Intuitiveness Level: Frustration-Based Methodology for Human–Robot Interaction Gesture Elicitation

CLEBESON CANUTO1, EDUARDO O. FREIRE2, LUCAS MOLINA2, ELYSON A. N. CARVALHO2, (Member, IEEE), AND SIDNEY N. GIVIGI3, (Senior Member, IEEE)

1Department of Electrical Engineering, Federal University of Espírito Santo, Goiabeiras, Vitoria, Espírito Santo 29075-910, Brazil
2Department of Electrical Engineering, Federal University of Sergipe, São Cristóvão, Sergipe 49100-000, Brazil
3School of Computing, Queen’s University, Kingston, ON K7L 3N6, Canada
Corresponding author: Clebeson Canuto (clebeson.canuto@gmail.com)

This work was supported in part by the Natural Sciences and Engineering Council of Canada under the Discovery Grant Program.

ABSTRACT For robotics to become more accessible to people not specialized in the area, it is of fundamental importance to improve and simplify the way people interact with robots. Despite human-robot interaction (HRI) being an effervescent research area, most of the works published so far on the use of gesture interfaces for human-robot communication do not clearly describe how the used gestures were elicited, thus hindering the reproducibility of those works. Considering this, we propose a new and reproducible Frustration-Based Approach (FBA), scientifically established on previous research, which can be used to obtain an intuitive and robust gesture vocabulary for HRI. To accomplish this, we propose Intuitiveness Level (IL), a score to rank gestures according with its intuitiveness. Using IL, it is possible to conceive a complex vocabulary, allowing an increasing of robustness, since more than one gesture can be associated to a task. In a general sense, the proposed methodology is not limited only for HRI, and it can also be used for human-machine interaction in general. In short, the contributions of this work are: (i) A complete methodology to elicit gestures to be used as intuitive communication interface between humans and robots. (ii) A metric of intuitiveness which takes into account at least three different characteristics about the elicited gestures.

INDEX TERMS Gesture elicitation, human–robot interaction, gesture vocabulary, intuitive gestures, psycholinguistics.

I. INTRODUCTION
Future ubiquity of robotics depends on the improvement of human-robot interfaces. As stated by Wexelblat [1], to be intuitive, the communicative interface must be as natural as possible, and this can be attained through speech and gestures, which, according to McNeill [2], are co-expressive, in other words, both come from the same shared semantic source, but they are not redundant.

More than 60% of communication between people is non-verbal (use of gestures) [3] and according to [2], gestures have more significance than speech, since they do not possess grammatical rules which limit their use. Therefore, it is possible that interfaces based only on gestures can be as intuitive as those based on speech. The study of gestures is formally part of a larger research field known as psycholinguistics [2].

Several works on Human-Robot Interaction (HRI) are concerned with the use of gestures as a communicative interface between humans and robots [4]–[8]. However, most of these works were concerned only about demonstrating that gestures offer a good alternative for intuitive interaction, but fail in
describing which gestures were used and the methodology adopted to elicit them. It is of crucial importance that works on HRI based on gestures be more complete in describing the tasks performed by the robot, the methodology adopted for choosing the vocabulary of gestures and the gestures chosen to represent each task. The present paper is a contribution towards a solution for this gap.

According to Powers [9], the Human-Computer Interaction (HCI) research area has strictly defined rules and concepts that are accepted by the academic community and could provide solutions to some of the unresolved issues in HRI. So, as did Wachs [10], we reviewed studies on HCI to find a suitable methodology to obtain gesture vocabularies. Despite finding some methodologies, none of them were properly based on psycholinguistic theory, jeopardizing the intuitiveness of the chosen vocabulary.

This paper focuses on the study of gestures as an intuitive and communicative interface for robots, based on psycholinguistic concepts. Following this study, a scientific basis for the question of the intuitiveness present in gestures will be presented. Therefore, the contributions of this work are: (i) A complete methodology to elicit gestures to be used as intuitive communication interface between humans and robots. (ii) A metric of intuitiveness which takes into account at least three different characteristics about the elicited gestures.

The remainder of the paper is organized as follows. Section II discusses correlated works with the aim of better detailing and explaining our proposal. Section III details our proposal of a new and reproducible methodology, scientifically established on previous research, which can be used to obtain an intuitive and robust gesture vocabulary for HRI. To accomplish this, we introduce the Intuitiveness Level (IL), a score to rank gestures according with their intuitiveness. In Section IV we show how the IL can be used to conceive a complex vocabulary, allowing for increased robustness, since more than one gesture can be associated to a task. The conclusions are presented in Section V.

II. RELATED WORKS

Since this paper proposes a new methodology to elicit intuitive gesture vocabularies for HRI, this section is organized in three subsections: a brief explanation of Psycholinguistics for Gestures, Intuitiveness of Gestures and Methodologies for Gesture Elicitation.

A. PSYCHOLINGUISTICS FOR GESTURES

Thought and language are linked to body actions, and, even though communication is performed in verbal and non-verbal ways, every human is first (as a baby or infant) in a pre-linguistic period, which is characterized by motions and gestures such as smiling, arm waving, hand gestures, and head motions, known as paralinguistic behaviours. This paralinguistic development shows different levels of complexity [11], [12].

The evolution of gesture types has a communicative function, and, therefore, several studies have demonstrated the relationship between gestures and cognitive processing, i.e., the interconnection between language, thought, and gesture [13]. Specifically, psycholinguistic studies found a correlation between linguistic abilities and hand gestures, as gestures are related to language development. Furthermore, hand gestures are important elements in organizing cognitive processes, being able to express a variety of thoughts [14].

Psycholinguistics supports the grouping of gestures into four types: iconic, metaphorical, rhythmic, and deictic, with different complexities and functions [2], [15]. In this paper, a mixture of these for types will be analyzed. This is so because, as McNeill [15] states, “we often find iconicity, metaphoricity, deixis and other features mixing in the same gesture”. As it will be discussed further in our results, rhythmic gestures will also involve pointing, which is deictic. Again, this should be seen more as dimensions [14] than categories.

