Analysis of Cross-Cultural Effect on Gesture-Based Human-Robot Interaction

Icaro V. Brito, Eduardo O. Freire, Elyson A. N. Carvalho, and Lucas Molina
Federal University of Sergipe, Robotics Research Group, São Cristóvão, Brazil
Email: icarovb@academico.ufs.br, {efreire, ecarvalho, lmolina}@ufs.br

***Abstract*—This paper presents a between-subjects elicitation study that aims to analyze the effect of culture on gesture-based human-robot interaction. In this study, participants from Brazil and other six countries are asked to perform gestures for controlling a mobile robot according to eight given tasks. The movements proposed are recorded and classified by the researchers, who statically assess the agreement level between the two groups. The results show that the culture does not influence the type of gesture used, but it may have an effect on the preferred gesture when the task has a cultural core. The study also highlights that the little experience of the public with robots, regardless of its cultural background, may hinder the interaction when it is required abstract commands.**

***Index Terms*—human-machine interaction, gestures, culture, social robotics, elicitation study**

I. Introduction

Bill Gates [1], in an article published in Scientific American Magazine, describes the birth scenario of a new industry. Based on cutting-edge technology, this industry is driven by large companies that sell highly specialized equipment, as well as by start-ups that target specific public, offering games and innovative gadgets. Nonetheless, the projects complexity, the slow progress, and the practical applicability generate uncertainty about the future and popularization of this industry. According to the Microsoft co-founder, this could be the description of the start of computer industry in the 1970s. However, Gates reveals that, this time, this is the picture of the robotics industry.

In fact, the challenge of popularizing robots today is similar to that faced by computers companies in the past. A decisive step towards the success of computers was the intense research for standardized and simplified interfaces for the general public. With this fact in mind, the number of scientists engaged in the development of efficient and intuitive interfaces for Human-Robot Interaction (HRI) has rapidly emerged [2]. Darling [3] further emphasizes that humans tend to interact more enthusiastically with objects that arouse affection, imagining them as if they were live beings. This fact has led to the development of robots with biological characteristics, such as the use of spoken and/or gestural language for communication. Thus, it becomes fundamental for robotics to develop interfaces with voice and gestures, surpassing the conventional ones, as mouse, keyboard, and joystick [2].

Although voice recognition finds a variety of applications, its use in HRI is limited, especially due to the influence of noise, language diversity, and distance from the microphone. In this matter, the recognition of gestures is able to overcome these barriers [4], which makes it a promising modality. Hence, several works related to human-robot interfaces through gestures have been proposed [5,6,7,8,9]. However, Burke and Lasenby [10] claim that the majority of traditional techniques tend to involve the use of pre-established gestures set, requiring the user to learn a specific gestural code for communication with the robot. This prior training implies impairment in the acceptability of the tool, since it makes interaction less fluid and it does not recognize the most relevant dimension of gestural communication: intuitiveness.

In order to overcome this problem, an important contribution was brought by [9]. The authors proposed a methodology, called elicitation studies, in which users provide suggestions for effecting commands in a system. In this way, the designer can create a more intuitive and natural interface because it is user-centric. In a later paper, the authors showed that, in an experiment conducted with 20 non-technical American participants, not even three interface specialists were able to generate a set of gestures that covered more than 60% of those suggested by the volunteers [11]. Since then, elicitation studies have been widely used in interaction research through gestures in different applications and conditions [12].

In robotics, Obaid *et al*. [13] pioneered this participatory design methodology in HRI systems. The authors invited technical and non-technical participants (all German) to perform gestures to control a humanoid robot. The results showed that technical volunteers had similar suggestions among themselves, but the gestures differed in relation to those performed by non-technical. Wobbrock, Morris and Wilson [11] also recognized that the results of their experiment with 20 non-technical American would be different if the participants were children, eastern or with lower educational level. This variation of gestures among distinct sample groups has already been observed by linguists. Kita [14], for example, reviewed studies that analyze cross-cultural differences in gestures and identified factors for such diversity. In psycholinguistic, the effect of cultural variations on the way of expressing oneself with gestures was also found in [15] and [16].

Manuscript received May 25, 2018; revised July 4, 2019.

Hence, this study aims to analyze the influence of culture on human-robot interaction through gestures. In order to assess the significance of such effect in HRI, a between-subjects elicitation study is conducted with two groups: natives (Brazilian) and foreigners. The participants are asked to freely use gestures to command a mobile robot, making it perform tasks such as *play music*, *clean the place*, *ask for help*. The robot always responds positively to users' gestures, because it role is to act as a distractor while a camera records the movements that the participants execute.

From the experimental results and the statistics proposed by [12], the researchers discuss what gestures are made for the tasks and which of these gestures vary by culture. Thus, this work contributes by offering base for studies, within similar application domain, that seek to extrapolate their findings to other populations. It also opens the path for works on user-centered interfaces that contemplate the pluralism of gestures across cultures.

The remainder of this paper is organized as follows: in section 2, it is presented a brief review of human gestures and the role of culture in gesture-based interactions. Sample profile, required materials, and experimental procedure are detailed in section 3. The corresponding results are discussed in section 4, and the study conclusion and future research directions are presented in section 5.

II. Related Work

A. Human Gestures

According to [17], the word *gesture* is derived from the Latin term *gestus*, past participle of *gerere* (to bear, carry on, perform). Therefore, *gesture* started to mean movement, attitude, beckon. Kendon [18] describes gestures as movement body phrases that convey a desire of communicative action. He states that gesture is an independent way of expression and it works in cooperation with the speech system.

