Article

Enhancing Interpretation of Ambiguous Voice Instructions based on the Environment and the User's Intention for Improved Human-Friendly Robot Navigation

M. A. Viraj J. Muthugala *, P. H. D. Arjuna S. Srimal and A. G. Buddhika P. Jayasekara

Robotics and Control Laboratory, Department of Electrical Engineering, University of Moratuwa, Moratuwa 10400, Sri Lanka; arjuna@elect.mrt.ac.lk (P.H.D.A.S.S.); buddhika@elect.mrt.ac.lk (A.G.B.P.J.)
* Correspondence: viraj@elect.mrt.ac.lk; Tel.: +94-71-629-6145

Abstract: Human-friendly interactive features are preferred for domestic service robots. Humans prefer to use verbal communication in order to convey instructions to peers. Those voice instructions often include uncertain terms such as "little" and "far". Therefore, the ability to quantify such information is mandatory for human-friendly service robots. The meaning of such voice instructions depends on the environment and the intention of the user. Therefore, this paper proposes a method in order to interpret the ambiguities in user instructions based on the environment and the intention of the user. The actual intention of the user is identified by analyzing the pointing gestures accompanied with the voice instructions since pointing gestures can be used in order to express the intention of the user. A module called the motion intention switcher (MIS) has been introduced in order to switch the intention of the robot based on the arrangement of the environment and the point referred by the gesture. Experiments have been carried out in an artificially-created domestic environment. According to the experimental results, the behavior of the MIS is effective in identifying the actual intention of the user and switching the intention of the robot. Moreover, the proposed concept is capable of enhancing the uncertain information evaluation ability of robots.

Keywords: understanding uncertain information; user intention; human–robot interaction; human-friendly robotics; service robotics

1. Introduction

An intelligent service robot is a machine that is able to perceive the environment and use its knowledge to operate safely in a meaningful and purposive manner [1]. Intelligent service robots are being developed as a solution for the widening gap between supply and demand of human caregivers for elderly/disabled people [2–5]. These service robots are intended to be operated by non-expert users in human-populated environments. Thus, human-friendly interactive features are preferred for domestic service robots [6].

Verbal communication is one of the most widely used communication modalities by humans in order to communicate with companions. Therefore, human-like verbal communication abilities are favored for domestic service robots with human-friendly interactive features [7,8]. The natural verbal communications phrases and utterances that indicate the distances often include uncertain terms such as "little" and "far". These uncertain terms are sometimes referred to as fuzzy linguistic information, qualitative terms and fuzzy predicates. The quantitative meaning of uncertain information depends on various factors such as environment, context and experience. As an example, consider a situation where a person is standing in front of wall with 5 m gap between the wall and the person. In this

case, the person may move about 1–1.5 m for the command, "move little forward". However, if the gap between the person and the wall is 1 m then the response of that person for the same command would be a movement of 20–30 cm. Therefore, service robot must possess human-like cognitive ability in understanding uncertain information in order to provide better interaction and service for the human users.

Methods for understanding natural language commands related to object manipulations and navigation have been developed and the systems are capable of understanding and successfully executing the robot actions required for fulfilling natural language commands [9,10]. However, the systems are not effective in quantifying the meaning of uncertain information in user instructions such as "little" and "far". The system proposed in [11] is capable of generating natural language spatial descriptors that include uncertain terms. However, the quantitative meanings of uncertain terms are fixed. Methods based on fuzzy inference systems for quantifying predetermined values for the uncertain terms in verbal instruction based on the current state of the robot have been proposed [12,13]. The method proposed in [14] adapts the meaning of uncertain information based on the immediate previous state of the robot. The method proposed in [15] evaluates a set of previous states instead of the immediate previous state for enhanced interpretation. The methods proposed in [16,17] use fuzzy neural networks that are capable of adapting the perception of uncertain information based on the user critics. However, these systems do not consider environmental factors for adaptation in a manner similar to humans, and they cannot adapt perception according to the environment even though the fuzzy implications related to the spatial information heavily depend on the environment. A method for manipulating objects through voice instructions with fuzzy predicates has been developed [18]. The method is capable of evaluating crisp values for fuzzy predicates in user instructions by evaluating the average distance between surrounding objects in its vision field. A concept to scale the fuzzy fluent related to positional information based on the size of the frame/point of view has been introduced in [19]. However, according to [20], the size of the frame (e.g., size of the room) is not enough for adapting the perception of uncertain information related to navigational commands. Hence, the method considers more environmental factors for the adaptation. Moreover, the proposed method evaluates the arrangement of the surrounding environment in a more rational way in order to adapt the perception of uncertain information. However, there are limitations in the proposed method in identifying the actual intention of the user and effectively acting in some scenarios (a detailed explanation is given in Section 3.3.1). According to [21,22] the understanding of voice instructions could be improved by fusing the information conveyed from gestures with the language instructions. However, the proposed systems are not capable of quantifying distance-related uncertain information in user instructions. Hence, the proposed methods cannot be adopted in order to improve the quantification ability of uncertain information in voice instructions.

Therefore, this paper proposes a method in order to switch the intention of the robot by identifying the actual intention of the user by analyzing the pointing gestures accompanied with voice instructions for enhanced interpretation of navigation instructions with uncertain information such as "move far forward". The overall functionality of the proposed system is explained in Section 2. The proposed method for switching the intention of the robot by identifying the actual intention of the user based on the pointing gestures is explained in Section 3 along with rationale behind the concept. Experimental results are presented and discussed in Section 4. Finally, the conclusions are presented in Section 5.

2. System Overview

A functional overview of the system is depicted in Figure 1. The goal of the system is to provide a way for effective identification of the intention of the user based on multimodal user commands for enhanced quantification of uncertain terms related to distances (e.g., "little" and "far") in motional navigation commands such as "move far forward". The voice recognition and understanding section converts voice into text and parses the commands with aid of the language memory. The voice response generation section is a text-to-speech converter that can be used in