Due to its link to cognitive and linguistic development, gestures have been widely used in the development of interactive interfaces. One specific area of active research is the use of gestures in robotics, both in the control of robots as well as an artifact supporting participation and enhancing communication with users. Based on psycholinguistic, the use of gestures is justified by the fact they are a natural and spontaneous occurrence in human communication and development. As such, correlating gestures to tasks is directly related to several cognitive processes used in communication, being very important in the development of technological tools for human-robot interfaces [16].

B. INTUITIVENESS OF GESTURES

According to Davidson [17], the word intuition originates from Latin, intuitionem, and has the following etymology: seeing through the eye, visual perception. Thus, intuition may be defined as an immediate perception of an external object as soon as it is seen, without the need for any previous reasoning to analyze it. According to Wachs [10], gestures intuitiveness can be defined as “the cognitive naturalness of associating a gesture with a command or intent”. This means that an intuitive gesture can be understood as a gesture that can be perceived and interpreted immediately, without the need of any inference of reasoning.

According to Wexelblat [1], to be intuitive, a gesture must be performed as naturally as possible. McNeill [2] explains that natural/spontaneous gestures have a core of meaning larger than gestures performed with restrictions, such as in the case of sign language, for example. In this way, a vocabulary of intuitive gestures cannot be restrictive, and should be conceived in a manner that maximizes their immediate execution by the users.

C. METHODOLOGIES FOR GESTURE ELICITATION

The coexistence and interaction between humans and robots is considered one of the most important questions in robotics [18]. In this subsection we present some works that
approach this question using gestures as a communication interface.

HRI can be seen as a subarea of Human Machine Interaction (HMI) and Human Computer Interaction (HCI), wherein methodologies used to elicit gestures or other symbolic vocabularies to be used in interacting with machines have already been proposed. For example, a formal metric for the guessability and agreement of a vocabulary of symbols was proposed in [19]. Despite a great contribution to the area of eliciting vocabularies of symbols (gestures can be seen as symbols related to activities or commands), both proposed metrics are based only on the frequencies of each symbol. In addition, nothing is said about the elicitation process itself, which is of primary importance for the quality of the resulted vocabulary, since the way in which instructions are passed during the elicitation process can influence the set of elicited symbols. An interesting statement of this work is that end-users, not experts, should participate of the elicitation process for the design of interactive systems. This can improve the quality of the vocabulary, but nothing can be stated about its intuitiveness. However, many works have used guessability and agreement to elicit gestures or other types of symbols to interact with machines [20]–[29].

As stated in [30], due to legacy-bias acquired by technical people during interaction with many devices, methods that aim to increase variations of elicited symbols can be more beneficial for Gesture Elicitation Systems (GES). An example of such method is the Production Principle [31]. The principle states that requiring users to produce multiple interaction proposals for each referent (task) may force them to move beyond simple, legacy-inspired techniques to ones that require more reflection. Although this approach brings enormous gains to the variability of the vocabulary of acquired gestures, it still requires specific and non-standard instructions for each new elicitation, which, again, may influence the obtained results, especially with regard to the intuitiveness of gestures. Other interesting open challenges [31] are the minimum number of symbols the participants must be asked to perform, and how this number can impact their creativity. Note that these issues can directly influence the intuitiveness of elicited vocabulary, and deserve to be taken into account.

Since a Production Principle based approach tends to collect many candidates gestures to each task, it is necessary to have a way to differentiate those gestures through metrics based on scores and votes that facilitate their ranking, in order to simplify the decision-making about which gesture should represent each task [30].

Nielsen et al. [32] proposed a procedure based on learning rate, ergonomics and intuition for gesture selection applied to HCI. The authors emphasized the importance of using ergonomics theory for choosing the gesture vocabulary. Despite considering intuitiveness as a key issue, the paper does not mention psycholinguistics theory, raising questions on the intuitiveness of the chosen vocabulary.

Several papers have focused specifically on HRI. Waldherr et al. [4] proposed the construction of a communicative interface for HRI based on dynamic gestures, solving problems of other works that addressed the gestural interface only with static gestures. Through using dynamic gestures, the interaction between humans and robots is more natural. However, the criteria/methods used to choose the gestures were not presented, resulting in doubts over the naturalness of their execution, which according to Wexelblat [1], is of vital importance for the production of an interface based on gestures.

Pereira et al. [33] showed that it is possible for a human to interact with a robot through gestures and that this helps in the carrying out of some specific tasks. To do the experiments, the authors proposed a combination of five manual and static gestures; each gesture representing a command attached to a specific task to be carried out by the robot. One issue to be considered in the execution of this study is that, despite being based only on gestures as interface for interaction, there was no discussion of why such gestures were chosen, raising concerns about their intuitiveness, as well as how the results could be reproduced in other scenarios.

Similar to [33], many other studies [18], [34], [35] also presented the gestures utilized, but do not consider their intuitiveness, nor the methodology used to select them.

For Wexelblat [1], if intuitiveness is not taken into consideration, it does not make sense to use natural interfaces, for users will have to learn the commands to be given as parameters for the entrance of the interface. Thus, in such situations, interfaces based on manual commands, such as the ones used in a joystick or a remote control (with a much higher success rate of success and lower implementation complexity), could be used.

Another methodology was proposed by Stern et al. [36]. It consists of an analytical methodology which aims to determine an optimal vocabulary of gestures for HRI through quadratic programming based on the following characteristics: intuitiveness of gestures, accuracy of recognition, and ergonomics. Although it is a good proposal, the methodology only considers the use of static gestures and only one gesture can be assigned to each task. Regarding intuitiveness, the work is based only on Nielsen’s methodology [32] and, therefore, is not based on any aspect of psycholinguistics.