McNeill [19] highlights the role of gesture in communication. Unlike speech, whose meaning is split into semantic parts, gestures can package the whole idea in one unit. Moreover, gestures have a high sensibility to distinguish information in the discourse context (e.g., inside *versus* outside) and their use is not restricted by grammatical rules.

In order to understand the gestural interaction and how gestures can be categorized, several researchers have conducted studies in human discourse. Although there is no universal standard classification for body gestures [13], Efron [15] was one of the first to build a taxonomy for gestures, classifying them into five categories: *physiographics*, *kinetographics*, *ideographics*, *deictics*, and *batons*. His taxonomy has led to different classifications, such as McNeill's: *iconic*, *metaphorics*, *deictics*, and *beats* [20], and Kendon's continuum: *gesticulation*, *language-like gestures*, *pantomimes*, *emblems* and *sign languages* [21].

Because most gesture studies are based on human discourse, the classical categories have limited applicability to human-robot interaction. Hence, the taxonomy used in this paper is based on the work of [13], who designed it for HRI contexts. The chosen taxonomy is described in Table I.

TABLE I. Taxonomy of Gestures for Human-robot Interaction

Taxonomy		Description
Form	Static gesture	A static body gesture is held after a preparation phase
	Dynamic gesture	The gesture contains movement of one or more body parts during the stroke phase
Body Parts	One hand	The gesture is performed with one hand
	Two hands	The gesture is performed with two hands
	Full body	The gesture is performed with at least one other body part than the hands
Nature	Deictic	The gesture is indicating a position or direction
	Iconic	The gesture visually depicts an icon
	Miming	The used gesture is equal to the meant action

B. Gesture-Based Human-Robot Interaction

Once researchers realized the nonverbal communication possibilities of gestures, the concept of gesture-based interaction started to receive attention among human-machine interface designers [19]. The definition of gesture-based interface used in this paper is as stated in [22], i.e., a user-centered way of interaction, in which the person performs body movements to interact and communicate with a digital system without touching a display.

An interactive system that perceives and responds to user gestures can provide a natural communication and it can be useful in situations when touch or speech are not possible or convenient [23]. Another reason for the popularization of this kind of interface is the decreasing prices of platforms such as Kinect and Leap Motion. In addition, touchless interfaces reduce injury and contamination risks [22], making them particularly useful in areas such as operating rooms [24] and industrial production lines [25].

All these advantages have motivated researchers in human-robot interaction to look for interfaces that enable communication through gestures [13]. Bodiroža, Stern and Eden [26] developed a set of gestures for users to interact with a robot waiter. In [5], a robot was programmed with gesture recognition to clean up an office. Burke and Lasenby [10] presented a pantomimic gesture interface for an unmanned aerial vehicle (UAV). In [8], the Kinect sensor was employed to recognize gestures for controlling a service robot.

Even though several studies have addressed gesture-based HRI, most of them focus on technical aspects, i.e., on sensing and recognition methods, and neglect users' cultural particularities. The insufficient progress in this perspective may hinder the engagement of the public with such interfaces [22].

C. Cultural Influence on Gesture-Based Interaction

According to [27], gesture-based interface designers need to think of cultural forms, which requires the analysis of social conventions and constructions associated with gestures. A wrong set of gestures may cause embarrassing and/or disruptive situations within a specific cultural context. Rico and Brewster [28] state that cultural aspects should not be ignored in interface projects, because they are decisive for motivating the user to interact through gestures, especially in public spaces. Decisions on the design of HRI frameworks must balance technical feasibility and psycho-social aspects [25], which may vary from one population to another.

Aware of these implications, some researchers have considered the sociocultural dimension in their studies. Vatavu and Wobbrock [12] analyzed the effect of gender on the elicitation study with the *Métamorphe keyboard* proposed by [29]. The results showed that women and men reach consensus over gestures in different ways, revealing a link between the nature of the gesture and the cognitive perception of each gender. Obaid *et al*. [13] perceived differences between technical and non-technical participants, when they were asked to control a humanoid robot.

Several studies in psycholinguistic have addressed the relationship between gestures and culture (considering *culture* from a geographic perspective). Efron [15] suggests that there is a consistency of gestures within a culture, but a variation from one to another. Calbris [30] also observes this pattern, and [16] claims that what is expressed through gestures and how they are coordinated with speech varies among speakers from different languages.

In human-machine interface studies [9,11,32], the effect of culture is considered worthy of investigation. There is, nonetheless, little research on how gestures in HRI vary by culture. Hence, this study addresses this issue, providing intra and intercultural analysis.

III. Method

The focus of this study is to analyze whether persons of different cultures and linguistic structures use different gestures in the interaction with a mobile robot, such as the Pionner 3-DX (Adept Technology). The methodology used in the experiment is based on the participatory design technique, proposed by [9], and on the classification of gestures shown in Table I. In this method, a camera records volunteers who are asked to suggest gestures (*proposals*) to command the robot, making it execute apparently random tasks (*referents*). After this stage, the researchers analyze the videos and discuss which kind of gesture was proposed for each referent.