order to generate the voice responses of the robot. The overall interaction between the robot and the user is managed by the interaction management module (IMM). These modules have been implemented similar to the system explained in [23]. The gesture evaluation module is deployed for identifying the non-verbal instructions accompanied with voice instructions by analyzing the skeleton of the user returned by the Kinect motion sensor attached to the robot. The analyzed body postures of the user are then fed into the intention identifier module (IIM) in order to identify the intention of the user related to the given voice instruction. Based on the intention of the user, the required actions for fulfilling a command may be switched by the motion intention switcher (MIS). Subsequently, the parameters required for the quantification of the uncertain information by the uncertain information understanding module (UIUM) [20] will be modified by MIS, if alterations are required.
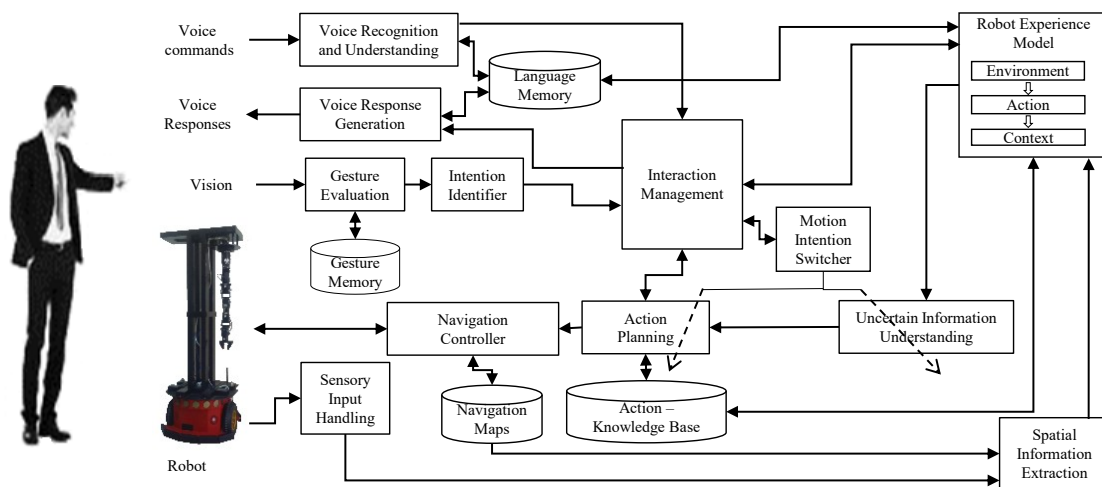

Figure 1. System overview.

The robot experience model (REM) [23] is a hierarchical structure that holds the knowledge about the environment, actions and context in a way that the knowledge can be used by the robot for fulfilling the actions in the robot's domain. The parameters required for interpretation of uncertain information by the UIUM is also retrieved from the REM. The low-level navigation controlling functionalities such as localization within a given navigation map are managed by the navigation controller. The required navigation maps can be created using the Mapper3 application. The information from the low-level sensors of the robot such as sonar sensors is retrieved by the sensory input handling module (SIHM). The spatial information extraction module (SIEM) extracts the information about the environment from the navigation maps and the SIHM. The extracted information is fed into the REM.

3. Quantification of Uncertain Information

3.1. Structure of the User Command

The command understanding ability of the system is similar to the system explained in [20]. It accepts the user commands that follow the following grammar structure:

$<$userCommand$>$ = $<$action$>$ $<$actionModifier$>$ $<$direction$>$;
$<$action$>$ = (go | move);
$<$actionModifier$>$ = (far | medium | little);
$<$direction$>$ = (forward | backward | left | right);

"actionModifer" decides the distance that has to be travelled and it is evaluated by the uncertain information understanding module (UIUM). "direction" decides the direction of the movement and the reference frame of the robot is considered for evaluating the direction. In addition to the component given in the above grammar structure, there can be redundant words in the command such as articles. If there are redundant words in a particular command, those redundant words are filtered out from the user command before parsing it in order to link the robot actions and the command. Furthermore, the system is capable of mapping the synonyms with the initial tokens of the grammar model as explained in [23].

3.2. Uncertain Information Understanding Module (UIUM)

The UIUM has been deployed in order to quantify the uncertain terms such as "little" and "far" in motional navigation commands such as "move little forward" and "move far left". It has been implemented with fuzzy logic as similar to the system explained in [20]. The inputs of the fuzzy system are action modifiers of a particular user instruction (i.e., the uncertain term) and the available free space of the room. The membership function for the input, free space, is modified according to the size of the room (S). The output is the quantified distance value of the uncertain term. The output membership function is modified according to the perceptive distance (D). The perceptive distance is decided based on the arrangement of the environment. The inputs and the output membership functions of the system are shown in Figure 2. The rule base of the system is given in Table 1. The default perceptive distance (D_r) is the distance to the object that obstructs the movement of the robot along a straight path towards the intended moving direction (this is illustrated in Figures 3 and 4).
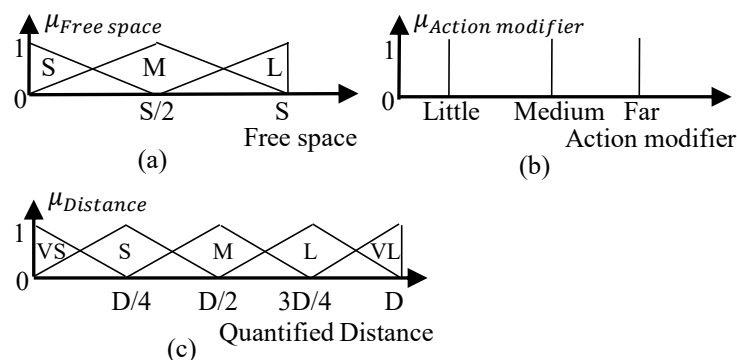

Figure 2. (**a**,**b**) represent the input membership functions of the uncertain information understanding module (UIUM). (**c**) represents the output membership function of the UIUM. The membership functions are defined similarly to the system explained in [20]. The fuzzy labels are defined as S: small; M: medium; L: large; VS: very small and VL: very large.
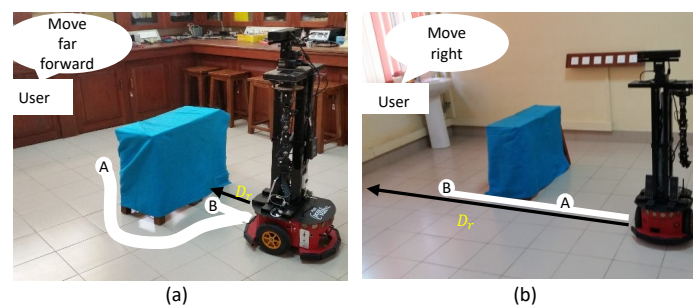

Figure 3. (**a**,**b**) show two example situations that exhibit the limitations of the method explained in [20]. The position requested by the user may be either positions 'A' or 'B'. However, the existing system considers only position 'B'. It should be noted that the annotated positions and paths are not exactly those generated from the systems and these are marked for the sake of explanation.
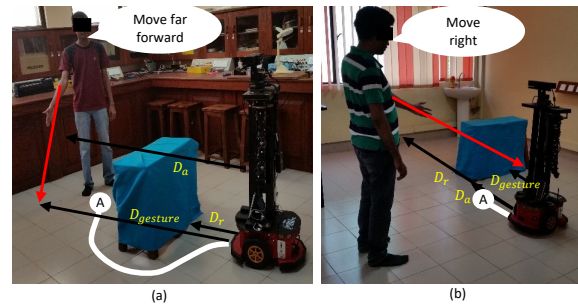
Figure 4. (**a**,**b**) show two example scenarios that explain the possibility of using pointing gestures in order to identify the intention of the user for switching the perceptive distance. It should be noted that the annotated positions, paths and vectors are not exactly those generated from the system and these are marked for the sake of explanation.