Some attempts on using psycholinguistic concepts have been attempted. For example, Wachs [10] introduced an approach to obtain an intuitive vocabulary of gestures for HRI based on psycholinguistics [2] and on HCI [32]. As previously mentioned, Wachs proposed a definition for intuitiveness and a methodology to choose the best vocabulary for its application. However, the proposed approach considered only static gestures, which sacrificed its naturalness. Another important issue is that a given task could not be represented by more than one gesture, thus reducing the robustness of the proposed vocabulary. Furthermore, despite citing psycholinguistics, the study on intuitiveness was done only based on Nielsen’s approach, which is only based on ergonomics theory.

The methodologies found in our review were not presented clearly enough to allow their exact reproduction by other researchers. Among the works reviewed, the clearest is the
one proposed by Nielsen et al. [32] for HCI, which also cannot be easily reproduced, since several steps were not completely described. So, the new methodology proposed in this paper comes to fill the gap of a fully reproducible methodology, scientifically established on previous research on psycholinguistics, to obtain intuitive gesture vocabularies for HRI.

III. THE PROPOSED METHODOLOGY
The methodology proposed to find gestures representing an intuitive vocabulary for HRI consists of four main steps: (a) selection of the tasks to be performed, (b) capture of gestural data for each task, (c) analysis of captured data, and (d) choice of gestures that better represent each task according to the Intuitiveness Level (IL), a score to rank gestures based on its intuitiveness.

A. CHOICE OF TASKS
This step is similar to the one described in the methodology presented by Nielsen et al. [32], where it is necessary to select a set of tasks to which the vocabulary of gestures should make reference. This step is of fundamental importance, since the next steps will depend directly on how many and which tasks are chosen.

B. CAPTURE OF GESTURAL DATA
According to Nielsen et al. [32] there are two different focus approaches to obtain intuitive gestures: focused on the user or on the machine. The former considers only the intuitiveness of gestures, while the latter considers characteristics of gestures that minimize recognition error. Therefore, to define the set of gestures which should compose the vocabulary in a way to maximize the effectiveness of the human-machine interaction, an user based approach should be used. This can be done in a conscious or subconscious way [32]. In the conscious way, the volunteer is asked to indicate or illustrate which gesture should represent a particular task or can choose among several gestures which best represents the regarded task. On the other hand, in the subconscious way, the volunteer is asked to command a machine to perform a particular task through an interface based only on gestures. Thus, the volunteers are induced to express their real perception of what gesture corresponds to a certain task, while in other words, it is not necessary to spend time thinking about what gesture should be done, since the gestures come directly from the subconscious. Gestures obtained in this way should be more spontaneous, which, according to McNeill [2], corresponds to gestures with a high degree of intuitiveness.

Since this paper is focused on finding a set of intuitive gestures, the subconscious approach is adopted, and it is applied through the use of experiments that make the volunteer perform gestures in the most spontaneous way possible, a concept from psycholinguistics (see Section II-A. To do this, a common methodology is that of the Wizard of OZ (WoZ) [37], [38], where the volunteer interacts with a machine while thinking that it is acting in an autonomous manner, when in fact the machine is, without the volunteers’ knowledge, being controlled remotely by another person. In this way, as the volunteers think that they are actually interacting with the machine, the interaction becomes more spontaneous and the volunteers tend to perform gestures that, subconsciously, are the most intuitive.

In the methodology described by Nielsen et al. [32] and Wachs [10], volunteers are instructed to perform just one gesture for each task. But, this approach may not be suitable to obtain an intuitive gesture vocabulary. In [39], an experiment following the previously mentioned methodology was performed, obtaining a gesture vocabulary to a set of tasks. A month later, another experiment was performed with the participation of the same volunteers. The main objective of the second experiment was to determine whether the volunteers would select, for each task, the same gesture as they did in the first experiment. To do so, before choosing a gesture to represent a particular task, volunteers could see all gestures that were selected, corresponding to that task, in the first experiment by all volunteers. The obtained results shown that 66% of the volunteers changed the chosen gesture and that 70% of the gestures of the vocabulary obtained in the second experiment were different from those of the first one. To validate such results, the second experiment was repeated, but this time with a set of volunteers that did not participate in the first two experiments. After analyzing the new results, it was found that most of the gestures of the third vocabulary were equal to those of the second, leading to conclusion that methodologies, which take into account only the repetition frequency of gestures performed in just one experiment, do not necessarily guarantee obtaining a vocabulary composed of popular gestures. This expresses the possibility for the existence of other gestures, also intuitive, which cannot be obtained when the volunteer is instructed to perform just one gesture for each task. So, to have a stronger guarantee that the captured gestures are the most intuitive for each volunteer, instead of performing the same experiment many times, it is more viable to capture all of the gestures, in a single experiment, that volunteers believe are suitable for each task.

To accomplish this, it is proposed in this work that each volunteer perform gestures that correspond to the task until they run out of ideas. For this, it is necessary that the communication interface never allow for the machine to comply with the commands of volunteers while they are instructed to perform gestures until the interface responds properly. In this way, volunteers will perform all of the gestures that they think are intuitive for the machine to perform a specific task, until they become frustrated at not having been able to interact. Such procedure addresses both the need to capture more user gestures [31], [39], as well as avoiding the need to establish a fixed number of gestures for each user, which is a problem for the Production Principle [31]. We name such approach Frustration Based Approach (FBA).

Note that each volunteer performs just one experiment for a single task because, once the frustration has been reached, the volunteer would not interact intuitively with the robot for another task. Despite this, a large number of gestures with
a relatively high level of intuitiveness corresponding to each task could be obtained from a single volunteer, as the gestures would be spontaneous until frustration. An important aspect to be addressed is that, even after carrying out the experiments, volunteers should not know its real purpose, since they could reveal it, even unintentionally, to other potential future volunteers.

All experiments were filmed and saved, so that in the next steps it is possible to analyze the gestures performed and to choose those that should be part of the vocabulary. Each video must be referenced by the task performed and listed with an identifier of the corresponding volunteer, so that afterwards, it is possible to identify to which task the video is associated with and link it with the respective volunteer.