Eight referents were selected from the ones that often appear in HRI papers and in service robot applications, namely:

1. *Establish contact;*
2. *Stop;*
3. *Task done;*
4. *Redo the task;*
5. *Play music;*
6. *Clean up the place;*
7. *Ask for help;*
8. *Follow me.*

Although the Pionner 3-DX is not capable of performing tasks as *play music* or *clean up the place*, such functions were chosen due to their practical applicability in an HRI context. The iRobot's Roomba (a vacuum cleaning robot), for instance, is one of the service robots most widely commercialized and has been used in several studies focusing on user needs [32].

A. Participants

This experiment involved the participation of 25 volunteers, arranged in two groups based on the person's country of birth. Those who were born in Brazil were classified as *natives* (13 participants); and the others, as *foreigners* (12 participants). The natives (9 females and 4 males) have an average age of 23 ($SD = 4.68$). The foreigners (8 females and 4 males) have an average age of 21 ($SD = 3.75$), and they are from USA (5), Europe (4), Taiwan (2), and East Timor (1). For this group, it was required a basic English reading skill. Overall, the participants are majority right-handed (23/25) and from several fields of study, such as Engineering, Health, Political Science, and Arts. Since each participant gives proposals for the eight referents, this study analyzes a dataset with 200 gestures.

B. Materials

The experiment is arranged in an area of 3 m wide by 4.5 m deep, as depicted in Fig. 1. The room is equipped with a Pionner 3-DX robot, a 32" monitor, a standard notebook computer, a wireless joystick and a Microsoft Kinect 2.0 sensor. The volunteer is positioned in front of the robot, being able to move freely in the designated *user area*. At the beginning of each experiment, the robot is placed at the center of the *driving area* and it is programmed to move towards any of the eight vertices of the octagon, at a maximum speed of 5 cm/s. Each vertex has a number related to the list of referents; e.g., the robot moves to the coordinates of the vertex 2 when the user executes a gesture for the referent *stop*.

The monitor is used to show the user which task the robot is supposed to "perform", and the Kinect records the gestures performed by the participants. During the experiment, the instructor holds the joystick and warns the participant that it is used to stop the robot for safety reasons. However, the buttons of the joystick (see Fig. 2), when conveniently pressed by the instructor, send wirelessly to the robot the coordinates of the octagon.

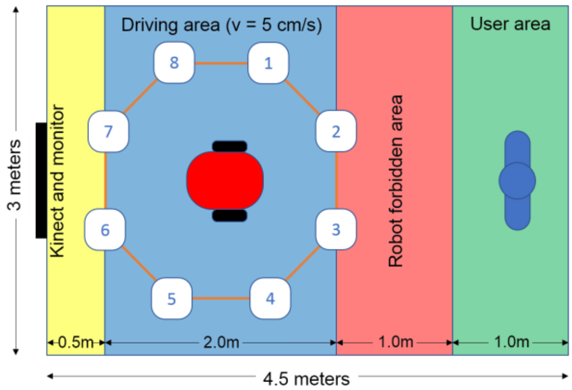

Figure 1. Experiment setup.

Figure 2. Command list and wireless joystick for teleoperation.

This technique, known as "Wizard of Oz" [33], is used in similar studies and it indents to distract the volunteers and to make them believe that their proposals work, performing the gestures as natural as possible. Therefore, the system has no recognition program, but a navigation script that responds to the joystick's buttons. None of the participants reported that they realized the manipulation; however, if such a case had happened, the sample would have been withdrawn.

C. Procedure

At the beginning of the experiment, the participant is invited to stay within the *user area* and to randomly select, from a given urn, one out of eight numbered cards. The participant is not given the list of referents, so he/she needs to show the instructor the card picked. The task corresponded to the card is, then, described on the screen with a sentence in English (for foreigners) or in Portuguese (for natives). Originally, the sentences were translated from Portuguese into English and they were revised by two English native speakers who are fluent in Portuguese. The sentences correspond to those on the list of referents, with the vocative *Pioneer* added, e.g., "Pioneer, ask for help."

As soon as the sentence is popped up on the monitor, the user is allowed to try any gesture to interact with the robot and to make it recognize the command. Regardless of the gesture performed, Pioneer executes the respective action, i.e., it goes to the corresponding coordinates by the instructor teleoperation. After each task, the participant should repeat the procedure for the remaining 7 commands.

From the recordings made by the Kinect, the researchers manually make the taxonomic classification (based on Table I) and number the gestures based on the similarities of the movements (two similar gestures receive the same number). Since this type of classification has a potential subjective bias, this stage is conducted independently by two researchers, who, ultimately, discuss the final classification.

IV. Results

This between-subjects study presents results that contemplate taxonomy, agreement measures, and subjective observations for each cultural group (native *versus* foreigners).

A. Gesture Taxonomy

All gestures are classified according to the three dimensions from Table I: *form*, *body parts*, and *nature*. *Form* distinguishes between static and dynamic gestures. Although a static gesture requires an initial and a retraction movement, its core is static, i.e., there is a significant amount of time in which the user holds a posture (in comparison with a dynamic gesture) [13]. The *body parts* dimension refers to the number of hands used by the participants (one or two hands), or if they used any other body part aside hands (full body). Finally, the *nature* of the gesture can have three categories: *deictic, iconic,* and *miming*. Deictic gestures indicate position or direction, e.g., pointing to the left, tilting the head. Iconic gestures are visual depictions, e.g., an open facing hand for *stop*, or holding the thumbs up for *task done*. The user may also execute miming gestures to guide the robot to reproduce the same movements/idea, e.g., pretending to play a guitar to make the robot *play music*.