Table 1. Rule base of the fuzzy system. S: small; M: medium; L: large; VS: very small and VL: very large.

Input Memberships		Free Space		
		S	M	L
	Little	VS	S	M
Action modifier	Medium	S	M	L
	Far	M	L	VL

3.3. Motion Intention-Switching by Identifying the Actual Intention of the User

3.3.1. Rationale behind the Evaluation of Gestures Accompanied with User Instructions

The two example scenarios given in Figure 3 are considered for investigating the limitations of the system proposed in [20]. In scenario (a), the user issues the command, "move far forward". In this situation the maximum quantified output of the system proposed in [20] will be less than D_r since the perceptive distance is limited to D_r. Therefore, the robot will move to position 'B'. However, there are situations where the intention of the user is to move the robot to a position similar to location 'A' since the user expects that the robot can see beyond the obstacle. In situation (b), the user issues the command, "move right". In this situation the quantified output of the system proposed in [20] will result in a movement of the robot to location 'B'. However, there are situations where the intention of the user is to move the robot to a location similar to the location 'A' since the user expects that the robot can consider the nearby obstruction for adapting the perception. Therefore, the system proposed in [20] is not capable of understanding the intention of the user effectively.

Typically, humans combine pointing gestures with voice instructions in order to convey the idea or the intention more clearly to the peers [21,22,24]. Therefore, the information conveyed from pointing gestures is analyzed by the intention identifier module (IIM) in order to identify the intention of the user effectively. The two example scenarios given in Figure 4 are considered for the explanation of the gesture-based user intention identification process that can be used in order to overcome the above-mentioned limitations. In case (a), the user is pointing to a location that is well beyond the default perceptive distance (i.e., D_r). Therefore, if the gesture is being pointed towards a location well beyond the default perceptive distance it can be concluded that the intention of the user is to navigate the robot beyond D_r (i.e., location 'A' instead of 'B' in Figure 3a). Similarly, in case (b), if the user is pointing to a location that is well within the default perceptive distance (D_r), then it can be concluded that the intention of the user is to move the robot to an alternative position 'A' instead of position 'B' in Figure 3b.

3.3.2. Pointing-Gesture Evaluation

Skeletal information that can be retrieved from the Kinect motion sensor attached to the robot is used in order to identify the pointing gesture and to estimate the pointing position. The vector drawn

from the elbow joint to the wrist joint is considered as the direction of pointing (marked in Figure 4 using red arrows). Then, this vector is extended until the vector crosses the plane of the floor. The point where the floor plane is crossed by the elbow-wrist vector is considered as the point that is referred by the user through the gesture. Then, the horizontal distance between the referred point and the position of the robot parallel to the intended direction of motion (i.e., parallel to D_r) is calculated (i.e., marked as $D_{gesture}$ in Figure 4). In order to consider a hand posture of the user as a pointing gesture, the joint positions should not be within the ranges defined for the rest positions and the elbow-wrist vector should point towards the floor plane. Furthermore, the pointing direction should be stable and the variation should be less than an experimentally decided threshold in order to consider it as a valid pointing gesture. Furthermore, the time duration for the perceiving the user through the Kinect is set as 5 s and the perceiving is triggered with the initiation of the voice instructions. It should be noted that the system has been designed and developed for single user situations and the system is only capable of detecting the gestures of a single person. If there are multiple people in the field of view of the Kinect, the system considers only the closest person. In this stage, it would be fine to consider only single user situations since the core contribution of the work is to addresses issues in resolving spatial ambiguity in spoken commands (speech involving example phrases such as "move a little bit to the right", "go far left", etc.) by incorporating user gestures and spatial information of the environment. The situations with multi-users are not considered in the scope of the work presented in the paper and methods for handling such situations are proposed for future work.

3.3.3. Motion Intention Switcher (MIS)

The desired position for the movement cannot be directly taken as the position referred from the gesture since the point referred from the gesture is not very accurate and typically it would not be the exact location that the user wants to navigate the robot. Moreover, the gesture instructions are often useful in enhancing the meaning of vocal instructions in human–robot interaction [22,24]. Therefore, it is only used for altering the perceptive distance (D) to an alternative perceptive distance (indicated as D_a in Figure 4) from the default (i.e., D_r) by identifying the actual intention of the user. The assigning of alternative perceptive distance (D_a) for the perceptive distance (D) is done by MIS if required. The decision as to whether the perceptive distance has to be altered to an alternative (D_a) is decided based on a rule-based approach that evaluates $D_{gesture}$ and D_r.

The procedure of assigning the perceptive distance D, is given in Algorithm 1. δ_{max} and δ_{min} are scalar constants used in order to avoid the false triggering of the intention switching due to the less accurate $D_{gesture}$. The alternative perceptive distance, D_a, has two cases where the $D_a > D_r$ and $D_a < D_r$. If $D_a > D_r$, then it is considered as $D_{a,max}$ and if $D_a < D_r$, then it is considered as $D_{a,min}$. Moreover, the MIS shifts the perception of robot between the alternative and default hypotheses based on the defined thresholds that depend on the pointing gesture issued by the user and the layout of the surrounding environment.

Algorithm 1 Assigning perceptive distance (D)

INPUT: D_r, $D_{gesture}$, D_a
OUTPUT: D

if $D_{gesture} > \delta_{max} D_r$ **then**
    $D = D_{a,max}$
else if $D_{gesture} < \delta_{min} D_r$ **then**
    $D = D_{a,min}$
else
    $D = D_r$
end if

