensus among researchers about the number of response levels. The most commonly used scale consists in five levels; but four, seven, or ten levels are also frequently used [23]. Adding levels results in obtaining more diverse valuations, as it avoids central tendency bias (CTB). CTB theory explains that in an item of only five levels, participants tend to avoid the two extreme options, obtaining very little variation.

The CTB effect could be softened by balancing positive and negative levels in the scale (symmetric scale) and letting the user respond to the test anonymously to avoid the pressure of being judged for selecting extreme options. A symmetric scale could also help to avoid acquiescence bias, which is a tendency of the respondent to agree or show positive feedback [24]. Since a five-level scale allows neutral response and two different levels of agreement and disagreement, which simplifies decision taking, it is selected for the user satisfaction test. Table 2 shows the displayed options in the final survey.

Table 2. Five-level Likert scale used in the user satisfaction questionnaire.

Scale Value	Opinion
−2	Strongly disagree
−1	Disagree
0	Neither agree nor disagree
1	Agree
2	Strongly agree

It is difficult to treat neutral responses, such as the "neither agree nor disagree" presented in the table, but it is recommended to offer the possibility of taking this option if the respondent is unsure about their opinion or cannot decide between a positive or a negative answer. About considering a middle option as "unsure" or "neutral", a study developed by R. Amstrong found the differences to be imperceptible [25].

3.4. Optional Questions

Some optional questions are provided in the delivered form at the end of each previous described test sections, to obtain further information regarding the respondent preferences. The three questions presented in the questionnaire are:

- Regarding human–robot interaction, would you prefer any alternative method to using sign language or reading subtitles?
- Why do you prefer the way of interacting with TEO that you selected?
- What would you improve about TEO signing performance?

The answers to these questions could provide additional details that would help us to understand some issues that need to be fixed in future developments.

4. Experimental Results

Experimental results were collected two weeks after delivering the online form to the institutions involved in its distribution. This limit on the period of time for receiving the form was established to assure distribution only within the reach of the target end-user group, as the link to the form was open and based on trust of anonymous user data. A total of 16 users participated up to that date.

4.1. Dactylology

Dactylology answers, provided the fact that robot movements were programmed by LSE non-experts, were surprisingly accurate and insightful. Table 3 depicts the confusion matrix that compares expected and obtained results. The elements in the main diagonal show the amount of correct answers for each specific letter. The 'Other" row contains the sum of answers that are not elements of the expected answers. It is noticeable at first sight that, except for the letters F and RR which will be commented within this section, all letters obtained a correct answer rate above 50%. One third of the alphabet was completely understood (10 letters), with no failed attempts (discarding outlier answers). Finally, the mean shows that approximately 82% of the answers were correct (369 correct answers over 450), which can be considered a successful outcome.

Table 3. Confusion matrix: dactylology. The elements of diagonal, which represent correct answers, are marked in bold. Elements with a shaded background mean 100% correct answers, discarding those of the "Other" row. In the Spanish alphabet, the CH, LL, RR and Ñ represent individual letters.

		Result Expected																													
		A	**B**	**C**	**CH**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**LL**	**M**	**N**	**Ñ**	**O**	**P**	**Q**	**R**	**RR**	**S**	**T**	**U**	**V**	**W**	**X**	**Y**	**Z**
Result Obtained	**A**	**14**	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	B	0	**13**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	C	0	0	**15**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	CH	0	0	0	**15**	0	0	0	0	4	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	D	0	0	0	0	**15**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	0	0	0	0	0	0	0
	E	0	0	0	0	0	**13**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	F	0	0	0	0	0	0	**6**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0
	G	0	0	0	0	0	1	0	**15**	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	H	0	0	0	0	0	0	0	0	**9**	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	I	0	0	0	0	0	0	0	0	0	**13**	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1
	J	0	1	0	0	0	0	0	0	0	0	**12**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	6
	K	0	0	0	0	0	0	0	0	1	0	0	**8**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	L	0	0	0	0	0	0	0	0	0	1	0	0	**15**	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	LL	0	0	0	0	0	0	0	0	0	0	0	0	0	**12**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	M	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**15**	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0
	N	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**15**	6	0	0	0	0	0	0	0	0	1	0	0	0	0
	Ñ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**9**	0	0	0	0	0	0	0	0	0	0	0	0	0
	O	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**14**	0	0	0	0	0	0	0	0	0	0	0	0
	P	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	**11**	0	0	0	0	0	0	0	4	0	0	0
	Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**15**	0	0	0	0	0	0	0	0	0	0
	R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**15**	6	0	0	2	0	0	0	0	0
	RR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**7**	0	0	0	1	0	0	0	0
	S	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	**14**	0	0	0	0	0	0	0
	T	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**13**	0	0	0	1	0	0
	U	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**11**	2	0	0	0	0
	V	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	**11**	0	0	0	0
	W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	**10**	0	0	0
	X	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**13**	0	0
	Y	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**13**	0
	Z	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	**8**
	Other	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

4.1.1. Individual Letter Error Analysis

Taking a deeper look at the matrix helps to clarify the source of errors in individual letter recognition. The most controversial letters, of which the initial frames are shown in Figure 7, are F, H, K, Ñ, RR and Z, with a correct answer rate less or equal to 75%. An independent study for each letter is convenient to identify causes of confusion.