Fig. 3 depicts the taxonomy distributions obtained for native and foreigner participants. It can be seen that there is a consistency across the groups for each dimension. Both groups preferred dynamic rather than static gestures. The expectation of responsiveness of the robot may have induced participants to think of kinetic proposals. Natives used more miming than foreigners, but both preferred iconic gestures in general. Since most referents are abstract (there is no pointing or navigating function), deictic gestures were rarely employed. Participants also rarely used the full body, and the choice between one or two hands was balanced. Although Fig. 3 shows slight variations across the groups, no statistically significant difference was found.

The taxonomic classification shows that culture does not influence the type of gesture used, which is important when it comes to the techniques applied to recognize gestures. The result highlights that, independently of culture, the sensing systems should be able to deal with dynamic, iconic and miming gestures, but tracking the full body may not be important.

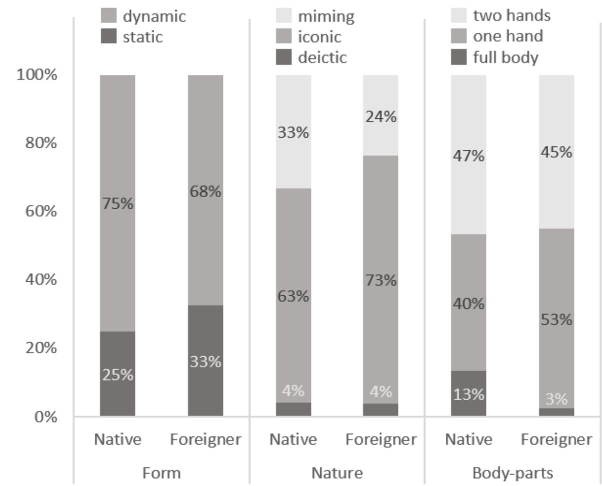

Figure 3. Taxonomy distribution for *natives* and *foreigners*.

B. Agreement Analysis

Vatavu and Wobbrock [34] proposed an agreement measure that is widely adopted in elicitation studies. The agreement rate (*AR*) calculates the probability of two participants choosing the same proposal for a given referent. Therefore, *AR* is a value in the range [0, 1] that defines the general agreement among users. The authors present the following equation to calculate the agreement score for a referent *r*:

$$AR(r,G_i)=\frac{a_i}{n_i}=\frac{\sum_{p=1}^{|G_i|}\sum_{q=p+1}^{|G_i|}\delta_{p,q}(r)}{\frac{1}{2}|G_i|\left(|G_i|-1\right)}, \quad (1)$$

in which a_i denotes the number of pairs in agreement in group G_i, and n_i is the total number of pairs of that group. The Kronecker's $\delta_{p,q}$ notation [35] is either 1 or 0, depending whether participants *p* and *q* are in agreement, i.e., if they chose the same gesture for task *r*. The notation $|G_i|$ represents the cardinality of set G_i.

For instance, let's assume that a referent *r* received proposals from two groups, $|G_1| = 7$ and $|G_2| = 5$. The 12 participants presented 4 different proposals, {♠, ♣, ♦, ♥} – these symbols might represent the gestures used in our study. If the set of proposals given by the participants from G_1 is {♠, ♥, ♠, ♣, ♣, ♥, ♣}, and from G_2 is {♦, ♦, ♠, ♠, ♦}, we have $a_1 = 5$ (pairs in agreement) and $a_2 = 4$; $n_1 = 21$ (possible pairs from G_1) and $n_2 = 10$. This leads to $AR_1(r) = 0.238$ and $AR_2(r) = 0.400$, which means there is 23,8% of chance that two persons from G_1 pick the same gesture, whereas this probability is 40% for G_2 group.

The difference between n_i and a_i equals the number of pairs in disagreement d_i, i.e., $d_i = n_i - a_i$. Using this notation, the data from *k* independent groups for referent *r* can be organized as a $2 \times k$ contingency table, as in Table II. This method is traditionally used to test the hypothesis of association between rows and columns. The null hypothesis, H_0, states that the proportions of a_i in relation to n_i are independent of the sets G_i, i.e., all groups have equal agreement rates. In the present study, rejecting H_0, for example, means that culture has a significant effect on gesture choices for a given *r*.

TABLE II. MODEL OF CONTINGENCY TABLE FOR AGREEMENT ANALYSIS

Agreement (*r*)	Groups				Total
	G_1	G_2	…	G_k	
Yes	a_1	a_2	…	a_k	***a***
No	d_1	d_2	…	d_k	***d***
Total	n_1	n_2	…	n_k	***n***

Although null hypotheses of contingency tables are usually assessed by techniques such as Pearson's Chi-Square [36] test or Fisher's exact test [37], [12] proposed a new approach focused on between-subject elicitation studies. Inspired by the principles of Fisher's exact test, the authors developed the V_b statistical test. Similarly to the χ^2 statistic, the value of V_b reflects the difference between observed (a_i) and expected (ε_i) agreement configurations, and it is given by:

$$V_b(a_1,a_2,\ldots,a_k)=\sum_{i=1}^{k}\left(a_i-\varepsilon_i\right)^2, \quad \varepsilon_i=\frac{n_i\sum_{j=1}^{k}a_j}{n}. \quad (2)$$

The expected number of pairs, ε_i, considers the total number of pairs, *n*, and the null hypothesis (same proportions for all groups). From the previous example with groups G_1 and G_2, $a_1 = 5$ and $a_2 = 4$ give $\varepsilon_1 = 6.10$ and $\varepsilon_2 = 2.90$, which leads to $V_b = 2.406$. On the other hand, a more extreme configuration for G_1, e.g., {♠, ♠, ♠, ♠, ♠, ♠, ♥}, would lead to $a_1 = 15$ ($AR_1(r) = 0.714$), $a_2 = 4$ ($AR_2(r) = 0.400$), $\varepsilon_1 = 12.87$, $\varepsilon_2 = 6.13$, and $V_b = 9.066$. Hence, the larger V_b is, the greater is the probability of having a variable (such as culture) that has a significant effect on the respective *AR*.