3.3.4. Estimation of Alternative Perceptive Distance (D_a)

The estimation of alternative perceptive distance, D_a is illustrated in Figure 5 considering the two possible cases where $D_{gesture} > \delta_{max} D_r$ and $D_{gesture} < \delta_{min} D_r$. A field angle of α in the intended moving direction is considered for estimating the D_a. The field angle, α, is considered as $30°$ since according to [20] the objects in that region have a higher impact for the human mobility. In case (a), D_a should be a value greater than D_r since $D_{gesture} > \delta_{max} D_r$. Therefore, $D_{a,max}$ exists and in order to estimate that, a vector parallel to the direction of the intended moving direction (i.e., the vector parallel to D_r) is extended until it reaches another obstruction for the movement inside the considered field. The magnitude of this vector is considered as $D_{a,max}$ in such cases (i.e., cases where $D_{a,max}$ is required as a result of $D_{gesture} > \delta_{max} D_r$). In case (b), D_a should be a value less than D_r since $D_{gesture} < \delta_{min} D_r$. Therefore, $D_{a,min}$ is required. The distance along a path parallel to the default intended moving path (i.e., parallel to D_r) to an obstacle within in the considered field from the robot is taken as the $D_{a,min}$ in such cases. If $\delta_{main} D_r \leq D_{gesture} \leq \delta_{max} D_r$ or a valid gesture is not detected (i.e., $D_{gesture} = null$) , the default perceptive distance (D_r) is considered as the perceptive distance (D) and hence the intention of the robot will not be switched from the default to an alternative intention in such cases.
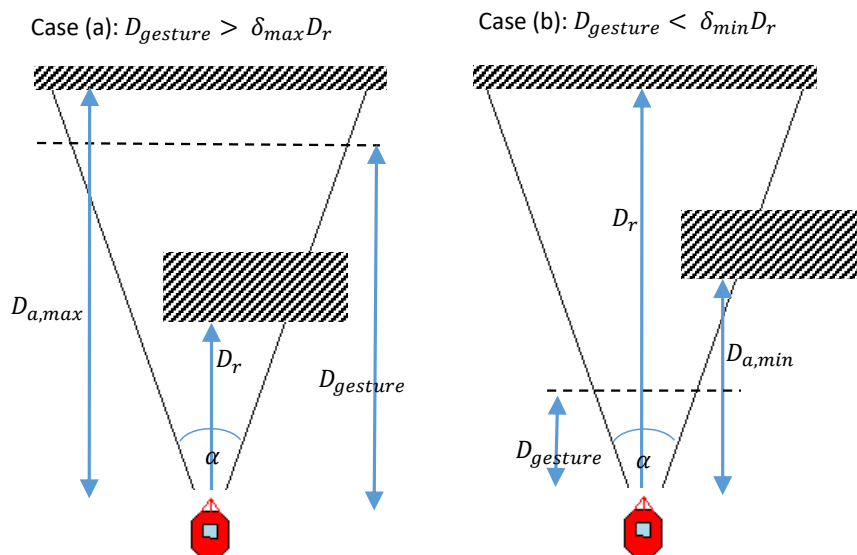

Figure 5. The ways to estimate the alternative perceptive distances are illustrated for the possible two scenarios. The shaded areas represent the obstacles/objects in the environment that are in near vicinity of the considered field view. The field angle is denoted as α. The dashed-line represents the perpendicular drawn to the intended moving path from the evaluated gesture pointing position in each scenario. $D_{gesture}$ is calculated based on the point referred by the gesture as explained in Section 3.3.2. D_r, $D_{a,min}$ and $D_{a,max}$ are computed based on the data of navigation map. This illustrates the parameter estimation considering the indented moving direction as forward. The same is applied for other directions too.

4. Results and Discussion

4.1. Experimental Setup

The proposed concept has been implemented on the MIRob platform [23] and experiments have been carried out in an artificially created domestic environment in order to validate the behavior of the proposed system in switching the perceptive distance according to the intention of the user based on the pointing gestures accompanied with verbal instructions. Furthermore, another set of experiments has been carried out in order to evaluate the performance gain of the proposed method over the work presented in [20] (i.e., system without the intention switching ability) which is not

capable of analyzing the information conveyed through gestures. The evaluation was carried out with five healthy participants (average and standard deviation of the age of the participant are 25.2 years and 1.7 years, respectively) and they were graduate students in the university. The experiments have been carried out based on the guidelines suggested in [25] for designing, planning and executing human studies for human–robot interactions in order to avoid the subjectivity of the experimental results. The scalar constants δ_{max} and δ_{min} are chosen experimentally as 1.5 and 0.75, respectively, in for achieving the desired characteristics.

4.2. Validation of the Behavior of the Motion Intention Switcher (MIS)

In order to validate the behavior of MIS in switching the intention based on the pointing gestures, experiments have been carried out in 10 different layout scenarios where such intention switching may be required in order to effectively evaluate the user instructions. Each participant was given the chance to perform the evaluation in any two of the previously unused arrangements among these 10 scenarios. The behavior of the proposed method (i.e., the system with MIS) and the system without the intention-switching ability (i.e., the system presented in [20]) have been analyzed in those situations. The sample results obtained from the experiment are given in Table 2. The views from the robot with tracked skeletons of the users in the sample cases are shown in Figure 6 along with the third person view of the scenarios. The corresponding positions of the robot during the execution of each case are marked on the map shown in Figure 7.

In case (a), the robot was initially placed on the location 'a_I' without deploying the MIS to the system. Then, the robot was commanded, "move far forward". The uncertain term in the command is "far" and the robot had to quantify the meaning of "far" for fulfilling the user command by navigating to the desired location. In this case, D_r was 33 cm since the robot only considers the immediate obstruction in its intended straight moving path. Therefore, the perceptive distance D was 33 cm and the quantified output generated from the UIUM was 29 cm, resulting a destination position in between the robot and the obstacle as explained in Section 3.3.1. Therefore, the robot moved to location 'a_B'. Then, the MIS was activated and the robot was again placed at the initial position (i.e., location 'a_I'). The robot was again commanded with the same voice instruction accompanied with a pointing gesture that expresses that the intention of the instruction is to navigate the robot to a position that is beyond the obstacle. The gesture evaluation system interpreted the gesture and calculated $D_{gesture}$ was 121 cm. In this situation, the perceptive distance was altered by MIS to alternative perceptive distance $D_{a,max}$ since $D_{gesture} > \delta_{max}D_r$. $D_{a,max}$ was evaluated as 252 cm and it was assigned to the perceptive distance (D). Therefore, the output of the UIUM was 199 cm that resulted a destination position beyond the obstacle and then robot moved to location 'a_A' by taking a curvy path generated by the navigation controller for avoiding the obstacle.

In case (b), the robot was initially placed in location 'b_I' with disabled MIS. Then it was commanded, "move medium right". The robot had to quantify the meaning of the uncertain term "medium" in order to move to the destination position requested by the user. Here, D_r was 272 cm. Subsequently, D and the quantified outputs were 272 cm and 181 cm, respectively. Therefore, the robot moved to location 'b_B' that is located well past the nearby obstacle. Then, the robot was again placed in the same initial position (i.e., 'b_I') with enabled MIS. This time the robot was commanded with the same voice instruction accompanied with a pointing gesture that expresses the intention of the user is not to move the robot to a location well past the nearby obstacle. Here, the $D_{gesture}$ was 57 cm, that lead to assigning of $D_{a,min}$ to D since $D_{gesture} < \delta_{min}D_r$. The evaluated $D_{a,min}$ was 72 cm since the robot considers the distance to the nearby obstacle in the considered field along the intended moving direction. Therefore, the quantified output was 48 cm, which resulted the movement of the robot to location 'b_A' where the robot is not required to move beyond the nearby obstacle.