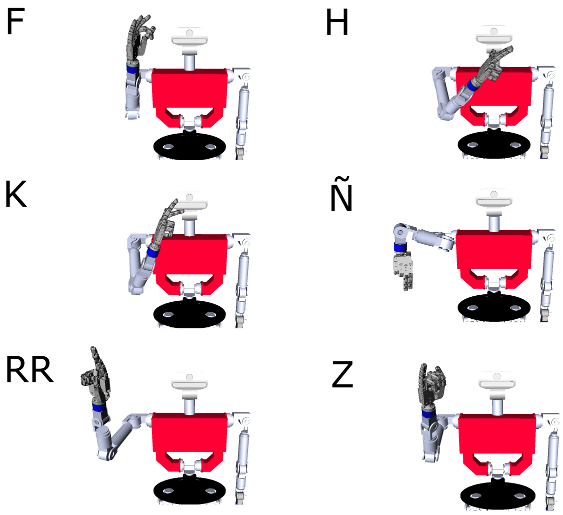

Figure 7. Initial frames of challenging letters in dactylology. *Dactylology obtained from CNSE Foundation for the Suppression of Communication Barriers images and signs LSE database.*

- Letter F is the most erratic letter of the experiment. It was mistaken for letter T in 60% of the attempts. These letters have relatively similar finger position configuration, as can be seen in Figure 3. It is remarkable that only 13% of users mistaken letter T for letter F. Reviewing the F simulation file its has been noticed that thumb position may have resulted confusing, and it has been modified for future experiments.
- Letter H is mistaken for letter CH in almost 27% of the attempts. Both letters have the same finger position configuration, but they differ in arm movement. Letter CH has not been mistaken for letter H in any attempt so the arm movement for CH was pronounced after this analysis.
- Letter K is mistaken for letters H and P in 20% of the attempts each. In this case, there is over a 7% of coincidence regarding letter H and no coincidence at all regarding letter P. Letter K is a specially complex case, since the position of the middle finger is not so evident as it is in other letters and there must have been some implementation errors that should be rectified with the help of an LSE expert.
- Letter Ñ is mistaken for letter N 40% of the attempts. The only difference between these letters is that letter N is static and letter Ñ requires movement. The solution provided to decrease this error is to make the movement more noticeable to avoid being confused with a letter transition.
- Letter RR is mistaken for letter R in 40% of the attempts. The casuistic is exactly the same as in the Ñ-N case. Therefore, the same solution is provided.
- Letter Z is mistaken for letter J in 40% of the attempts. Both letters need motion and they share finger position configuration with letter I. Letter J performs a circular movement while letter Z performs a zig-zag movement. The second one was developing this movement in an almost horizontal plane, so it was not easily understandable. The solution was to change the angle of movement execution.

Some other letters show small inaccuracies of which sources are not as immediately perceptible as these previous ones, so further analysis is required to find new root causes.

4.1.2. Age Influence in Dactylology

Figure 8 shows the relation between the number of correct answers and the age of the users. The negative slope of the linear trendline shows a light tendency towards misunderstanding the dactylology developed by TEO in relation to age increase. To measure letter transition understanding, the answer is considered correct only if the user is able to understand the complete set of three letters, which means that the movements between letters have not influenced the correct perception of the dactylology.

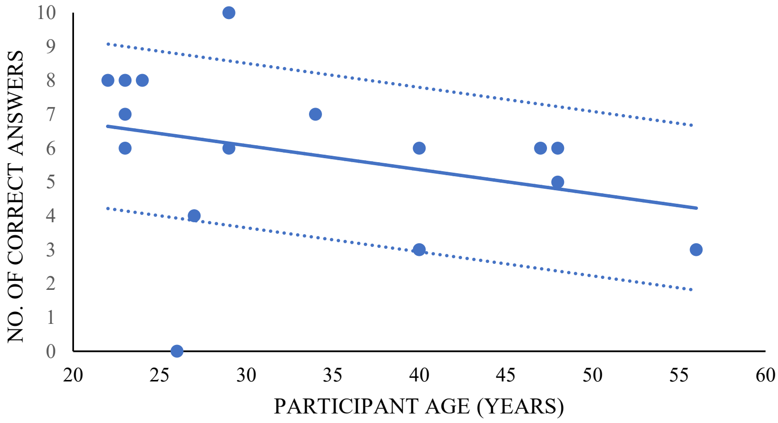

Figure 8. Graph of number of correct answers per participant in dactylology test versus participant's age (years) with linear trendline and linear regression channel which contains approximately 68% of all answers.