The authors provide a software tool [12] to run the V_b test, which also includes the computation of $\Pi a_1,a_2,..,a_k|n_1,n_2,..,n_k$, the cumulative probability of observing the set of proposals a_1, a_2, …, a_k or more extreme configurations (those with larger V_b). The null hypothesis is rejected if Π is smaller than a *p* level (usually, $p = 0.05$ or lower). The V_b statistic is reported as $V_{b(k,N=total)} = V_b$, $p = \Pi$, in which *k* is the degree of freedom (in this case, the number of groups), and *N* is the total number of participants. From the first example, for referent *r*, the V_b test shows no significant difference between AR_1 and AR_2 (0.238 *versus* 0.400, $V_{b(2,N=12)} = 2.406$, $p = 0.373$).

One can notice that, athough AR_1 and AR_2 are not statistically different, there is a clear disagreement *between* G_1 and G_2 in relation to the proposals. To capture the actual preference of independent groups, [12] also proposed an additional measure, called between-groups coagreement rate (CR_b). The CR_b evaluates how much consensus is shared *between* independent groups and it is given by:

$$CR_b(G_1,G_2,\ldots,G_k)=\frac{\sum_{i=1}^{k}\sum_{j=i+1}^{k}\sum_{p=1}^{|G_i|}\sum_{q=1}^{|G_j|}\delta_{p,q}}{\sum_{i=1}^{k}\sum_{j=i+1}^{k}|G_i|\cdot|G_j|}, \quad (3)$$

in which $\delta_{p,q}$ is the Kronecker's notation from (1). The sum goes for all pairs of participants selected from all pairs of groups G_i and G_j ($1 \le i < j \le k$). From the first example, whose proposals was {♠, ♥, ♠, ♣, ♣, ♥, ♣} for G_1, and {♦, ♦, ♠, ♠, ♦} for G_2, only 4 out of the 35 possible pairs across the two groups are in coagreement, i.e., $CR_b(G_1,G_2) = 0.114$. Therefore, high *ARs* and low CR_b means consensus within groups but disagreement between them [12].

All this statistical background provides the tools to analyze the present elicitation study between natives (G_1) and foreigners (G_2), with $|G_1| = 13$ and $|G_2| = 12$. To better understand the results, Fig. 4 depicts the gestures more often chosen per group for each referent, while in Fig. 5 it can be seen the found agreement rates, as well as the CR_b, with the referents in ascending order of the exact *p* value of the V_b test.

From the statistical analysis, it was found significant difference between natives and foreigners for the function *play music* (0.438 *versus* 0.156, $V_{b(2,N=25)} = 158.42$, $p = 0.013$). Natives reached a higher agreement because

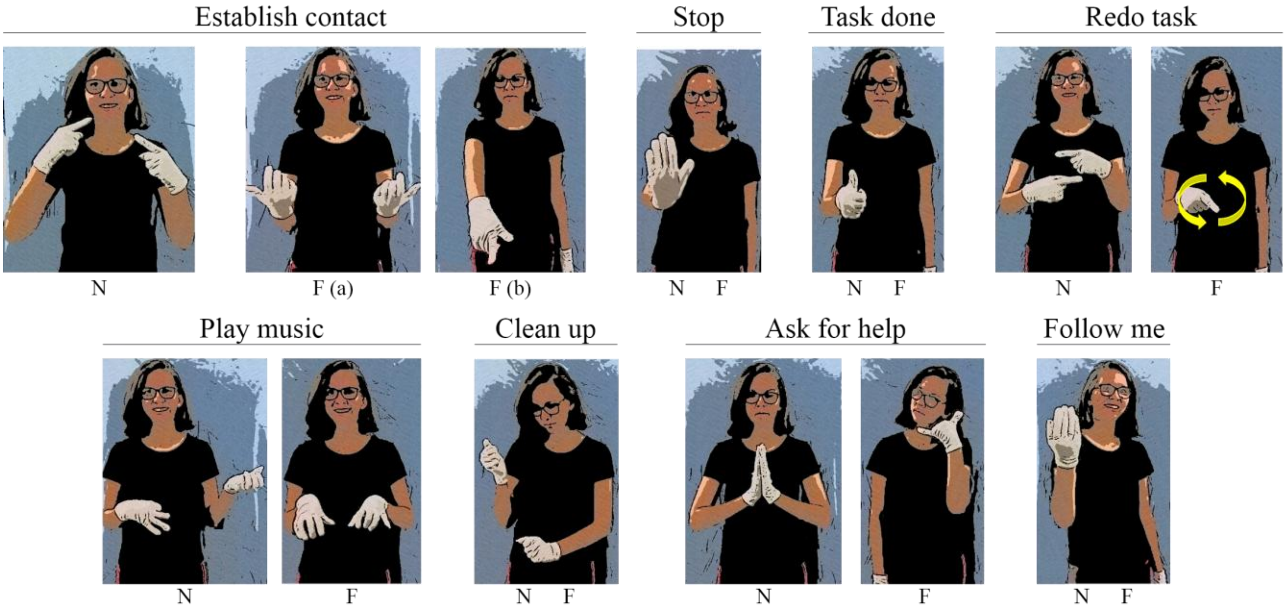

Figure 4. Most frequently gestures used by native (N) and foreigner (F) participants to interact with Pioneer. For *establish contact*, there was a tie of two proposals, F(a) and F(b), among foreigners.