In case (c), the initial position is location 'c_I' and it was commanded, "move far forward". The system without the MIS quantified the meaning of "far" as 42 cm by considering the default perceptive distance and the robot moved to location 'c_B'. The quantified output of the system with MIS was 218 cm since it considered the $D_{a,max}$ as D since the evaluated gesture indicated a request to change the default intention.

Figure 6. The view of the robot and the third person view of sample scenarios are shown with the corresponding case (**a**–**e**) given in Table 2. The tracked skeletons of the users are also superimposed with the RGB view of the robot for better clarity.
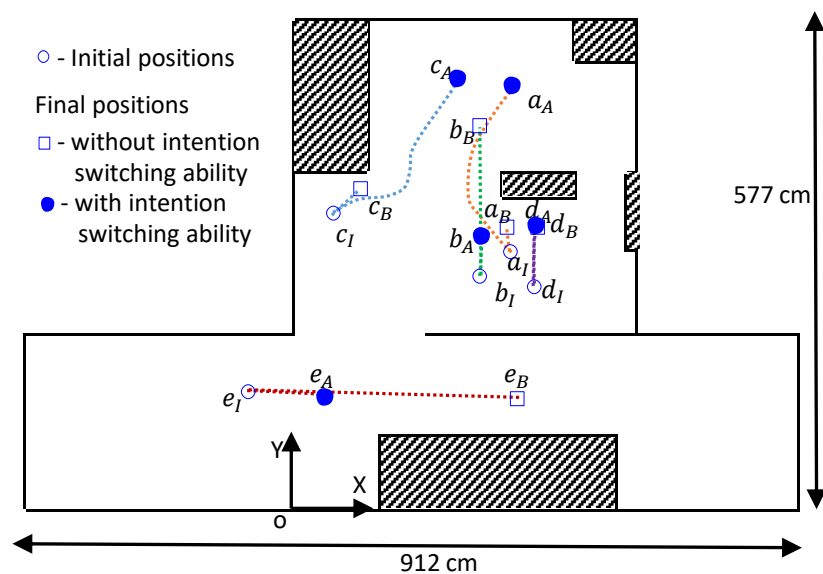
Figure 7. The initial and final positions of the robot during the experiment for identifying the behaviors of the proposed method are marked on the map with corresponding case letters. The shaded areas represent the objects in the environment. The map is drawn to a scale. However, it should be noted that the markers do not represent the actual size of the robot.

In case (d), the initial location was 'd_I' and it was commanded, "move far forward". The quantified output of the system without the MIS was 70 cm and the robot moved to location 'd_B'. In the system with the MIS case, D should be altered to $D_{a,min}$ since $D_{gesture} < \delta_{min} D_r$. However, in this situation $D_{a,min}$ and D_r were the same. Therefore, D was not altered and the quantified output is the same as the system without MIS. Therefore, the robot moved to location 'd_A' which was almost the same as 'd_B' (due to navigational errors there is a very small different in position coordinates). In this case, the intention of the user was to express his intention of navigating the robot to a location that is in between the obstacle and the robot without altering the default intention. Moreover, the proposed system is capable of successfully handling such situations.

In case (e), similarly to the case (b) the robot with MIS switched the intention by identifying the actual intention of the user by analyzing the instructions conveyed from pointing gestures given along with voice instructions. Similarly, the behavior of the MIS was found to be capable of effectively switching the intention of the robot according to the actual intention of the user in all the test cases. An explanatory video that shows the behaviors of the two systems in a similar kind of experimental scenario is provided as a supplementary material. It shows the video footage from a third person's view along with the traced location of the robot within the navigation map. Furthermore, parameters used in the interpretation process of the commands are also given with annotated explanations.

Table 2. Sample results of the experiment for validating the behaviors of the motion intention switcher (MIS).

	User Command	Initial Position [1]			Uncertain Term	Room Size (m²)	Free Space (m²)	Without MIS			Final Position [1]			With MIS				Final Position [1]		
		X	Y	θ				D_r (cm)	D (cm)	Output (cm)	X	Y	θ	D_r (cm)	$D_{gesture}$ (cm)	D (cm)	Output (cm)	X	Y	θ
a	move far forward	254	302	88	far	15.08	12.77	33	33	29	252	329	95	33	121	252	199	254	500	90
b	move medium right	220	272	179	medium	15.08	12.77	272	272	181	218	452	87	274	57	72	48	217	319	93
c	move far forward	46	344	49	far	15.08	12.77	64	64	42	78	375	50	66	180	269	218	189	509	49
d	move far forward	285	260	87	far	15.08	12.77	86	86	70	289	330	87	85	−140	85	70	283	334	88
e	move medium forward	−53	135	−5	medium	18.55	16.33	470	470	313	262	125	−3	470	100	130	86	33	127	−5

[1] It should be noted that these positions are given in (X cm, Y cm, $θ°$) format with respect to the coordinate system marked on the map shown in Figure 7. $θ$ is measured with respect to the positive X-axis in a counter-clockwise direction.

4.3. Evaluation of Performance Gain of the Proposed Method

A set of experiments has been carried out in order to compare the performance gain of the system with MIS (i.e., the proposed system) over the system without MIS (i.e., the system explained in [20]). For this experiment, the users were asked to navigate the robot from a given initial position to a given goal position marked on the floor as shown in Figure 8. The number of steps taken for navigating the robot towards the goal has been considered as the parameter for the evaluation work based on the experimental evaluation carried out in the work presented in [26]. The same task was repeated for both systems and the information related to the systems was recorded. Ten different layout arrangements (i.e., with different initial and goal positions) have been selected by randomly choosing the initial and goal positions. The initial and goal positions for a particular layout scenario have been kept within the same room since it is impractical to navigate the robot from one room to another room using only this kind of simple motion command. Furthermore, such navigation tasks could be deduced into this kind of problem by using the ability of the robot to understand a command like "move to the kitchen" as explained in [23]. All the participants have been given the chance to perform one by one in all 10 layout arrangements and the results have been analyzed in order to evaluate the value addition of the proposed MIS. It should be noted that this experimental scenarios are independent of the experimental scenarios discussed in experiment 1 (i.e., in Section 4.2).

The data of the experiments for user 1 in layout arrangement 1 (i.e., named as case 1) and user 1 in layout arrangement 2 (i.e., named as case 2) are given in Table 3 as sample results. The corresponding positions of the robot after executing each user instruction are marked on the map shown in Figure 9. The positions are annotated with the corresponding indexes given in Table 3.