C. ANALYSIS OF CAPTURED GESTURES
After collecting gesture data in the form of video, it is necessary to follow some steps to analyze the filmed gestures.

1) ANALYSIS OF THE GESTURES FOUND IN EACH VIDEO
The recorded experiments need to be analyzed with the aim of identifying each of the gestures performed. For this, it is necessary to create two tables, the first one, called Gesture Description Table (Table 1), with two columns: gesture description and gesture identifier (gesture ID); and the second table, called Performed Gestures Table, with four columns: the task corresponding to the gesture, the volunteer identifier (volunteer ID), the gesture ID, and the time when the gesture was performed. With these tables, the analysis of gestures can be made following these steps:

- **Step 1.** For each new gesture performed, the evaluator (the person responsible for the analysis of captured gestures) judges if the gesture is compatible with one of the gestures previously described in the Gesture Description Table. If yes, move on to step 3, if not, move on to step 2.
- **Step 2.** Since the performed gesture does not correspond to none of the described gestures, a new entry is created in the Gesture Description Table, in which the gesture receives an identifying number and a description. Move to Step 4.
- **Step 3.** After identifying the gesture, a new entry should be added to the Performed Gestures Table containing the name of the task associated with the performed gesture, the volunteer ID, the gesture ID and the time (in seconds) in which it was performed.
- **Step 4.** Return to step 1 until all gestures of all volunteers have been analyzed.

The gesture analysis presented in this subsection should be performed with much caution, since it is highly subjective and the results will directly influence the obtained vocabulary. In this sense, it is recommended that such analysis be carried out by more than one person.

2) CALCULATION OF THE OCCURRENCE RATES FOR EACH TASK
With the aim of making the selection of the gestures that will compose the vocabulary as objective as possible, we propose the use of three different occurrence rates for each gesture associated to each task.

1) **General Occurrence Rate (GOR).** This is the ratio of the number of times a certain gesture is performed to the total number of performed gestures (repeated or not) for each task (taking into consideration all gestures performed by all volunteers). With this, it is possible to measure the global importance of each gesture in relation to all gestures performed for a certain task.

2) **Volunteer Occurrence Rate (VOR).** Even after the calculation of the GOR, the percentage of volunteers that performed a gesture is still unknown. Therefore, the Occurrence Rate per Volunteer is defined by the ratio of the number of volunteers that performed a certain gesture and the number of volunteers that participated in the experiments, allowing us to know which gesture was performed by a larger or smaller percentage of volunteers.

3) **Occurrence Rate by Time (ORT).** It is possible that gestures performed first are more intuitive. Therefore, it is necessary to attribute different importance for each gesture in a manner that the gestures performed first are of greater importance. To get a fair judgment with relation to the occurrence of each gesture, ORT considers the time in which the gesture was performed and the duration of the experiment.

For a more clear description of the occurrence rates for a specific task, consider the following definitions:

- \( G_i \) is the gesture identified by ID \( i \) (see Table 1);
- \( N_e \) is the number of experiments (volunteers) performed;
- \( N_{ei} \) is the number of volunteers that performed \( G_i \);
- \( N_i \) is the number of occurrences of \( G_i \) for all experiments;
- \( N \) is the number of occurrences of all gestures performed for all experiments;
- \( ST_{ij} \) is the sum of differences between the duration of the \( j \)-th experiment and all occurrence times of \( G_i \) in the same experiment;
- \( ST_j \) is the sum of differences between the duration of the \( j \)-th experiment and all occurrence times of all gestures in the same experiment;

Therefore, the occurrence rates of a gesture \( G_i \) for a certain task can be calculated as follows:

\[
GOR = \frac{N_i}{N}, \tag{1}
\]

\[
ORV = \frac{N_{ei}}{N_e}; \tag{2}
\]

\[
ORT = \frac{\sum_{j=1}^{N_e} ST_{ij}}{\sum_{j=1}^{N_e} ST_j}. \tag{3}
\]

D. CHOICE OF VOCABULARY (INTUITIVENESS TABLE)
The vocabulary is composed by the gestures that best represent each task. Therefore, for each task, it is necessary to construct an Intuitiveness Table performing the following steps:
TABLE 1. Gesture description table (first gestures for the experiments described in Section IV).

| Gesture ID | Gesture Description                                                                 |
|------------|--------------------------------------------------------------------------------------|
| 1          | One flat hand slowly moving top-down                                                 |
| 2          | Flat hands moving horizontally in opposite directions from the center of the chest to the sides of the body (two times) |
| 3          | Both flat hands slowly moving top-down                                               |

... ... ...

- **Step 1.** The first four columns of the Intuitiveness Table contain the gestures ID, GOR, VOR and ORT;
- **Step 2.** To allow a scaled composition of occurrence rates, each one of them is normalized. To do so, for each occurrence rate, the value assigned to all gestures is divided by the highest one of them. The normalized occurrence rates are added to the Intuitiveness Table;
- **Step 3.** The Intuitiveness Level (IL) of each gesture is defined as the arithmetic mean of the normalized occurrence rates. The IL is the last column of the Intuitiveness Table.

Since so far there is not enough information to properly assess the relative importance of the three proposed indices, it was decided to attribute the same importance to each one of them in the calculation of the IL. That is why in Step 3 the arithmetic average is taken instead of a weighted one. Therefore, the higher the IL for a gesture, more intuitive is the gesture for a task, a concept based on psycholinguistic concepts. An example of an Intuitiveness Table is presented in Table 2, using real experimental data obtained for task Pointing (see Section IV).

**IV. APPLICATION OF THE PROPOSED METHODOLOGY**

To validate the proposed methodology, a vocabulary of intuitive gestures for interaction between humans and mobile robots was designed. For this, each step described in Section III was performed as presented in the next subsections.

