Naturally Occurring Gestures in a Human-Robot Teaching Scenario

Nuno Otero, Chrystopher L. Nehaniv, Dag Syrdal and Kerstin Dautenhahn

Abstract— **This paper describes our general framework for the investigation of how human gestures can be used to facilitate the interaction and communication between humans and robots. More specifically, a study was carried out to reveal which "naturally occurring" gestures can be observed in a scenario where users had to explain to a robot how to perform a specific home task. The study followed a within-subjects design where ten participants had to demonstrate how to lay a table for two people using two different methods for their explanation: utilizing only gestures or gestures and speech. The experiments also served to validate a new coding scheme for human gestures in human-robot interaction, with good inter-rater reliability. Moreover, annotated video corpus was produced and characteristics such as frequency, duration, and co-occurrence of the different gestural classes have been gathered in order to capture requirements for the designers of HRI systems. The results regarding the frequencies of the different gestural types suggest an interaction between the order of presentation of the two methods and the actual type of gestures produced. Moreover, the results also suggest that there might be an interaction between the type of task and the type of gestures produced.**

Index Terms — **Human-Robot Interaction, gestures, classification system, observation study**

I. Introduction

Human-Robot Interaction (HRI) is regarding its conceptual and theoretical foundations still a recent research field. Kiesler and Hinds [1] consider that the study of design alternatives to facilitate Human-Robot Interaction is a new focus of Human-Computer Interaction (HCI). However, some researchers consider that HRI will probably need to develop its own discipline specific methods due to the embodied nature of interaction with robots [1-4]. A lot of ground work needs to be done in HRI in order to establish conceptual and theoretical foundations e.g. fleshing out to what extent results from human-human interaction studies can be translated to HRI studies. Also, experimental frameworks and methodologies need to be adapted from other fields and/or newly developed for human-robot interaction.

Dautenhahn [5] points out that the idea of agents being able to interact with humans in a "natural" way is considered attractive. As robots start acting in human environments issues of agency, believability and sociality become very important. Robots that inhabit human social spaces will need to be designed to conform as much as possible to human expectations. Fong, Nourbakhsh and Dautenhahn [6] state that the design of sociable robots needs input from research concerning social learning and imitation, gesture and natural language communication, emotion and recognition of interaction patterns. Moreover, according to the authors, three primary types of dialogue are crucial to foster the robots' abilities to interact with humans: low level interaction mechanisms, non-verbal communication, and natural language.

This paper describes our investigation regarding how human gestures can be used to facilitate the interaction and communication between humans and robots. More specifically, a study was carried out to illuminate which "naturally occurring" gestures can be observed in a scenario where users had to explain to a robot how to perform a specific home task. "Natural" refers to a situation where no scripts or pre-defined gestures to use were given.

This work is being developed as part of the research for the European funded project, "The Cognitive Robot Companion" (COGNIRON). Within the COGNIRON project, the line of research described here is part of the overall goal to capture requirements for contextual interpretation of body postures and human activities for purposes of human-robot interaction. Results from this work inform research by other COGNIRON partners into developing computational algorithms for detecting and recognizing human activities, including body postures and gestures (see [7] for an overview of the on-going investigation collaboration). In order to capture these requirements, a coding scheme to classify gestures people produce was developed. The coding scheme is an essential part of our strategy to systematically study the frequency, duration and sequence of different gestures in people's task

The work described in this paper was conducted within the EU Integrated Project COGNIRON ('The Cognitive Robot Companion' - www.cogniron.org) and was funded by the European Commission Division FP6-IST Future and Emerging Technologies under Contract FP6-002020.

All authors are with Adaptive Systems Research Group, School of Computer Science, and Science and Technology Research Institute, University of Hertfordshire Hatfield Herts AL10 9AB, United Kingdom (emails: (N.R.Otero, C.L.Nehaniv, D.S.Syrdal, K.Dautenhahn)herts.ac.uk). For further correspondece phone: +44 1707281032; e-mail: N.R.Otero@herts.ac.uk.

1-4244-0565-3/06/$20.00 ©2006 IEEE.

demonstrations. The classification system and corresponding coding scheme are in line with the conceptual framework developed by Nehaniv et al. [4].

The rest of this paper is structured as follows. First, we will summarize relevant research concerning the role of gestures in the interaction process with a particular emphasis on how the results from human-human interaction informed (or can inform) the development of robots and other computational artifacts. Secondly, the current exploratory study will be described. The third part proceeds with the presentation of the results. Finally, a general discussion and topics for future research conclude the paper.