Figure 8. This shows the experimental scenario of the case 1 of the experiment for comparing the performance of the system with the MIS and the system without the MIS. The user was asked to navigate the robot to the goal position marked on the floor by implementing both the system in the robot. The goal area is annotated as "goal" in here.
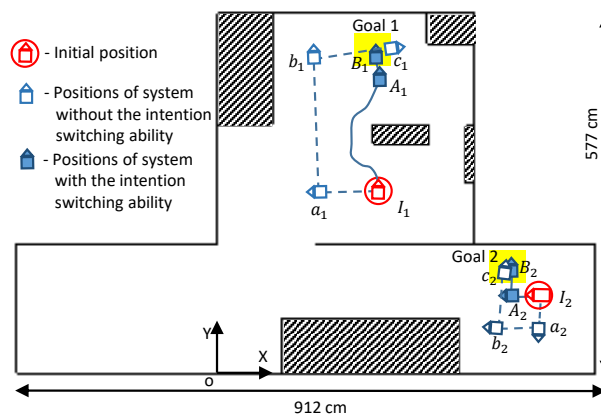

Figure 9. The positions of the robot after executing each user instructions given in Table 3 for cases 1 and 2 are marked on the map with the corresponding indexes. The shaded areas represents the objects in the environment. The light color solid areas represent the positions of the goals. The map is drawn to a scale. However, it should be noted that the markers do not represent the actual size of the robot.

Table 3. Sample results of the experiment for evaluating the performance gain of the system with the motion intention switcher (MIS).

	User Command	Uncertain Term	Room Size (m^2)	Free Space (m^2)	D_r (cm)	$D_{gesture}$ (cm)	Intention Switched	D (cm)	Distance Moved (cm)	Position [1] (X, Y, θ)
Case 1	Initial position [1]									I_1 (247, 283, 89)
with MIS	A. move medium forward	medium	15.08	12.77	57	128	True	275	183	A_1 (250, 466, 89)
	B. move little forward	little	15.08	12.77	87	Not detected	False	87	36	B_1 (249, 502, 89)
without MIS	a. move little left	little	15.08	12.77	206	-	-	206	86	a_1 (160, 283, −179)
	b. move far right	far	15.08	12.77	270	-	-	270	219	b_1 (149, 502, 92)
	c. move medium right	medium	15.08	12.77	149	-	-	179	117	c_1 (270, 519, 8)
Case 2	Initial position [1]									I_2 (504, 117, 179)
with MIS	A. move little forward	little	18.55	16.33	470	60	True	102	42	A_2 (462, 118, 179)
	B. move medium right	medium	18.55	16.33	63	Not detected	False	63	42	B_2 (460, 159, 87)
without MIS	a. move medium left	medium	18.55	16.33	98	-	-	98	64	a_2 (504, 57, −89)
	b. move medium right	medium	18.55	16.33	106	-	-	106	71	b_2 (434, 68, 175)
	c. move far right	far	18.55	16.33	110	-	-	110	89	c_3 (447, 152, 83)

[1] It should be noted that these positions are given in (X cm, Y cm, $\theta°$) format with respect to the coordinate system marked on the map shown in Figure 9. θ is measured with respect to the positive X-axis in a counter-clockwise direction.

In this case, the initial position of the robot was 'I_1' and the goal position is annotated as 'goal 1' in the map. In the system with the MIS event, first the robot was commanded, "move medium forward" while being shown a gesture that expresses the requirement of switching the intention to navigate the robot beyond the obstacle in the front. D_r and $D_{gesture}$ were 57 cm and 128 cm, respectively. The intention of the robot was switched by the MIS since $D_{gesture} > \delta_{max}D_r$ and $D_{a,max}$ was assigned to the perceptive distance (D). Therefore, D was 275 cm and subsequently the quantified distance output was 183 cm which resulted the movement of the robot to location 'A_1'. Then the robot was commanded, "move little forward" and a pointing gesture was not detected by the system since a pointing gesture was not issued by the user. Therefore, the intention of the robot was not switched and the robot moved 36 cm by considering D_r as perceptive distance (D). The moved position was 'B_1' that was inside the given goal area. Therefore, this was considered as the completion of the task. Then, the robot was placed on the same initial position (i.e., 'I_1') after disabling the MIS (i.e., system similar to [20]) and again the user was asked to navigate the robot to the goal. In this event, if the user had commanded the robot "move medium forward" similar to the earlier event, the robot would have moved to a point between the obstacle and the robot (due to the limitation of the system without MIS discussed in Section 3.3.1). However, that movement would be a waste since the user cannot navigate the robot beyond the obstacle without changing the moving direction. Therefore, with this in mind, the user first issued the command "move little left" in order to take away the robot from the barrier. The robot quantified the distance meant by "little" as 86 cm by considering the default perceptive distance and moved to position 'a_1'. Then the robot was commanded "move far right" and robot moved to position 'b_1' in order to fulfill the request of the user. Then, the robot was commanded "move medium right" and the robot moved to position 'c_1' which was inside the goal area. Therefore, the task was completed. In order to complete the task with the system with the MIS, the user had to issue only two user instructions while with the system without the MIS, the user had to issues three instructions in order to complete the tasks. Moreover, the work overhead of the user is comparative less when the MIS is deployed into the robot.

In case 2, the initial position of the robot was 'I_2' and the goal is annotated as 'goal 2' in the map. In the system with the MIS event, the user first issued the command "move little forward" accompanied with a pointing gesture that express the requirement for the intention switching. If such a gesture had not been issued, the robot would have moved to a location that is well past the nearby table. Therefore, the robot moved to position 'A_2' by switching the perception to the alternative perception. Then the robot was commanded, "move medium right" without giving a pointing gesture. Therefore, the robot moved to position 'B_2' considering the default intention. Therefore, the task was completed with 2 user instructions. In the event of the system without the MIS, first the command, "move medium left" was issued by the user and the robot moved to location 'a_2'. If the command "move little forward" had been issued in this case, the robot would have moved to a location that is well past the intended moving position due to the limitation of the system (without MIS) and the user already knew this from his past experience. That is the reason for issuing the command "move medium right" instead of "move little forward" similar to the system with the MIS case. Then with the next voice instruction, the robot moved to position 'b_2'. After the next instruction, the robot moved to 'c_2' that is inside the goal area . Therefore, in order to navigate the robot in this situation, three user instructions were required which is higher than for the event with the MIS.