The regression channel, which is the area between dotted lines in Figure 8, is based on the linear regression that represents a simple trendline that is projected using the least squares method. Consequently, this line turns out to be an average line of the correct answer rate that is changing. It can be considered as an "equilibrium" result line, while any deviation from it up or down indicates the higher activity of correct or wrong answers, respectively [26]. The distance between the channel bands and the regression line is equal to the standard deviation value of the correct answer rate with respect to the regression line. The upper and lower channel lines therefore contain between themselves approximately 68% of all user answer data.

For this dactylology test d, the trendline equation and the coefficient of determination R_d^2 obtained by the least squares method are shown in Equations (4) and (5), respectively. Since the regression line is relatively far from some of the points, the R_d^2 of the regression is quite low.

$$y_d = -0.071x_d + 8.2056 \tag{4}$$

$$R_d^2 = \frac{\sigma^2_{X_d Y_d}}{\sigma^2_{X_d}\sigma^2_{Y_d}} = 0.1076 \tag{5}$$

The standard deviation s_d for the regression channel included in Figure 8 is shown in Equation (6). Using this value and Equation (4), upper and lower lines on the regression channel are drawn.

$$s_d = \sqrt{\frac{\sum_{i=1}^{N}(y_{d_i} - \overline{y_d})^2}{N-1}} = 2.4281 \tag{6}$$

For these sets of three words, this standard deviation shows that a range of dactylology understanding approximately between 42% and 90% can be expected from users in their early twenties, in comparison to the 18–67% approximated range for middle aged people.

It should be taken into consideration for this analysis that there is an outlier due to one user which, through a manual review of the answers, can be determined to have answered the comprehension tests arbitrarily. Consequently, if this outlying data is omitted, the data presented in Equations (4)–(6) presents the variations presented in Equations (7)–(9), respectively.

$$y_{d'} = -0.0996x_{d'} + 9.608 \tag{7}$$

$$R^2_{d'} = 0.3453 \tag{8}$$

$$s_{d'} = 1.9346 \tag{9}$$

As expected, this change presents a steeper negative slope and the coefficient of determination R^2_d has increased more than three times while the standard deviation s_d has decreased. As this model is more adjusted to the variable when the outlier is omitted, it can be concluded that the tendency to misunderstand the dactylology in relation to age is more pronounced than previously stated.

4.2. Basic House Vocabulary

Vocabulary test results were significantly more positive than the dactylology ones. The correct answer rate can be checked in the confusion matrix shown in Table 4.

The average of correct answers per user is 13.3, which means a 83% of success rate (133 correct answers over 160 answers). The lowest understood word achieved a 62.5% of correct answers (10 correct answers over 16 answers), so almost two thirds of the users were able to understand even the most challenging words. This result in vocabulary understanding was expected, since word signing, in this particular case, does not require a high level of detail and it is more figurative than fingerspelling.

4.2.1. Vocabulary Error Analysis

In order to detect some irregularities and check if the groups of similar words produced confusion among users, a detailed error analysis is developed.

- “Machine” and “clothes” are two words in LSE that are similar, since the main difference between them is the position of the hand, but the arms develop relatively the same movement. “Machine” was mistaken for “clothes” in almost 19% of the answers. The word “clothes” was however never mistaken for “machine”. Since the difference between both words is a matter of open/close fist variation it is possible that users are not so used to the word “machine” or even that the first word that appeared in the drop-down list was “clothes”.
- “Door”, “kichen” and “closet” are words that require a similar arm and hand movement, with some variations in the order the hands are positioned. “Door” was mistaken for “closet” 25% of the attempts and only a 13% in the inverse order. This difference may be attributed to simulation, since the hand position order was correctly developed. “Kitchen” was mistaken for “door” over a 6% of the attempts but not a single time in the opposite way, so it is not considered significant. There were no connections at all between “closet” and “kitchen” in either direction.
- “Bedroom” and “table” could have been confused since they share similar movements, but they were not confused at any time.
- “Living room” and “telephone” are two words that require signing in the head area. “Telephone” was identified 100% of the attempts, which is an interesting rate, considering there is one user who submitted most of their answers wrong. The reason for this accuracy may be explained through the fact that the Spanish sign for “telephone” could be understood internationally without any LSE knowledge. “Living room” was not confused at any time with “telephone”, but it was confused approximately 13% of the attempts with “machine”, which is an outcome that cannot be explained from the consulted LSE signing database point of view.

Table 4. Confusion matrix: basic house vocabulary. The elements of diagonal, which represent correct answers, are marked in bold. Elements with a shaded background mean 100% correct answers (15 correct, and 16 in the exceptional case where there was a correct answer within the outlying data).