most of them executed a miming gesture of playing guitar. Foreigners also executed miming movements, which explains similar results in the taxonomy distribution (see Fig. 3), but they varied in their choices of instrument (which went from violin to drums and piano, mostly the latter). One hypothesis for the preference of miming is that most participants have no technical background. In [13], this miming trend was found with German nontechnical users, which suggests that the nature of the gesture is not significantly associated with culture. On the other hand, the instrument choices may reflect the popularity of guitar in Brazil, whereas in other countries there should be more exposure to different sorts of instruments. Despite the cultural difference among the chosen gestures, the found instrument miming pattern could be explored in robots that play music, such as the humanoid Lynx (UBTECH Robotics), without relying only on voice recognition or touch displays.

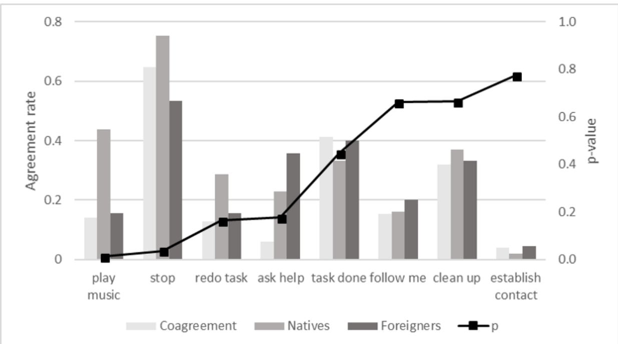

Figure 5. Agreement rates computed for *native* and *foreigner* participants.

The referent *stop* also showed significant difference (0.752 *versus* 0.533, $V_{b(2,N=25)} = 95.220$, $p = 0.037$). However, the CR_b was the highest among all referents (0.647), indicating that 64,7% of all pairs of participants across the two groups were in agreement about how to stop the robot. Hence, even though there were more agreement among natives, it is very likely that two persons choose the same gesture for *stop*, regardless of their culture. The disagreement among foreigners is, in fact, low, and the difference to natives is due to the body parts chosen: natives used one hand with higher frequency for this specific referent.

The remaining referents did not present significant differences. In fact, the proposals for *stop*, *task done* and *follow me* are classical iconic gestures. *Clean up* is as abstract as *play music*, so most participants chose miming gestures with a cleaning object, as a virtual broom (see Fig. 4), which reveals a pattern to be explored in service robots design.

Redo the task received similar proposals (hand spinning movements), although natives agreed slightly more on two-handed gestures. This result is particularly important for works such as [25], in which a gesture set was developed for HRI in industrial context. The authors defined the gestures (see Fig. 6) based on safety and environmental constraints, without any elicitation study. From our results, the gestures for *redo the task* and *task done* in Fig. 6, for example, are likely to be well accepted across other cultures. However, a special attention should be taken with *ask for help* and *establish contact*.

Ask for help showed a CR_b (0.060) much lower than the corresponding *ARs* (0.229 *versus* 0.356, $V_{b(2,N=25)} = 32.000$, $p = 0.176$). This means that, albeit the groups have regular within agreement rate, only 6% of all pairs across G_1 and G_2 would choose the same gesture. While natives simulated a drowning, or folded their hands together as they were praying (depicted in Fig. 4), foreigners chose to simulate calling an emergency number (Fig. 4), or they put their hands around the mouth, as they were screaming for help.

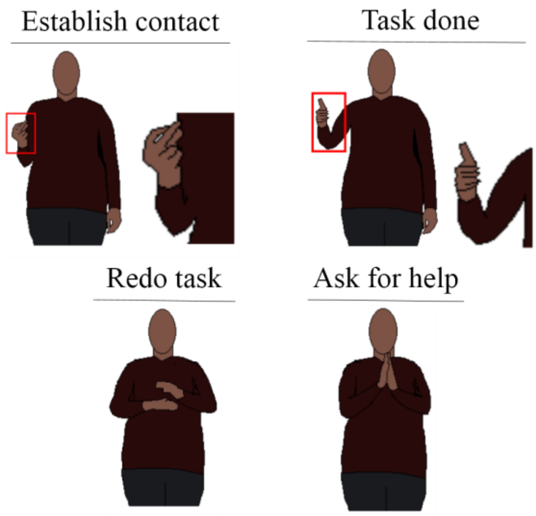

Figure 6. Gestures proposed by [25] for HRI in industrial context.