Similarly, the experiments have been carried out in all the layout arrangements by all the participants. The average number of steps required for fulfilling the navigation task in each layout arrangement for the system with the MIS and without the MIS is given in the graph shown in Figure 10.
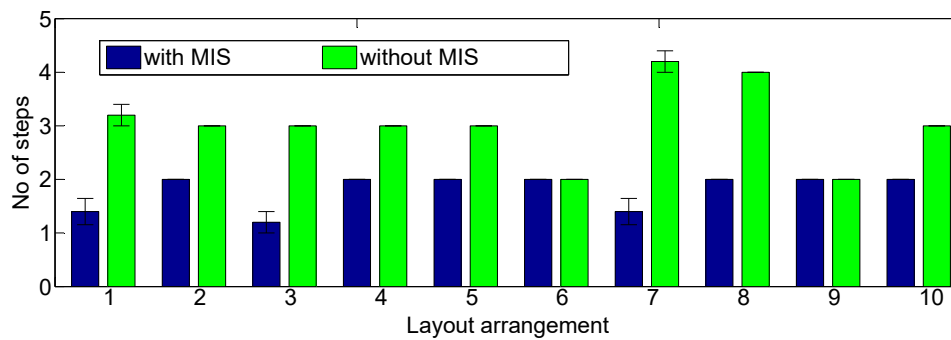
Figure 10. This graph shows the average number of steps/instructions taken in order to navigate the robot to the goal positions in different experimental layout arrangements during the experiment for evaluating the performance gain of the proposed MIS. The error bars represent the standard error.

In all the layout arrangements except 6 and 9, the robot with the MIS was able to be navigated to the goal positions with a fewer number of voice instructions compared to the robot without the MIS and the difference is statistically significant ($p < 0.05$) according to the results of two sample t tests. Moreover, the system with MIS has better abilities in understanding the intention of the user over the system without the MIS. Therefore, the deployment of the MIS enhances the evaluating ability of the ambiguous language instructions by the robot. However, in layout arrangements 6 and 9, the number of steps taken by both the system are the same. The reason behind this was in those two arrangements, the ability of the MIS was not required and the robot was navigated without switching the perception from the default perception. In all other layout arrangements, the intention of the robot was changed only once in each case which leads to a reduction of required total number of steps. Therefore, the number of user instructions or steps required to navigate the robot to a desired location in this kind of situations can be reduced by deploying the MIS. Even though the step number reduction in this kind of task is small (about 1–3 steps), a robot that is used as supportive aid in a caring facility such as a nursing home would be required to perform this sort of navigation task a large number of times per a day and hence there would be a noticeable reduction of the work load in real-world applications. Moreover, this validates the potential of the MIS in enhancing the human-friendliness of the robot and interpretation of ambiguous voice instructions.

Furthermore, a user study has been carried out similarly to the performance analysis carried out in [2,27]. In here the participants were asked to rate the ability of robot based on the effectiveness of interpreting uncertain information in user instructions with MIS (a system similar to [20]) and without MIS (the system proposed in this paper) situations on a scale of 0 to 10 as similar to the evaluation approach of the work presented in [2]. The mean values obtained from this user rating the two systems are given in the graph shown in Figure 11 with the standard error bars. The ratings for the system with MIS and without the MIS are 8.0 and 6.4, respectively. According to the two sample t test, the system with the MIS has a statistically significant ($p < 0.05$) higher rating than the system without the MIS. Therefore, this validates the enhancement in uncertain information interpretation ability of the proposed concept. Moreover, these results validate the potential of the MIS in enhancing the interpretation of voice instructions with uncertain information and subsequently the improvement of the human-friendliness of the robot.
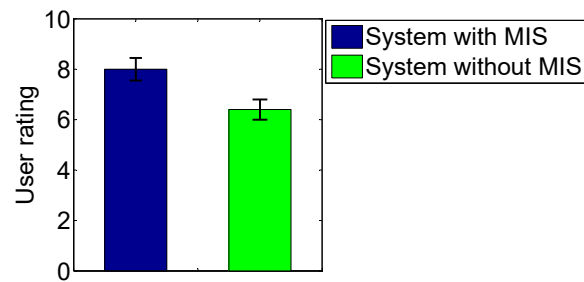
Figure 11. This graph shows the mean values of the user rating for the effectiveness of uncertain information evaluation of the systems in the two cases: system with MIS and without MIS. The error bars represent the standard error.

5. Conclusions

A method has been introduced in order to enhance the effectiveness of interpretation of verbal instructions with uncertain information such as "move far forward" by identifying the actual intention of the user. The ability for effectively interpreting such voice instructions by a service robot is useful in accomplishing typical daily activities and human–robot collaborative tasks that involve navigation of the robot. Therefore, the proposed method will improve the abilities of human-friendly service robots.

The main improvement of the proposed method over the existing approaches is that the system is capable of switching the intention of the robot by identifying the actual intention of the user. The actual intention of the user is identified by analyzing the information conveyed from pointing gestures that can be accompanied with voice instructions. Moreover, the interaction ability has been improved by integrating multimodal interaction ability in order to guess the intention of the user for improved interpretation of uncertain information in user instructions.

The intention of the robot is switched by the proposed motion intention switcher (MIS) by altering the perceptive distance from the default to an alternative. The position referred from the pointing gesture and the arrangement of the environment in that scenario are analyzed by the MIS in order to decide the alternative perceptive distance. Moreover the MIS shifts the perception of the robot between the default and the alternative hypotheses based on a set of predefined rules. It would be interesting for future work to consider a probabilistic approach instead of this rule-based approach for intention switching.

Experiments have been carried out in an artificially-created domestic environment in order to analyze the behavior of the the proposed MIS. The behavior of the MIS has been found to be effective according to the experimental results. Furthermore, experiments have been carried out in order to evaluate the performance gain of the proposed concept. The experimental results validates the potential of the proposed concept in enhancing the human-friendliness of service robots by effective interpretation of ambiguous voice instructions.

Author Contributions: M. A. Viraj J. Muthugala and A. G. Buddhika P. Jayasekara conceived and designed the proposed concept and experiments; M. A. Viraj J. Muthugala and P. H. D. Arjuna S. Srimal performed the experiments; M. A. Viraj J. Muthugala and A. G. Buddhika P. Jayasekara analyzed the data; P. H. D. Arjuna S. Srimal contributed to designing and implementing the gesture evaluation module; M. A. Viraj J. Muthugala and A. G. Buddhika P. Jayasekara wrote the paper.