**A. DEFINITION OF TASKS**

As this paper is focused only on presenting and validating the proposed methodology, and not on solving a specific problem, generic tasks in HRI were searched in the literature [4], [18]. As a result, seven tasks were selected: abort, pointing, attention, slower, faster, ok, follow me. These are common tasks for a mobile robot and can be understood as commands that should be passed through the interface to the robot. Thus, the obtained vocabulary may be used in future research where such tasks are also included.

**B. CAPTURING THE GESTURAL DATA**

To be able to capture gestures that most intuitively represent each of the above tasks, 84 experiments (volunteers) were performed; 12 for each task. The volunteers were undergraduate students from three different universities (88% engineering students), aged from 18 to 49 years old and about 80% of them male.

Before performing the experiments, the maximum time of duration of the experiments was determined taking into consideration that within this time the volunteer should reach frustration. Thus, pilot experiments were performed with seven members of the Robotics Research Group (GPR-UFS), where it was observed that it took around 18 seconds for volunteers to become frustrated for being unable to interact with the robot. So, to increase the likelihood of frustration and at the same time to minimize the cost of storing the captured data, a minimum time of 20 seconds and a maximum time of 30 seconds was imposed for each experiment.

In each experiment, volunteers were invited to interact with a mobile robot (Pioneer 3-DX) and to perform one of the 7 chosen tasks. Each volunteer was told that the robot should perform the selected task and that it was represented by a single and predefined gesture. Thus, volunteers had to perform gestures which they thought were suitable for the desired task until they performed the one recognized by the robot. However, as the robot was controlled by a wireless joystick (using the WoZ approach), the robot would never respond in a satisfactory manner, which would take the volunteer to intuitively perform every gesture that seemed suitable for that task until reaching frustration (FBA) and giving up to interact with the robot. At this point, the objective of the experiment was reached. Frustration could be noticed when the volunteer questioned the functioning of the interface or showed a lack of ideas on what gestures to perform. Thus, each experiment was finalized after a time between 20 and 30 seconds from the beginning, depending on robot’s velocity or if the volunteer became frustrated. In a joint analysis, the experiments had an average duration of 23.47 ± 3.44 seconds. The experiments setup can be seen in Fig. 1.

Before the beginning of the experiment, in order to persuade volunteers that the robot is really able to interact through gestures, the volunteer was told that, to verify if the gesture interface was already working properly, he should perform a certain gesture, demonstrated by the researcher, that would tell the robot to go back for about a half meter.

When the researcher observed the volunteer performing the defined gesture, a command was sent to the robot through a wireless joystick, telling it to go back, thus convincing volunteers that they were actually commanding the robot through the gesture.

As indicated in the proposed methodology, each volunteer made just one experiment and was instructed in the following way about each task.
TABLE 2. Intuitiveness table with real experimental data for task pointing, sorted by IL. Abbreviation captions: general occurrence rate (GOR), volunteer occurrence rate (VOR), occurrence rate by time (ORT), intuitiveness level (IL).

| Gesture ID | GOR  | VOR  | ORT  | GOR Norm. | VOR Norm. | ORT Norm. | IL  |
|------------|------|------|------|-----------|-----------|-----------|-----|
| 76         | 0.41 | 0.67 | 4.55 | 1.00      | 1.00      | 1.00      | 1.00|
| 78         | 0.16 | 0.42 | 2.21 | 0.39      | 0.63      | 0.49      | 0.50|
| 77         | 0.05 | 0.33 | 0.78 | 0.13      | 0.50      | 0.17      | 0.27|
| 79         | 0.07 | 0.33 | 0.60 | 0.16      | 0.50      | 0.13      | 0.26|
| 27         | 0.08 | 0.25 | 0.96 | 0.18      | 0.37      | 0.21      | 0.26|
| 81         | 0.04 | 0.17 | 0.76 | 0.11      | 0.25      | 0.17      | 0.17|
| 31         | 0.03 | 0.17 | 0.54 | 0.08      | 0.25      | 0.12      | 0.15|
| 5          | 0.03 | 0.17 | 0.21 | 0.08      | 0.25      | 0.05      | 0.12|
| 82         | 0.04 | 0.08 | 0.47 | 0.11      | 0.12      | 0.10      | 0.11|
| 80         | 0.01 | 0.08 | 0.33 | 0.03      | 0.12      | 0.07      | 0.07|
| 3          | 0.01 | 0.08 | 0.26 | 0.03      | 0.12      | 0.06      | 0.07|
| 83         | 0.01 | 0.08 | 0.13 | 0.03      | 0.12      | 0.03      | 0.06|
| 37         | 0.01 | 0.08 | 0.08 | 0.03      | 0.12      | 0.02      | 0.06|
| 84         | 0.01 | 0.08 | 0.07 | 0.03      | 0.12      | 0.01      | 0.06|
| 26         | 0.01 | 0.08 | 0.04 | 0.03      | 0.12      | 0.01      | 0.05|
| 59         | 0.01 | 0.08 | 0.01 | 0.03      | 0.12      | 0.00      | 0.05|

FIGURE 1. Environment where the experiments were performed. (a) Researcher’s view. (b) Volunteer’s view. (c) View of an outsider observer.

1) **Abort.** The purpose of this task is to tell the robot to stop performing a task, independent of what stage of the task it is at. Volunteers were told that the robot was assigned a task consisting of moving towards them, and they should perform gestures that the robot could understand as an abort command. However, to induce frustration, the robot would continue to move towards volunteers until the end of the experiment.

2) **Attention.** The objective of this task is to drive robot’s attention to someone who is trying to interact with it. So, while the robot remained still, volunteers were instructed to perform gestures until the robot notices their presence and responded by moving towards them. However, to induce frustration, the robot would remain still until the end of the experiment.

3) **Pointing.** The purpose of this task is to tell the robot to go to a specific point of the environment. While the robot remained still, volunteers were instructed to perform gestures indicating the robot to which point of the environment it should go. However, the robot would never move, frustrating the volunteers.