In the case of *establish contact*, it showed the lowest consensus (0.019 *versus* 0.044, $V_{b(2,N=25)} = 1.280$, $p = 0.775$). A possible explanation is that the Pionner is a navigation robot, without the features of a humanoid one, such as human-like body parts. The participants hesitated as they were not sure about how to request the robot's attention. Hence, they came up with all sorts of proposals, from bowing down to simulating a hug. Few proposals for this referent repeated, so the *establish contact* gestures in Fig. 4 do not represent a significant set.

The results for *ask for help* and *establish contact* indicate that designers should devise parametric gestures, as the ones in Fig. 6. Surprisingly, the proposal to *ask for help* in Fig. 6 agreed with the natives' choice in Fig. 4. Nonetheless, the low CR_b and the relevance of this task reinforce the need of a conventionalized sign, regardless culture. In such cases, designers may also consider other ways of interaction; e.g., participants suggested saying "SOS" for *ask for help* and to clap hands for *establish contact*.

V. Conclusion

In this paper, it was analyzed the cross-cultural effect on gesture-based human-robot interaction. A between-subjects elicitation study was conducted, allowing a comparison between natives (Brazilian) and foreigners towards their gesture preferences.

The results of this study show that culture does not affect significantly the class of gesture employed (considering the taxonomy presented), but it may influence the choice of specific gestures, particularly if the function is related to the culture itself, as in *play music*.

The referent *ask for help*, which would be an important function in many HRI contexts, should receive attention from the HRI designers due to its low coagreement. The overall low agreement for *establish contact*, a primary function in HRI, indicates that the type of robot may play an important role on intuitiveness of the gestures. In both cases, standard gestures are recommended.

The referents *stop*, *task done*, *redo the task*, *clean up* and *follow me*, whose proposals did not significantly vary with culture, revealed patterns that should be considered by interface designers. These findings reinforce the capability of gestures to overcome the barriers of language and to become an inclusive way for human-robot interaction.

Further work will include more participants, so other sociocultural aspects, as gender, age, and revenue, may be analyzed. We hope the contributions of this work will provide researchers with insights about the relevance of users' culture on interface design.

Acknowledgment

This study was financed in part by the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil* (CAPES) - Finance Code 001; and by the Foundation for Research and Technological Innovation in the State of Sergipe (FAPITEC/SE).

References

[1] B. Gates, "A robot in every home," *Sci. Am. sp*, vol. 18, no. 1, pp. 4–11, 2008.

[2] D. H. Pal and S. M. Kakade, "Dynamic hand gesture recognition using kinect sensor," in *Proc. 2016 Int. Conf. Glob. Trends Signal Process. Inf. Comput. Commun.*, pp. 448–453, 2016.

[3] K. Darling, "Children beating up robot inspires new escape maneuver system," *IEEE SPECTRUM*, Aug-2015. [Online]. Available: http://spectrum.ieee.org/automaton/robotics/artificial-intelligence/children-beating-up-robot. [Accessed: 25-Aug-2017]

[4] C. Yang, D. K. Han, and H. Ko, "Continuous hand gesture recognition based on trajectory shape information," *Pattern Recognit. Lett.*, vol. 0, pp. 1–9, 2017.

[5] S. Waldherr, R. Romero, and S. Thrun, "Gesture based interface for human-robot interaction," *Auton. Robots*, vol. 9, no. 2, pp. 151–173, 2000.

[6] K. Nickel and R. Stiefelhagen, "Visual recognition of pointing gestures for human-robot interaction," *Image Vis. Comput.*, vol. 25, no. 12, pp. 1875–1884, 2007.

[7] A. Butalia, D. Shah, and R. V Dharaskar, "Gesture recognition system," *Int. J. Comput. Appl.*, vol. 1, no. 5, pp. 61–67, 2010, .

[8] T. Fujii, J. H. Lee, and S. Okamoto, "Gesture recognition system for human-robot interaction and its application to robotic service task," in *Proc. Int. MultiConference Eng. Comput. Sci. 2014 Vol I, IMECS 2014*, vol. I, pp. 1–6, 2014.

[9] J. Wobbrock *et al.*, "Maximizing the guessability of symbolic input," in *CHI '05 extended abstracts on Human factors in computing systems*, 2005, no. January, pp. 1869–1872.

[10] M. Burke and J. Lasenby, "Pantomimic gestures for human-robot interaction," *IEEE Trans. Robot.*, vol. 31, no. 5, pp. 1225–1237, 2015.

[11] J. O. Wobbrock, M. R. Morris, and A. D. Wilson, "User-defined gestures for surface computing," in *CHI '09 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2009, pp. 1083–1092.

[12] R. D. Vatavu and J. O. Wobbrock, "Between-subjects elicitation studies: Formalization and tool support," in *Proc. 2016 CHI Conf. Hum. Factors Comput. Syst. - CHI '16*, pp. 3390–3402, 2016.

[13] M. Obaid, F. Kistler, M. Häring, R. Bühling, and E. André, "A framework for user-defined body gestures to control a humanoid robot," *Int. J. Soc. Robot.*, vol. 6, no. 3, pp. 383–396, 2014.

[14] S. Kita, "Cross-cultural variation of speech-accompanying gesture: A review," *Lang. Cogn. Process.*, vol. 24, no. 2, pp. 145–167, 2009.

[15] D. Efron, *Gesture and Environment*, Oxford, England: King'S Crown Press, 1941.

[16] A. Özyürek, "Cross linguistic variation in children's multimodal utterances," in *Sources of Variation in First Language Acquisition: Languages, contexts, and learners*, no. March, M. Hickmann, E. Veneziano, and H. Jisa, Eds. Amsterdã John Benjamins, 2018, pp.