		Result Expected									
		Table	**Door**	**Bedroom**	**Closet**	**Telephone**	**Machine**	**Kitchen**	**Clothes**	**Iron**	**Living Room**
Result Obtained	**Table**	**14**	0	0	0	0	0	0	1	0	0
	Door	0	**10**	2	2	0	0	1	0	0	0
	Bedroom	0	2	**14**	0	0	0	0	0	0	0
	Closet	0	4	0	**11**	0	0	0	0	0	1
	Telephone	0	0	0	0	**16**	0	0	0	0	1
	Machine	0	0	0	0	0	**12**	0	0	1	2
	Kitchen	1	0	0	0	0	0	**14**	0	0	0
	Clothes	0	0	0	0	0	3	0	**15**	0	0
	Iron	0	0	0	0	0	0	1	0	**15**	0
	Living Room	1	0	0	3	0	1	0	0	0	**12**

An unexpected result was the confusion between "closet" and "living room", with about 19% error rate in the mentioned order, and just a 6% in the inverse order. There is no relation at all between the way both words were simulated, so it may indicate an implementation error or may be biased by frequency of everyday use.

The only independent word, which is "iron", did not represent any challenge for users, since it presented a 94% of success rate. Some other words also shown small inaccuracies, which sources are relatively difficult to determine.

4.2.2. Age Influence in Vocabulary

Figure 9 shows the relation between the rate of correct answers and the age of the users. The negative trendline slope shows a even lighter tendency to misunderstand the vocabulary in relation to age than the one presented in Figure 8.

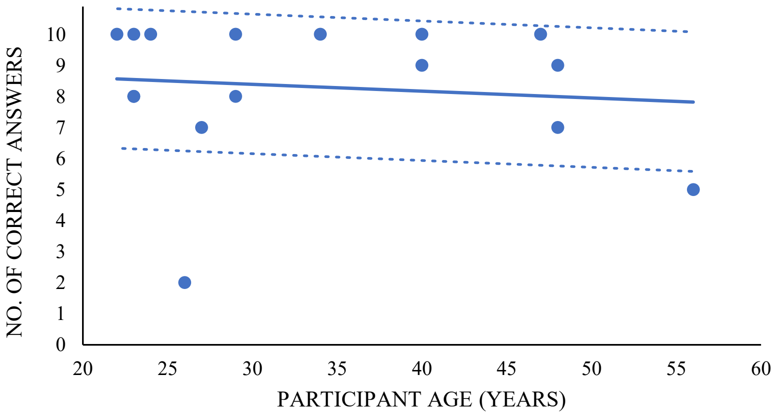

Figure 9. Graph of number of correct answers per participant in vocabulary test versus participant's age (years) with linear trendline and linear regression channel which contains approximately 68% of all answers.

For this vocabulary test v, the resulting equation used to draw the trendline and the coefficient of determination R_v^2 are shown in Equations (10) and (11). As expected, the slope of the negative trendline is less than one third of the dactylology trendline slope, which means that the tendency to misunderstand sign language using words in relation to age is almost insignificant. Since the regression line is relatively far from a high percentage of the points, the R_v^2 of the regression is quite low.

$$y_v = -0.022x_v + 9.0529 \tag{10}$$

$$R_v^2 = \frac{\sigma_{X_vY_v}^2}{\sigma_{X_v}^2\sigma_{Y_v}^2} = 0.0121 \tag{11}$$

The standard deviation for the regression channel included in Figure 9 is shown in Equation (12). Using this value and Equation (10), upper and lower lines on the regression channel can be drawn.

$$s_v = \sqrt{\frac{\sum_{i=1}^{N}(y_{v_i} - \overline{y_v})^2}{N-1}} = 2.2425 \tag{12}$$

This standard deviation shows that a range of vocabulary understanding between 63% and 100% can be expected from users in their twenties, quite similar to the 56–100% range for middle aged people.

If the outlier is also not being considered in this case, the data presented in Equations (10)–(12) presents some variations, which are presented in Equations (13)–(15), respectively.

$$y_{v'} = -0.0511 x_{v'} + 10.482 \tag{13}$$

$$R^2_{v'} = 0.1447 \tag{14}$$

$$s_{v'} = 1.533747356 \tag{15}$$

As occurred with dactylology, this change presents over a double steeper negative slope, and the coefficient of determination R^2_v has increased almost twelve times, while the standard deviation s has decreased significantly. As this model is more adjusted to the variable when the outlier is omitted, it can be concluded that the tendency to misunderstand the dactylology in relation to age is slightly more pronounced than previously stated.

Considering the obtained data, it can be concluded that this high vocabulary understanding correct answer rate and this small pronounced slope in comparison with the dactylology outcomes may be due, not only to the signing simplicity, but also to the fact that letters were displayed in sets of three, while words were tested independently and not in a sentence in order to simplify analysis, so it is understandable that the error rate decreases.

4.3. User Satisfaction

Table 5 shows the satisfaction questionnaire individual results, sorted by age and measured in a [−2, 2] Likert scale. Average overall user satisfaction over this experimental work results in a promising 0.78 (69.5%), roughly between a neutral and positive position.

This data is grouped and analysed with the purpose of drawing relevant conclusions. Figure 10 gives a breakdown of this outcome, where no negative mean values can be observed, but some relevant different satisfaction levels are found.