References

1. Arkin, R.C. *Behavior-Based Robotics*; MIT Press: London, UK, 1998.
2. Jayawardena, C.; Kuo, I.; Broadbent, E.; MacDonald, B.A. Socially Assistive Robot HealthBot: Design, Implementation, and Field Trials. *IEEE Syst. J.* **2016**, *10*, 1056–1067.
3. Chu, M.T.; Khosla, R.; Khaksar, S.M.S.; Nguyen, K. Service Innovation through Social Robot Engagement to Improve Dementia Care Quality. *Assist. Technol.* **2016**, *29*, 8–18.
4. Fischinger, D.; Einramhof, P.; Papoutsakis, K.; Wohlkinger, W.; Mayer, P.; Panek, P.; Hofmann, S.; Koertner, T.; Weiss, A.; Argyros, A.; et al. Hobbit, a care robot supporting independent living at home: First prototype and lessons learned. *Robot. Auton. Syst.* **2016**, *75*, 60–78.
5. Johnson, D.O.; Cuijpers, R.H.; Juola, J.F.; Torta, E.; Simonov, M.; Frisiello, A.; Bazzani, M.; Yan, W.; Weber, C.; Wermter, S.; et al. Socially Assistive Robots: A comprehensive approach to extending independent living. *Int. J. Soc. Robot.* **2014**, *6*, 195–211.
6. Smarr, C.A.; Mitzner, T.L.; Beer, J.M.; Prakash, A.; Chen, T.L.; Kemp, C.C.; Rogers, W.A. Domestic robots for older adults: Attitudes, preferences, and potential. *Int. J. Soc. Robot.* **2014**, *6*, 229–247.
7. Kleanthous, S.; Christophorou, C.; Tsiourti, C.; Dantas, C.; Wintjens, R.; Samaras, G.; Christodoulou, E. Analysis of Elderly Users' Preferences and Expectations on Service Robot's Personality, Appearance and Interaction. In Proceedings of the International Conference on Human Aspects of IT for the Aged Population, Toronto, ON, Canada, 17–22 July 2016; Springer: Berlin, Germany, 2016; pp. 35–44.
8. Wang, N.; Broz, F.; Di Nuovo, A.; Belpaeme, T.; Cangelosi, A. A user-centric design of service robots speech interface for the elderly. In *Recent Advances in Nonlinear Speech Processing*; Springer: Berlin, Germany, 2016; pp. 275–283.
9. Tellex, S.; Kollar, T.; Dickerson, S.; Walter, M.R.; Banerjee, A.G.; Teller, S.; Roy, N. Understanding natural language commands for robotic navigation and mobile manipulation. In Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 7–11 August 2011; AAAI Press: Palo Alto, CA, USA, 2011; pp. 1507–1514.
10. Hemachandra, S.; Duvallet, F.; Howard, T.M.; Roy, N.; Stentz, A.; Walter, M.R. Learning models for following natural language directions in unknown environments. In Proceedings of the 2015 IEEE International Conference on Robotics and Automation (ICRA), Seattle, WA, USA, 26–30 May 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 5608–5615.
11. Skubic, M.; Perzanowski, D.; Blisard, S.; Schultz, A.; Adams, W.; Bugajska, M.; Brock, D. Spatial language for human-robot dialogs. *IEEE Trans. Syst. Man Cybern. C (Appl. Rev.)* **2004**, *34*, 154–167.
12. Kawamura, K.; Bagchi, S.; Park, T. An intelligent robotic aid system for human services. In *NASA Conference Publication*; NASA: Washington, DC, USA, 1994; pp. 413–413.
13. Pulasinghe, K.; Watanabe, K.; Izumi, K.; Kiguchi, K. Modular fuzzy-neuro controller driven by spoken language commands. *IEEE Trans. Syst. Man Cybern. B* **2004**, *34*, 293–302.
14. Jayawardena, C.; Watanabe, K.; Izumi, K. Controlling a robot manipulator with fuzzy voice commands using a probabilistic neural network. *Neural Comput. Appl.* **2007**, *16*, 155–166.
15. Jayasekara, A.G.B.P.; Watanabe, K.; Kiguchi, K.; Izumi, K. Interpreting Fuzzy Linguistic Information by Acquiring Robot's Experience Based on Internal Rehearsal. *J. Syst. Des. Dyn.* **2010**, *4*, 297–313.
16. Lin, C.T.; Kan, M.C. Adaptive fuzzy command acquisition with reinforcement learning. *IEEE Trans. Fuzzy Syst.* **1998**, *6*, 102–121.
17. Jayasekara, A.G.B.P.; Watanabe, K.; Kiguchi, K.; Izumi, K. Interpretation of fuzzy voice commands for robots based on vocal cues guided by user's willingness. In Proceedings of the 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems, Taipei, Taiwan, 18–22 October 2010; pp. 778–783.
18. Jayasekara, A.G.B.P.; Watanabe, K.; Izumi, K. Understanding user commands by evaluating fuzzy linguistic information based on visual attention. *Artif. Life Robot.* **2009**, *14*, 48–52.
19. Schiffer, S.; Ferrein, A.; Lakemeyer, G. Reasoning with qualitative positional information for domestic domains in the situation calculus. *J. Intell. Robot. Syst.* **2012**, *66*, 273–300.
20. Muthugala, M.A.V.J.; Jayasekara, A.G.B.P. Interpretation of Uncertain Information in Mobile Service Robots by Analyzing Surrounding Spatial Arrangement Based on Occupied Density Variation. In Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Daejeon, Korea, 9–14 October 2016; pp. 1517–1523.

21. Matuszek, C.; Bo, L.; Zettlemoyer, L.; Fox, D. Learning from Unscripted Deictic Gesture and Language for Human-Robot Interactions. In Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence (AAAI), Québec City, QC, Canada, 27–31 July 2014; pp. 2556–2563.
22. Whitney, D.; Eldon, M.; Oberlin, J.; Tellex, S. Interpreting multimodal referring expressions in real time. In Proceedings of the 2016 IEEE Intenational Conferece on Robotics and Automation (ICRA), Stockholm, Sweden, 16–21 May 2016; pp. 3331–3338.
23. Muthugala, M.A.V.J.; Jayasekara, A.G.B.P. MIRob: An intelligent service robot that learns from interactive discussions while handling uncertain information in user instructions. In Proceedings of the 2016 Moratuwa Engineering Research Conference (MERCon), Moratuwa, Sri Lanka, 5–6 April 2016; pp. 397–402.
24. Mavridis, N. A review of verbal and non-verbal human–robot interactive communication. *Robot. Auton. Syst.* **2015**, *63*, 22–35.
25. Bethel, C.L.; Murphy, R.R. Review of human studies methods in HRI and recommendations. *Int. J. Soc. Robot.* **2010**, *2*, 347–359.
26. Jayasekara, A.B.P.; Watanabe, K.; Habib, M.K.; Izumi, K. Visual evaluation and fuzzy voice commands for controlling a robot manipulator. *Int. J. Mech. Manuf. Syst.* **2010**, *3*, 244–260.
27. Lee, M.K.; Forlizzi, J.; Kiesler, S.; Rybski, P.; Antanitis, J.; Savetsila, S. Personalization in HRI: A longitudinal field experiment. In Proceedings of the 2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Boston, MA, USA, 5–8 March 2012; pp. 319–326.