4) **Slower.** The purpose of this task is to make the robot slow down. Thus, volunteers were instructed to perform gestures that commanded the robot to reduce its speed. However, this would never happen and the robot would move at a constant speed of 12 cm/s and would only stop when the experiment ended.

5) **Faster.** This task is the opposite of the latter, as the idea is to speed the robot up. For this, volunteers were instructed to perform gestures that commanded the robot to speed up. However, this would never happen, and the robot would move towards the volunteer at a constant speed of 5 cm/s, and would only stop at the end of the experiment.

6) **Ok.** The purpose of this task is to give the robot positive feedback on the last performed task, indicating the completion of its execution. Volunteers were told that the robot was assigned a task to move towards them. So, while the robot moved, volunteers should perform gestures indicating that the task had been completed with success. The robot should respond by stopping, since
the task of moving towards the user was successfully completed. However, the robot would never stop, and would continue in the direction of the volunteers until the end of the experiment.

7) **Follow me.** The objective of this task is to tell the robot to follow volunteers. So, while the robot remained still, volunteers were asked to perform gestures that commanded the robot to follow them. However, the robot kept its position until the end of the experiment, thus driving the volunteer to frustration.

RGB and depth images, and skeleton data acquired by a Kinect 360 sensor were captured, processed and stored as separated files, which were organized by experiments, to be later used to choose the gestures that composed the vocabulary.

When applying the WoZ methodology, volunteers should not notice that someone was controlling the robot. However, in this data capture, the researcher who conducted the experiment also played the role of the “Wizard of Oz”. This was possible because the researcher followed the experiment through the image captured by Kinect and viewed on the computer screen. In addition, as mentioned earlier, the wireless joystick used to control the robot was out of the sight of volunteers at all times, as seen in Fig. 1(a).

**Remark 1:** Although the experiments involve humans, images or any other sensitive data of the volunteers were not disclosed. In addition, all volunteers signed a consent form, through which they expressed their agreement to participate in the experiments after knowing the content of the research being carried out. Under these conditions, our research is exempt from the need for approval by an ethics committee.

### C. Analysis of Gestures

Video data recorded in all experiments were analyzed to create the tables described in Section III). During the analysis, 97 different gestures were found, and each of them was identified and described carefully in an effort to reduce the inherent subjectivity of the process. Thus, for the total of 36 minutes of captured video for all 84 experiments, a total time of 18 hours were spent in analysis. For the analysis, a specific application was developed, which loaded the video referring to each experiment, allowed pauses, and allowed to forward and backward frames. Each frame shown on the screen was timestamped relative to the first frame of the video, so that the evaluator could identify the exact moment when each gesture was initiated.

To identify which gesture was performed, the evaluator used the Gesture Description Table (Table 1) and verified whether any of the gestures described there corresponded to the observed gesture. If so, a spreadsheet received the volunteer’s identification, the identification of the gesture performed and its start time. If the gesture was new, its description was included in the table and an incremental ID was assigned to it. Note that in the analysis of the first video, the Gesture Description Table was empty and was filled as the videos of all volunteers were evaluated and new gestures were identified.

| Task       | Gesture ID | GOR | VOR | ORT | GOR Norm. | VOR Norm. | ORT Norm. | IL |
|------------|------------|-----|-----|-----|-----------|-----------|-----------|----|
| Abort      | 2          | 0.38| 0.92| 5.14| 1.00      | 1.00      | 1.00      | 1.00|
| Abort      | 5          | 0.15| 0.50| 1.74| 0.40      | 0.55      | 0.34      | 0.43|
| Abort      | 3          | 0.13| 0.50| 0.82| 0.36      | 0.55      | 0.16      | 0.35|
| Attention  | 28         | 0.22| 0.67| 2.94| 1.00      | 1.00      | 1.00      | 1.00|
| Attention  | 26         | 0.11| 0.42| 1.05| 0.52      | 0.63      | 0.36      | 0.50|
| Attention  | 4          | 0.08| 0.33| 1.14| 0.39      | 0.50      | 0.39      | 0.43|
| Faster     | 39         | 0.14| 0.42| 2.09| 1.00      | 1.00      | 1.00      | 1.00|
| Faster     | 27         | 0.13| 0.33| 1.36| 0.93      | 0.80      | 0.65      | 0.79|
| Faster     | 48         | 0.12| 0.33| 1.48| 0.86      | 0.80      | 0.71      | 0.79|
| Follow     | 27         | 0.31| 0.83| 4.18| 1.00      | 1.00      | 1.00      | 1.00|
| Follow     | 36         | 0.13| 0.42| 1.92| 0.44      | 0.50      | 0.46      | 0.47|
| Follow     | 22         | 0.10| 0.50| 1.10| 0.31      | 0.60      | 0.26      | 0.39|
| Ok         | 43         | 0.18| 0.75| 3.02| 0.96      | 1.00      | 1.00      | 0.99|
| Ok         | 64         | 0.19| 0.75| 2.14| 1.00      | 1.00      | 0.71      | 0.90|
| Ok         | 5          | 0.12| 0.75| 1.85| 0.65      | 1.00      | 0.61      | 0.75|
| Pointing   | 76         | 0.41| 0.67| 4.55| 1.00      | 1.00      | 1.00      | 1.00|
| Pointing   | 78         | 0.16| 0.42| 2.21| 0.39      | 0.63      | 0.49      | 0.50|
| Pointing   | 77         | 0.05| 0.33| 0.78| 0.13      | 0.50      | 0.17      | 0.27|
| Slower     | 5          | 0.28| 0.67| 3.26| 1.00      | 1.00      | 1.00      | 1.00|
| Slower     | 85         | 0.15| 0.33| 1.83| 0.54      | 0.50      | 0.56      | 0.53|
| Slower     | 86         | 0.11| 0.33| 1.25| 0.39      | 0.50      | 0.38      | 0.43|
For future applications of the proposed methodology, to mitigate the subjectivity of this analysis, the videos of the experiments could be analyzed by two different researchers, and a third researcher could act as a referee in cases where the two disagree.