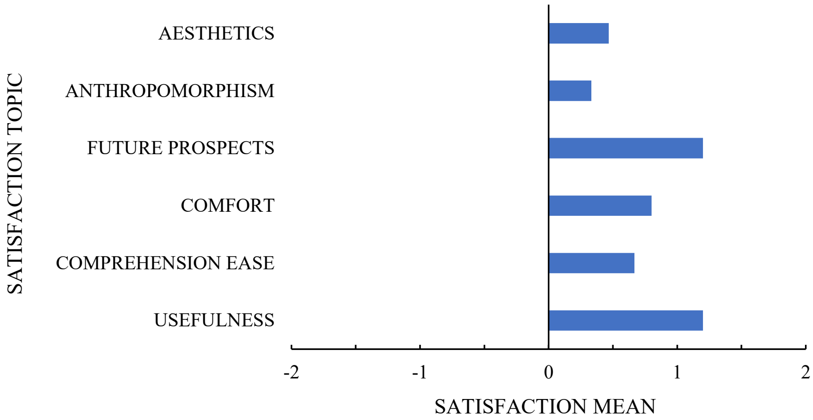

Figure 10. Overall satisfaction classified by topic. A five level Likert scale has been used, where the values mean: (−2) Strongly disagree, (−1) Disagree, (0) Neither agree nor disagree, (1) Agree, (2) Strongly agree.

Table 5. Satisfaction topics outcome classified by age in ascending order. A five level Likert scale has been used, where the values mean: (−2) Strongly disagree, (−1) Disagree, (0) Neither agree nor disagree, (1) Agree, (2) Strongly agree.

		Satisfaction topics						**Satisfaction mean**
		Usefulness	**Comprehension ease**	**Comfort**	**Future prospects**	**Anthropomorphism**	**Aesthetics**	
Age	**22**	1	2	0	1	1	1	1.00
	23	2	2	2	2	2	2	2.00
	23	2	2	2	2	2	2	2.00
	23	1	−1	1	2	−1	−1	0.17
	24	2	1	1	2	1	0	1.00
	26	1	1	1	1	1	1	1.00
	27	1	0	2	2	1	2	1.33
	29	2	2	2	2	2	2	2.00
	29	2	2	0	1	1	-1	0.83
	34	2	2	2	2	1	2	1.83
	40	1	−1	−1	1	−1	−1	−0.33
	40	0	−1	−1	−1	−1	0	−0.67
	47	1	1	1	1	1	1	1.00
	48	2	−1	2	2	0	−1	0.67
	48	1	1	−1	0	−1	−1	−0.17
	56	−1	0	−1	−1	−2	−2	−1.17

Top valued topics were future prospects and usefulness, with a 1.2 average satisfaction or, which is the same, an 80% positive feedback. This result demonstrates the user willingness to use this technology, and the high level of expectation the use of LSE with a humanoid robot this first contact has generated.

Comfort and comprehension ease, with 0.8 (70%) and 0.7 (67%) of average satisfaction, respectively, occupy the following positions in the ranking. A reasonable explanation to find these topics lower rated than the previous ones is that there are various letters and vocabulary which have presented some understanding difficulties and have led to confusion. In any case, as proved, these minor inconveniences have not influenced the user expectation. Finally, the least favourable marks are aesthetics and anthropomorphism, with a 62% and 58%. These topics are closely associated to the robot appearance. Since TEO is still being developed at both software and hardware level, it is comprehensible that there are divergent opinions about the way it looks. In either case, this nearly neutral anthropomorphism user perception should not be interpreted as a negative outcome, since resemblance to a human being is not only unnecessary, but also a characteristic to be avoided in assistive robotics.

4.3.1. Age Influence in Satisfaction

Age influence in overall user satisfaction is related to its influence in dactylology and vocabulary understanding. Figure 11, where the satisfaction-age relation is shown, presents a negative trendline that goes through the neutral line, so it is the first graphic in which the trend drops almost a 50% from the youngest to the oldest age.