I. Background

Our present research focus can be described as an investigation of the specificities of gestures for interacting with robots. As a starting point we are considering the following points: How do humans use gestures in their communication processes? How should we proceed to find design solutions for robots that take advantage of these human abilities? Let us turn our attention to the first question.

A. Gestures in Human-Human Interaction: Brief summary

Gestures are closely linked with the accompanying speech in terms of timing, meaning and communicative function [see, for example, 8, 9, 10]. Adam Kendon's and David McNeill's work on the study of human communicative gestures are considered to be landmarks in the field [for an overview see, 9, 10]. McNeill [9] takes a very restricted definition of gestures as he considers them to be only the non-manipulative hand/arm movements that occur during speech. Kendon [10], however, considers it difficult to specify what kinds of body movements (in a broad sense) should be called gestures. Nevertheless, he presents boundaries for what he considers to be gestures: "...only actions that are treated by co-participants interaction as part of what a person meant to say will be included: conventional gestures, gesticulations, and signing are included, but posture shifts, self-touchings, and incidental object manipulations are not." [10, pag. 110]. Kendon's proposal defines a broader set than the strict definition put forward by McNeill. In fact, McNeill [9] refers to Kendon's definition as the *Kendon's continuum*: Gesticulation → Language-like Gestures → Pantomimes → Emblems → Sign Languages. According to McNeill [9]:

"As we move from left to right: (1) the obligatory presence of speech declines, (2) the presence of language properties increases, and (3) idiosyncratic gestures are replaced by socially regulated signs." (pag. 37).

Gestures can be classified into distinct types, and different classificatory systems do exist. A review of the different classification systems is beyond the scope of this work. For a starting point into the topic see, for example, McNeill [9]; Kendon [9], Cassell [8], Kipp [11]. Research on the function of gestures has been diverse. Gestures have been studied in relation to child development and language acquisition [12, 13], or teaching and learning strategies [14-16]. Research in relation to problem solving has found not only that gestures can convey specific information and reveal thoughts not revealed in speech but also that observing gestures can be a useful extra to speech when trying to uncover cognitive processes [17, 18]. However, research investigating the function of gestures in relation to speech has produced contrary results or divergent opinions. For example, some authors defend the importance of gestures semantics independently from speech [10, 19] while others consider the primary function of gestures is not to convey semantic information [20]. Lozano and Tversky [21] argue that gestures might be of benefit to both communicators and recipients. The authors report two distinct studies to show how gestures and speech combine. More specifically, they investigated (a) if the same type of gestures helps communicators and recipients, (b) why do gestures help, (c) are gestures helping indirectly by enhancing the quality of the speech or do gestures convey specific semantic content? Their experimental work shows that gestures do help the communicators by giving valuable motor knowledge and experience. In relation to benefits for the recipients, it seems that when exposed to a demonstration that only used gestures (without speech) people were more aware of the actual actions crucial for the accomplishment of the task (which was the assembly of a TV cart). The authors suggest that, by watching the demonstration, the gestures might have provided the participants with perceptual knowledge and motor experience that was important for their own learning about the task.

B. Gestures in Human-Robot Interaction

The use of gestures in the communication loop between humans and computer artifacts has been explored. A simplified overview of research on multimodal interfaces and human-robot interaction seems to indicate the following approaches for the inclusion of gestures in the interaction process:

- Some research has explored the use of sets of pre-defined gestures and speech to communicate with robots [22-24].
- Other researchers consider that the use of gestures for communication with computer artifacts can and should be explored beyond the confines of a set of pre-defined gestures [4, 8, 11, 25-27].

Robots may need to recognize human gestures and movements to infer limited intent regarding these human gestures and movements, but also to communicate their own internal state and plans to humans [4]. Nehaniv et al. [4] analyze the specificities of analyzing and incorporating gestures in the interaction loop between humans and robots and argue for the need to consider a broad definition of

gestures, the reason being to avoid:

"...inherent limitations of approaches working with a too narrow notion of gesture, excluding entire classes of human gesture that should eventually be accessible to interactive robots able to function well in a human social environment." (Nehaniv et al.[4], pag. 372).