D. CHOOSING THE VOCABULARY

After analyzing and calculating the Intuitiveness Level (IL) of each gesture using the proposed methodology, a list of gestures for each task ordered in a descending manner by IL (Table 3) was compiled. Fig. 2 illustrates the vocabulary obtained by choosing gestures with the higher IL for each task. In this work, gestures with IL above 0.90 were selected to compose the vocabulary. So, task “Ok” is represented by two gestures, as seen in Fig. 2 with the IDs in Table 3. If the gestures in Fig. 2 is performed with only one arm, they are considered the same, no matter what arm is used.

Note that the FBA, in addition to solving the problem of capturing several gestures from each volunteer without having to determine a fixed number of them, is a user-centered approach, which, according to [32], is the most suitable approach to elicit intuitive gestures. In addition to these aspects, the use of the WoZ approach, as is done in several works that aim to elicit commands for interaction with computational devices and software interfaces, tends to induce the volunteer to perform those gestures that are most representative for each task. These characteristics contribute both to the practical aspect of the proposed methodology and to the intuitiveness of the gesture vocabulary elicited with it.

V. CONCLUSION

Recognizing the importance of interactive interfaces for the popularization of robotics and the lack of studies that prioritize the intuitiveness of them, this paper proposes a new methodology scientifically based on psycholinguistics, that is thoroughly described to allow its fully reproduction, which can be used to obtain an intuitive and robust gesture vocabulary for human-robot interaction.

According to psycholinguistics [2], spontaneous gestures are naturally more intuitive. Therefore, the proposed methodology consists in the realization of user-based subconscious experiments and the use of the Frustration Based Approach (FBA), proposed in this work, in order to obtain as many spontaneous gestures as possible from each volunteer. Besides, three occurrence rates which consider different aspects of gesture’s intuitiveness were introduced and used to calculate the Intuitiveness Level (IL) of each gesture, a score to rank gestures according with its intuitiveness, also proposed in this work. Another interesting aspect of the proposed methodology is the possibility to get a complex vocabulary, where more than one gesture can be assigned to a certain task, allowing an increase in robustness.

The methodology has the ability to acquire several gestures for each task (following the “Production Principle” [31]), in addition to offering an intuitiveness metric based on three sub-metrics that rank the most intuitive gestures for each task, an open problem also addressed by [31]. Furthermore, the frustration approach does not require that a specific number of gestures to be stipulated for the experiment, another issue in this field [31].

The intuitiveness level (IL) is calculated using an arithmetic average of the three proposed metrics. However, it is possible that one of those metrics is more important for IL than the others, which opens the possibility of a future study about the best way to estimate the weights to ponder the three metrics.

Another work that can be done in the future is carrying out the experiments in an environment closer to the real situation where humans will interact with robots to better contextualization of operation. This will possibly encourage volunteers to perform even more gestures, or even eliminate some gestures that would not be intuitive for such situations.

Finally, the proposed methodology is not limited only to HRI, and it can also be used for HMI.

REFERENCES

[1] A. Wexelblat, “Research challenges in gesture: Open issues and unsolved problems,” in Gesture and Sign Language in Human-Computer Interaction (Lecture Notes in Computer Science), vol. 1371, I. Wachsmuth and M. Fröhlich, Eds. Berlin, Germany: Springer, 1998, ch. 1, pp. 1–11.
[2] D. McNeill, Hand and Mind: What Gestures Reveal About Thought. Chicago, IL, USA: Univ. Chicago Press, 1992.
[3] E. T. Hall, The Silent Language. New York, NY, USA: Anchor Book, 1959.
[4] S. Waldherr, S. Thrun, R. Romero, and D. Margaritis, “Template-based recognition of pose and motion gestures on a mobile robot,” in Proc. AAAI/IAAI, 1998, pp. 977–982.
[5] J. L. Burke, R. R. Murphy, E. Rogers, V. J. Lumelsky, and J. Scholtz, “Final report for the DARPA/NSF interdisciplinary study on human-robot interaction,” IEEE Trans. Syst., Man, C, Appl. Rev., vol. 34, no. 2, pp. 103–112, May 2004.
[6] V. Kalyukin, “Human-robot interaction through gesture-free spoken dialogue,” Auto. Robots, vol. 16, no. 3, pp. 239–257, May 2004.
[7] P. Brenner, A. Pipe, C. Melhuish, M. Fraser, and S. Subramanian, “Conversational gestures in human-robot interaction,” in Proc. IEEE Int. Conf. Syst., Man Cybern., Oct. 2009, pp. 1645–1649.
[8] A. Uribe and S. Alves, “Task planning for human-robot interaction,” in Proc. Robot. Symp., 2011, pp. 81–85.
[9] A. Powers, “What robotics can learn from HCl,” Interactions, vol. 15, no. 2, p. 67, 2008.
Why We Gesture: The Surprising Role of Hand Movements in

D. McNeill, I. Vilà-Giménez, N. Dowling, Ö. E. Demir-Lira, P. Prieto, and S. Goldin-Dastick, "Recurrent gestures: How the mental reflects the social," J. Wachs, "Optimal hand gesture vocabulary design methodology for

T. Fujii, J. H. Lee, and S. Okamoto, "Gesture recognition system for

D. McNeill, Why We Gesture: The Surprising Role of Hand Movements in Communication. Cambridge, U.K.: Cambridge Univ. Press, 2015.

D. McNeill, Gesture and Thought. Chicago, IL, USA: Univ. Chicago Press, 2005.

M. Filho, A. Inaldo, M. M. Teixeira, and I. M. O. Maia, “Gestures as a cognitive means of human-computer interaction,” Blucher Des. Proc., vol. 1, no. 2, pp. 1034–1042, 2014.[Online]. Available: www.proceedings.blucher.com.br/article-details/8761

W. L. Davidson, “Definition of intuition,” Mind, vol. 7, no. 26, pp. 304–310, 1882.

T. Fuji, J. H. Lee, and S. Okamoto, “Gesture recognition system for human-robot interaction and its application to robotic service task,” in Proc. Int. Conf. Eng. Comput. Scientists, vol. 1, Hong Kong, 2014.