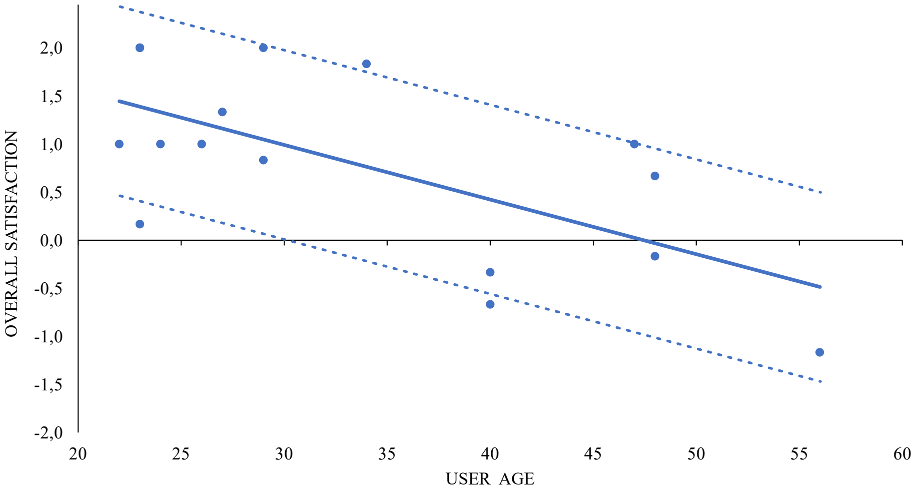

Figure 11. Graph of user satisfaction relation versus age (years) with linear trendline and linear regression channel which contains approximately 68% of all answers. A five level Likert scale has been used, where the values mean: (−2) Strongly disagree, (−1) Disagree, (0) Neither agree nor disagree, (1) Agree, (2) Strongly agree.

For this satisfaction questionnaire s, the resulting equation used to draw the trendline and the R_s^2 value are shown in Equations (16) and (17).

$$y_s = -0.0568 x_s + 2.6947 \tag{16}$$

$$R_s^2 = \frac{\sigma^2_{X_s Y_s}}{\sigma^2_{X_s}\sigma^2_{Y_s}} = 0.42 \tag{17}$$

The standard deviation s_s for the regression channel included in Figure 11 is shown in Equation (18). Using this value and Equation (16), the upper and lower lines on the regression channel can be drawn.

$$s_s = \sqrt{\frac{\sum_{i=1}^{N}(y_{s_i} - \overline{y_s})^2}{N-1}} = 0.9827 \tag{18}$$

This standard deviation shows that between 62% and 100% user satisfaction can be expected from users in their twenties, in comparison to a 27% to 62% range for middle aged people. The input from the user that previously provided outlying data is not affecting these results dramatically, and is therefore taken as a valid part of the experiment.

Taking a deeper look into Table 5, further conclusions can be drawn. For users between 20 and 40 years old, the main disadvantages of this work are related to anthropomorphism and aesthetics; while users between 40 and 60 years old, also find quite inconvenient the comprehension ease and comfort. This satisfaction distribution helps to identify what fields need to be improved to reach all users. Although it is reasonable to expect that people around this second group may not be so used to technology as the millennial generation, and that their answers could also be influenced by the online test format, it is convenient to reach a universal level of ease of understanding for the wide range of people that could need to communicate with the robot.

4.4. User Additional Comments

Approximately 63% of users left additional optional comments with their personal opinions. This feedback is highly valuable, since it offers the opportunity to detect further aspects that need improvement, while it sets the basis for communication error understanding.

4.4.1. Alternatives for Human–Robot Interaction

End-users are asked to provide alternative possibilities to the two options considered in this paper: subtitles and sign language. This question was answered by 31% of the participants, of which only one user actually provided an alternative solution: using a robotic mouth which is able to enunciate accurately while using sign language. The rest of opinions encouraged the use of sign language as the best available option.

4.4.2. Justification of Preference

User preference justification must be divided between users which finally selected sign language and the one that selected subtitles as the ideal option after TEO signing demonstration.

The reasons for selecting sign language can be summed up in the following points:

- Sign language is clearer.
- Sign language is more understandable.
- Many deaf people find it difficult to read or interpret a text.
- Interpreting signing is effortless for the user, since they use sign language in a daily basis.

Two points were shared by one user to justify the subtitles selection:

- Lack of facial expression and lip-speaking.
- Depending on the context, a sign can have several different meanings.

Overall, there are various reasons regarding comfort which lead people to prefer interacting with the robot via sign language. Nevertheless, it must be taken into consideration that there are some drawbacks, such as the lack of facial expression, which may hinder the communication.

4.4.3. Proposals for Improvement

The submitted suggestions were of major importance to detect which areas required improvement. There were several individual comments which stressed that the representation of LSE was clear and understanding the robot is a matter of practice. The remaining comments are listed bellow.

- Some words are not sign language or they are not used anymore, such as "living room". This note highlights the importance of working with people specialised in LSE to implement this language.
- Human-like appearance is demanded. Human misconception of what to expect from a robot may be biased due to science fiction culture and may lead some users to feel disillusioned by the humanoid appearance or "behaviour".
- Hand motion seems too rigid. The robot is made of hard materials and actuated by electrical actuators, so it is complicated to reproduce a smooth motion as human muscles can perform.
- Bigger size of the images in the form would be required. This is an important point since it could justify the increasing failure tendency in understanding the robot related to age, considering that the decline of vision generally associated with age.

5. Discussion

Regarding the developed research, detected several challenges that may be addressed in order to enhance the analysis of the results can be described. As mentioned, only the user's age was considered as the main characteristic to evaluate the tests and questionnaire outcome. However, it would be highly useful to ask the participants for their education level and to let them rate their frequency of use and familiarity with technology. These elements could help factor out some outlying responses and classify the data more accurately, especially in future studies where a larger sampling group will be managed.