123–137.
[17] A. B. Carmo, “Consultório Etimológico.” [Online]. Available: http://origemdapalavra.com.br/artigo/gesto/. [Accessed: 20-Aug-2017].
[18] A. Kendon, “An agenda for gesture studies,” *Semiot. Rev. Books*, vol. 7, no. 3, pp. 1–22, 1996.
[19] D. McNeill, *Hand and Mind: What gestures reveal about thought*. University of Chicago press, 1992.
[20] D. McNeill, “So you think gestures are nonverbal?,” *Psychol. Rev.*, vol. 92, no. 3, p. 350, 1985.
[21] M. Studdert-Kennedy, “A review of McNeill, D. (1992). hand and mind: What gestures reveal about thought,” *Haskins Lab. Status Rep. Speech Res.*, no. 1992, pp. 149–153, 1993.
[22] M. Lou Maher and L. Lee, “Designing for gesture and tangible interaction,” *Synth. Lect. Human-Centered Informatics*, vol. 10, no. 2, p. i-111, 2017.
[23] H. Jetter, “A cognitive perspective on gestures, manipulations, and space in future multi-device interaction,” no. May, 2014.
[24] K. Jurewicz, D. M. Neyens, K. Catchpole, and S. T. Reeves, “Developing a 3D gestural interface for anesthesia-related human-computer interaction tasks using both experts and novices,” *Hum. Factors*, 2018.
[25] P. Barattini, C. Morand, and N. M. Robertson, “A proposed gesture set for the control of industrial collaborative robots,” in *Proc. IEEE Int. Work. Robot Hum. Interact. Commun.*, pp. 132–137, 2012.
[26] S. Bodiroža, H. I. Stern, and Y. Edan, “Dynamic gesture vocabulary design for intuitive human-robot dialog,” *Proc. seventh Annu. ACM/IEEE Int. Conf. Human-Robot Interact - HRI’12*, p. 111, 2012.
[27] M. S. Horn, “The role of cultural forms in tangible interaction design,” pp. 117–124, 2013.
[28] J. Rico, A. Crossan, and S. Brewster, “Gesture-based interfaces: practical applications of gestures in real world mobile settings BT - whole body interaction,” D. England, Ed. London: Springer London, 2011, pp. 173–186.
[29] G. Bailly, T. Pietrzak, J. Deber, and D. J. Wigdor, “MéTamorphe: Augmenting hotkey usage with actuated keys,” in *Proc. the SIGCHI Conference on Human Factors in Computing Systems*, 2013, pp. 563–572.
[30] G. Calbris, “From left to right: Coverbal gestures and their symbolic use of space,” *Metaphor gesture*, vol. 3, pp. 27–53, 2008.
[31] D. Mauney, J. Howarth, A. Wirtanen, and M. Capra, “Cultural similarities and differences in user-defined gestures for touchscreen user interfaces,” in *Proc. 28th Int. Conf. Ext. Abstr. Hum. factors Comput. Syst. - CHI EA ’10*, p. 4015, 2010.
[32] J. Fink, V. Bauwens, F. Kaplan, and P. Dillenbourg, “Living with a vacuum cleaning robot: A 6-month ethnographic study,” *Int. J. Soc. Robot.*, vol. 5, no. 3, pp. 389–408, 2013.
[33] N. Dahlbäck, A. Jönsson, and L. Ahrenberg, “Wizard of Oz studies - why and how,” *Knowledge-Based Syst.*, vol. 6, no. 4, pp. 258–266, 1993.
[34] R. D. Vatavu and J. O. Wobbrock, “Formalizing Agreement Analysis for Elicitation Studies,” in *Proc. 33rd Annu. ACM Conf. Hum. Factors Comput. Syst. - CHI ’15*, pp. 1325–1334, 2015.
[35] R. C. James, *Mathematics Dictionary*, Springer Science & Business Media, 1992.
[36] M. L. McHugh, “The chi-square test of independence,” *Biochem. Medica Biochem. Medica*, vol. 23, no. 2, pp. 143–149, 2013.
[37] R. A. Fisher, “Statistical methods for research workers,” in *Breakthroughs in statistics*, Springer, 1992, pp. 66–70.

Icaro V. Brito received the title of Electrical Engineer (2016) from the Federal University of Sergipe (UFS). He is currently a Master student at the Department of Electrical Engineering of the Federal University of Sergipe (UFS). His main research interests are: robotics, human-robot interaction, and social robots.

Eduardo O. Freire received the title of Ph.D. in Electrical Engineering (2002) from the Federal University of Espírito Santo (UFES). He is currently an Associate Professor at the Department of Electrical Engineering of the Federal University of Sergipe (UFS). His main research interests are: robotics, control, and artificial intelligence.

Elyson A. N. Carvalho received the title of Ph.D. in Electrical Engineering (2012) from the Federal University of Campina Grande (UFCG). He is currently an Associate Professor at the Department of Electrical Engineering of the Federal University of Sergipe (UFS). His main research interests are: robotics, nonlinear systems, instrumentation, and signal processing.

Lucas Molina received the title Ph.D. in Electrical Engineering (2014) from the Federal University of Campina Grande (UFCG). He is currently an Assistant Professor at the Department of Electrical Engineering of the Federal University of Sergipe (UFS). His main research interests are: robotic manipulators, mobile robotics, and control.