J. O. Wobbrock, H. H. Aung, B. Rothrock, and B. A. Myers, “Maximizing the guessability of symbolic input,” in Proc. CHI Extended Abstr. Hum. Factors Comput. Syst., Apr. 2005, pp. 1869–1872.

J. O. Wobbrock, M. R. Morris, and A. D. Wilson, “User-defined gestures for surface computing,” in Proc. SIGCHI Conf. Hum. Factors Comput. Syst., Apr. 2009, p. 1083.

R. D. Vatavu and J. O. Wobbrock, “Formalizing agreement analysis for elicitation studies: New measures, significance test, and toolkit,” in Proc. 33rd Ann. ACM Conf. Hum. Factors Comput. Syst., Apr. 2015, pp. 1325–1334.

E. Chan, T. Seyed, W. Stuerzlinger, X.-D. Yang, and F. Maurer, “User elicitation on single-hand microgestures,” in Proc. CHI Conf. Hum. Factors Comput. Syst., May 2016, pp. 3403–3414.

B. F. Gheran, J. Vanderdonckt, and R.-D. Vatavu, “Gestures for smart rings: Empirical results, insights, and design implications,” in Proc. Designing Interact. Syst. Conf., Jun. 2018, pp. 623–635.

H. Wu and L. Yang, “User-defined gestures for dual-screen mobile interaction,” Int. J. Hum.–Comput. Interact., vol. 36, no. 10, pp. 978–992, Jun. 2020.

L. H. Kim, D. S. Drew, V. Domova, and S. Follmer, “User-defined swarm robot control,” in Proc. CHI Conf. Hum. Factors Comput. Syst., Apr. 2020, pp. 1–13.

J. L.-Pérez-Medina, S. Villarreal, and J. Vanderdonckt, “A gesture elicitation study of nose-based gestures,” Sensors, vol. 20, no. 24, p. 7118, Dec. 2020.

A. Ali, M. R. Morris, and J. O. Wobbrock, “‘I am iron man’ priming improves the learnability and memorability of user-elicited gestures,” in Proc. CHI Conf. Hum. Factors Comput. Syst., 2021, pp. 1–14.

Z. Cui, H. Gong, Y. Wang, C. Shen, W. Zou, and S. Luo, “Enhancing interactions for in-car voice user interface with gestural input on the steering wheel,” in Proc. 13th Int. Conf. Automot. User Interface Interact. Veh. Appl., Sep. 2021, pp. 59–68.

Y. Takayama, Y. Ichikawa, B. Shizuki, I. Kawaguchi, and S. Takahashi, "A user-based mid-air hand gesture set for spreadsheets," in Proc. Asian CHI Symp., May 2021, pp. 122–128.

S. Villarreal-Narvaez, J. Vanderdonckt, R.-D. Vatavu, and J. O. Wobbrock, “A systematic review of gesture elicitation studies: What can we learn from 216 studies?” in Proc. ACM Designing Interact. Syst. Conf., Jul. 2020, pp. 855–872.

M. R. Morris, A. DanieleCsu, S. Drucker, D. Fisher, B. Lee, M. C. Schraefel, and J. O. Wobbrock, “Reducing legacy bias in gesture elicitation studies,” Interactions, vol. 21, no. 3, pp. 40–45, May 2014.

M. Nielsen, M. Störring, T. B. MoeSlund, and E. Granum, “A procedure for developing intuitive and ergonomic ges- ture sets for the HCI,” Gesture-Based Commun. Hum.–Comput. Interact., vol. 2915, pp. 409–420, 2004.

F. G. Pereira, R. F. Vassallo, and E. O. T. Salles, “Human–robot interaction and cooperation through people detection and gesture recognition,” J. Control. Autom. Electr. Syst., vol. 24, no. 3, pp. 187–198, Jun. 2013.

C. Canuto et al.: Intuitiveness Level: Frustration-Based Methodology for HRI Gesture Elicitation

C. Canuto received the B.S. degree in information systems from Tiradentes University, Brazil, in 2013, the M.Sc. degree in computer science from Federal University of Sergipe, Brazil, in 2016, and the Ph.D. degree in electrical engineering from the Federal University of Espírito Santo, Brazil, in 2020. His research interests include computer vision, gesture recognition, and human–robot interaction.

E. O. Freire received the bachelor’s degree in electrical engineering from the Federal University of Pará, in 1995, and the M.A.Sc. and Ph.D. degrees in electrical engineering from the Federal University of Espírito Santo, Brazil, in 1997 and 2002, respectively. He is currently a Titular Professor with the Federal University of Sergipe.

L. Molina received the bachelor’s degree in electronic engineering from the Federal University of Sergipe (UFS), in 2007, the master’s degree in control, automation, and robotics from the Federal University of Rio de Janeiro, in 2010, and the Ph.D. degree in electrical engineering from the Federal University of Campina Grande, in 2014. He is currently an Adjunct Professor with UFS. He is also a Researcher with the Robotics Research Group of UFS (GPR-UFS).

A. N. Carvalho (Member, IEEE) received the bachelor’s degree in electronic engineering from the Federal University of Sergipe, in 2006, and the master’s and Ph.D. degrees in electrical engineering from the Federal University of Campina Grande, in 2007 and 2012, respectively. He is currently an Associate Professor III with the Federal University of Sergipe. He is also a permanent member of the Graduate Program in Electrical Engineering (PROEE).

S. Givigi (Senior Member, IEEE) received the Ph.D. degree in electrical and computer engineering from Carleton University, Ottawa, ON, Canada, in 2009. He is currently an Associate Professor with the School of Computing, Queen’s University, Kingston, ON, Canada. His current research interests include autonomous systems and robotics.

17154 VOLUME 10, 2022