A ten choice drop-down list has been used in the vocabulary test to measure the performance of the robot. This was done to provide ease and avoid fatigue of the respondents, while simultaneously avoid obtaining a high proportion of outlying responses that could negatively affect the confusion matrix. Possible redesign alternatives in relation to the format of the test within the Participatory Design process essentially fall into one of the following two categories: (1) to have the robot perform more actions, forming complex sentences aiming at completing a full dialogue, or (2) having the respondents provide more custom or personalised answers, moving from a set of closed responses to an open interview format. While these options are not mutually exclusive are definitively appealing, they are prone to lead into the same kind of pitfall, which is: how to quantitatively analyse and evaluate the respondent's answers to obtain statistically relevant results. However, there is an incentive for focusing on how to circumvent these challenges, as the long-term goal of this study is to establish a complete and effective human–robot communication.

Even though there are some potential limitations that need to be handled, such as the need for sign language expertise and the development of a more complex sign language reproduction by TEO, the excellent results obtained with simulation show the importance of focusing on making further advances towards full communication via sign language. One of the considered paths to face these issues is to develop machine learning algorithms to learn from LSE datasets that contain collections of signs performed by professional interpreters. The developed system would additionally enable learning new signs –or in different languages– from data obtained by low-cost sensors.

6. Conclusions

Given the worldwide need for user accessibility and UD in assistive robotics, this work provided a pioneer study of end-user interest, comprehension and satisfaction regarding the reproduction of sign language by a humanoid robot.

The willingness of the end-users of the study towards using sign language with a humanoid robot was almost 94% positive, which is reaffirmed in the user satisfaction questionnaires after the comprehension tests, where usefulness and future prospects are valued with the highest marks. Both dactylology and vocabulary tests resulted in 82% and 83% correct answer rate respectively, with a relatively pronounced tendency to acceptance in relation to a younger age. Most errors encountered on dactylology and vocabulary should be mendable by modifying finger joint configuration or pronouncing the movement, so further iterations of experiments could be performed to prove if the confusing signs are fixed. Most users distinguished the robot appearance as its most remarkable inconvenience, which is a reasonable outcome since the robot used for testing is an experimental platform and its appearance is constantly changing.

The most challenging issue regarding this project has been attempting to reproduce sign language with the lack of facial expressions and other non-manual markers. This circumstance may cause understanding problems to some users and would be a potential barrier regarding the development of more complex communication. Concerning basic instructions communication, the tests have shown a proficient human–robot interaction.

The data collected over these experiments has provided quantitative measurements on end-user satisfaction, as well as useful insight regarding user needs. The experimental results shed light towards new improvements and developments to make assistive robotics and CPS more usable for deaf and hearing-impaired users.

Author Contributions: Conceptualization, J.J.G. and J.G.V.; Data curation, J.J.G.; Formal analysis, J.G.V.; Funding acquisition, C.B.; Investigation, J.J.G. and J.G.V.; Methodology, J.J.G.; Project administration, J.G.V. and C.B.; Resources, C.B.; Software, J.J.G. and J.G.V.; Supervision, J.G.V. and C.B.; Validation, J.J.G.; Visualization, J.G.V.; Writing—original draft, J.J.G.; Writing—review and editing, J.G.V. and C.B.

Funding: The research leading to these results has received funding from the RoboCity2030-III-CM project (Robótica aplicada a la mejora de la calidad de vida de los ciudadanos. fase III; S2013/MIT-2748), funded by Programas de Actividades I+D en la Comunidad de Madrid and cofunded by Structural Funds of the EU.

Acknowledgments: The authors thank CILSEM (Spanish Sign Language Interpreters of Madrid Association) and Signapuntes Lengua de Signos for their kind collaboration with this project.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BOE	Official Gazette of the Spanish Government
CPS	Cyber–Physical Systems
DOF	Degrees of Freedom
LSE	Spanish Sign Language
PD	Participatory Design
TEO	Task Environment Operator
UC3M	University Carlos III de Madrid
UD	Universal Design

References

1. Story, M.F.; Mueller, J.L.; Mace, R.L. *The Universal Design File: Designing for People of All Ages and Abilities*, Revised ed.; Center for Universal Design: Washington, DC, USA, 1998.
2. Spanish Law 27/2007, October 23rd, Which Recognizes the Spanish Sign Languages and Regulates the Means of Support for Oral Communication of Deaf People, Hearing Impaired and Deafblind. [Online]. Boletín Oficial del Estado (BOE), October 2007, No. 255, pp. 43251–43259. Available online: https://www.boe.es/buscar/doc.php?id=BOE-A-2007-18476 (accesed on 1 September 2017).
3. Metaxas, D.; Liu, B.; Yang, F.; Yang, P.; Michael, N.; Neidle, C. *Recognition of Nonmanual Markers in American Sign Language (ASL) Using Non-Parametric Adaptive 2D-3D Face Tracking*; LREC: Boston, MA, USA, 2012.
4. Kelion, L. Toshiba's Robot Is Designed to Be More Human-Like. *BBC News*, 9 March 2016. Available online: https://www.bbc.com/news/technology-35763917 (accessed on 18 September 2018).