Therefore we need to adopt a broad notion of gesture. To illustrate the point made, take the example of manipulative gestures which would not be considered according to narrow definitions of gesture. These gestures involve the manipulation of objects (usually without concomitant speech) and seem central to HRI since a robot, at times, will have to recognize or even to co-ordinate object manipulations with humans. Thus, we decided in our work to include hand/arm movements that concern the exchange of objects between partners involved in interaction as a type of gesture. Nehaniv et al. [4] propose the following five classes of gestures:

- *Irrelevant* - these are gestures that do not have a primary communicative or interactive function (e.g. adjusting ones hair or rubbing the eye).
- *Manipulative gestures* - these are gestures that involve the displacement of objects (e.g. picking a cup).
- *Side Effect of Expressive Behavior* - these are gestures that occur as side-effects of people communicative behavior. They can be motion with hands, arms, face etc but without specific interactive, communicative, symbolic or referential roles.
- *Symbolic gestures* - these are gestures that follow a conventionalized signal. Their recognition is highly dependent on the context, both current task and cultural milieu (e.g. the thumbs up or the ring to convey "OK").
- *Interactional Gestures* - this category classifies gestures used to regulate interaction with a partner. Thus are can be used to initiate, maintain, invite, synchronize, organize or terminate an interaction behavior between agents (e.g. head nodding, hand gestures to encourage the communicator to continue).
- *Referencing/pointing gestures* - the gestures that fall into this category are gestures used to indicate objects or loci of interest.

Nehaniv et al. [4] stress the importance of knowing the context in which gestures are produced since this is crucial to disambiguate meaning. They suggest that data on the *interaction history* and *context* may help the classification process. The authors also point out the need to consider to whom or what the gesture targeted (identify *target*) and who, if anyone, is supposed to see it (identify the *recipients*). Certain gestures in particular situations might be multipurpose [4]. For example, a gesture of bringing an object conspicuously toward an interaction partner is manipulative but it may also be classified as interactional since it might comprise a solicitation for the partner to take the object.

C. The research questions

The following questions frame our investigation:

A. What is the level of inter-rater agreement in the coding of the video recordings?
B. Are there any differences regarding the frequency and/or duration of the different types of gestures produced when people are asked to use gestures only or are allowed to gesture and speak to explain a routine home task?
C. The previous point, B, implies two different methods of explaining, using gestures only and gestures and speech. If people are asked to used both methods one after the other (basically following a within subjects design) is it possible that any differences concerning B) can be due to the effect of the order of presentation? In other words, is it possible that by starting to use gestures only to explain the task will affect the way one will explain the task using gestures and speech afterwards (or the other way around)?
D. Again, considering a within subjects design, are there any differences concerning B) due to the effect of fatigue on the production of gestures?
E. What is the distribution of co-occurrence of different types of gestures?
F. What method of explanation (gestures only or gestures and speech) did the participants' prefer?
G. Which method (gestures only or gestures and speech) did they think produced the better explanation?

II. METHODOLOGY

A. The Design

This exploratory study followed a within-subjects design. The participants had to demonstrate how to lay a table for two people utilizing two different methods: using only gestures or gestures and speech - these were the two experimental conditions. The particular task was chosen since it is considered a relevant task for a robot companion in the home. Considering that there were two conditions, the order of appearance was counterbalanced to try to cancel effects of order. Thus we had 2 arrangements of the two conditions: five participants started by using gestures only and the other five started by explaining the task using gestures and speech.

B. The Participants

The sample consisted of 6 female and 4 male subjects, all from our university. For this exploratory study the number of participants was considered suitable in order to provide input for future studies involving larger numbers of subjects. The participants' occupations are: two researchers (one from Computer Science and one from Psychology), six PhD students (from different technical subjects and also social

sciences) and two members of the administrative staff.

C. The Materials/Apparatus

At the beginning of the session, the participants had to fill in a pre-session questionnaire that consisted of a consent form and some questions regarding demographics and general issues regarding their familiarity with robots. After the main task, a post-session questionnaire was given to tap into the participants' experiences concerning the task they performed. The questions covered the following issues: (a) to what extent they could see themselves doing something similar in the real life, (b) what method of demonstration did they prefer, (c) which method they thought made them think harder and to what extent that was helpful for their demonstration, (d) which method of explanation they thought produced the best explanation, (e) to what extent did they plan their actions, (f) to what extent were they willing to learn sets of predefined gestures or words to communicate with a robot.