5. Kose, H.; Yorganci, R. Tale of a robot: Humanoid Robot Assisted Sign Language Tutoring. In Proceedings of the 2011 11th IEEE-RAS International Conference on Humanoid Robots, Bled, Slovenia, 26–28 October 2011.
6. Uluer, P.; Akalın, N.; Köse, H. A New Robotic Platform for Sign Language. Tutoring Humanoid Robots as Assistive Game Companions for Teaching Sign Language. *Int. J. Soc. Robot.* **2015**, *7*, 571–585. [CrossRef]
7. Goossens, M. Optimisation of a Humanoid Sign Language Robot. Bachelor's Thesis, Universiteit Antwerpen, Antwerpen, Belgium, 2016.
8. Asaro, P.M. Transforming society by transforming technology: The science and politics of participatory design. *Account. Manag. Inf. Technol.* **2000**, *10*, 257–290. [CrossRef]
9. Estevez, D.; Fernandez-Fernandez, R.; Victores, J.G.; Balaguer, C. Improving and evaluating robotic garment unfolding: A garment-agnostic approach. In Proceedings of the 2017 IEEE International Conference on Autonomous Robot Systems and Competitions, Coimbra, Portugal, 26–28 April 2017.
10. Estevez, D.; Victores, J.G.; Fernandez-Fernandez, R.; Balaguer, C. Robotic ironing with 3D perception and force/torque feedback in household environments. In Proceedings of the IEEE/RSJ IROS, Vancouver, BC, Canada, 24–28 September 2017; pp. 6484–6489. [CrossRef]
11. Gago, J.J.; Victores, J.G. Desarrollo e Integración de Mano Robótica Antropomórfica en el Robot Humanoide TEO. Bachelor's Thesis, Dept. Sist. Autom., Universidad Carlos III de Madrid (UC3M), Madrid, Spain, 2018.
12. McNemar, Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* **1947**, *12*, 153–157. [CrossRef] [PubMed]
13. McCrum-Gardner, E. Which is the correct statistical test to use? *Br. J. Oral Maxillofac. Surg.* **2008**, *46*, 38–41. [CrossRef] [PubMed]
14. Fagerland, M.W.; Lydersen, S.; Laake, P. The McNemar test for binary matched-pairs data: Mid-p and asymptotic are better than exact conditional. *BMC Med. Res. Methodol.* **2013**, *13*. [CrossRef] [PubMed]
15. Edwards, A. Note on the "correction for continuity" in testing the significance of the difference between correlated proportions. *Psychometrika* **1948**, *13*, 185–187. [CrossRef] [PubMed]
16. Virzi, R.A. Refining the test phase of usability evaluation: How many subjects is enough? *Hum. Factors* **1992**, *34*, 457–468. [CrossRef]
17. Lewis, J.R. Evaluation of procedures for adjusting problem-discovery rates estimated from small samples. *Int. J. Hum.-Comput. Interact.* **2009**, *13*, 445–479. [CrossRef]
18. Turner, C.W.; Lewis, J.R.; Nielsen, J. Determining usability test sample size. In *International Encyclopedia of Ergonomics and Human Factors*, 2nd ed.; CRC Press: Boca Raton, FL, USA, 2006; Volume 3, pp. 3084–3088.
19. Huete, A.J.; Victores, J.G.; Martinez, S.; Gimenez, A.; Balaguer, C. Personal Autonomy Rehabilitation in Home Environments by a Portable Assistive Robot. *IEEE TSMC Part C (Appl. Rev.)* **2012**, *42*, 561–570. [CrossRef]
20. Mori, M.; MacDorman, K.F.; Kageki, N. The Uncanny Valley [From the Field]. *IEEE Robot. Autom. Mag.* **2012**, *19*, 98–100. [CrossRef]
21. Likert, R. A Technique for the Measurement of Attitudes. *Arch. Psychol.* **1932**, *22*, 1–55.
22. Burns, A.; Burns, R. *Basic Marketing Research*, 2nd ed.; Prentice Hall: Upper Saddle River, NJ, USA, 2007; p. 250.
23. Dawes, J. Do Data Characteristics Change According to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales. *Int. J. Mark. Res.* **2008**, *50*, 61–77. [CrossRef]
24. Watson, D. Correcting for Acquiescent Response Bias in the Absence of a Balanced Scale: An Application to Class Consciousness. *Sociol. Methods Res.* **1992**, *21*, 52–88. [CrossRef]
25. Armstrong, R. The midpoint on a Five-Point Likert-Type Scale. *Percept. Motor Skills* **1987**, *64*, 359–362. [CrossRef]
26. Seal, H.L. The historical development of the Gauss linear model. *Biometrika* **1967**, *54*, 1–24. [CrossRef] [PubMed]