The following additional materials were also utilized: (a) two sets of cup, plate, and cutlery in order to simulate laying a table for two persons and (b) a video-recording camera to capture the participants' activities. The camera being used was representative of a camera mounted on a robot which is observing and analyzing the subjects' behavior. Since no feedback from the robot to the subject was considered, it was not deemed necessary to involve an actual robot in the experiment, although, as our previous work has shown [28] the actual appearance of a robot might influence how subjects teach a robot.

D. The coding scheme to classify the participants' gestures

The coding scheme follows Nehaniv et al. [4]. The categories formulated by the authors were extended and specified further into some sub-categories to facilitate the coding of the video recordings. Furthermore the following coding heuristics were developed:

- Eye gaze: only code eye gaze when there is informational value for the interactional gestures category.
- Symbolic gestures: if the episode shows more than one gesture symbolization choose the one you consider more important for the episode and make comment regarding the other. Coding a gesture that performs an action involves the choice between symbolic or manipulative categories. The coder may need to see what the following gesture is and also evaluate to what extent the gestural action is used symbolically from the context.
- Interactional gestures: whenever two similar events follow each other consecutively code as one long episode.

The coding scheme was implemented in the video-coding software ANVIL [11], and this tool was used for the coding of the video-recordings of the participants demonstrations.

E. The procedure

1) The Physical Setting

Two rectangular tables were positioned perpendicular to each other. The participants would start their demonstration behind one, in front of a video camera. The different objects were placed on one table and the participants had to lay the table for two people on the other table.

2) Spoken Instructions and Sequencing of Activities

The instructions followed a three step script:

- Beginning of the experiment: *instruction* - "This experiment has two parts. In the first part you will be asked to demonstrate how to accomplish a certain task in front of a video camera. The data collected will be used to train a robot in understanding human explanations of home tasks. In the final part, we will ask you to complete a questionnaire concerning your experience with our experiment." The nature of the different materials in the room was explained to the participant and he/she was asked to fill the pre-session questionnaire.
- The main experimental task: *instruction* - "Your performance is not being assessed and there is no right or wrong way of doing things. There are two tasks. The first one is to pick the plates and cutlery from the first table and lay the table in the second one. The second task is to put plates and the cutlery back into the first table. Only one object at a time can be manipulated." Depending on the conditions we explained that only gestures or speech and gestures could be used in the demonstration: (a) *gestures only condition* - "For this demonstration only the visual system of the robot will be used. Hence, you should only use gestures" and (b) *speech and gestures condition* - "For this demonstration both the visual and speech recognition systems of the robot will be used. So feel free to use gestures and speech".
- Finally, the participants were asked to complete the post-session questionnaire.

III. Results

A. Inter-rater agreement

Categories	Cases	% of Agreement	Kappa	Sig.
Irrelevant	1113	99	.27	$p < .000$
Manipulative	1113	98	.95	$p < .000$
Side effect exp.	1113	98	.59	$p < .000$
Symbolic	1113	99	.60	$p < .000$
Interactional	1113	92	.67	$p < .000$
Pointing	1113	99	.93	$p < .000$

Table 1 Intercoder Agreement for the Coding Scheme. Number of cases under analysis, percentage of agreement between the coders, the kappa statistic and its level of significance

Table 1 shows the inter-rater agreement statistics

including the Cohen's kappa for six distinct main categories used for the codification of the video recordings.

The observation of agreement concerns second by second agreement between two coders with an error of plus or minus one second. The number of cases corresponds to the sum of all the subjects' video recordings taken together second by second. The results from Table 1 show that the agreement reached is good except on the irrelevant gestures category.

B. Are there any differences regarding the frequency and/or duration of the different types of gestures between the two experimental conditions (gestures only or gestures and speech)?

Table 2 shows the descriptive statistics for all the categories concerning the frequencies of occurrence per experimental condition (*Gestures only* or *Gestures and Speech*).

Conditions	Categories	N	Mean	SD	Min.	Max.
Gestures	Pointing	10	.20	.2	0	1
	Interactional	10	7.60	5.15	2	20
	Irrelevant	10	.40	.70	0	2
	Manipulative	10	20.00	9.03	11	39
	Side effect expr.	10	.20	.42	0	1
	Symbolic	10	.20	.42	0	1
Gestures and Speech	Pointing	10	1.50	3.48	0	11
	Interactional	10	5.30	6.80	1	24
	Irrelevant	10	.30	.95	0	3
	Manipulative	10	16.20	6.30	8	32
	Side effect expr.	10	1.20	1.75	0	4
	Symbolic	10	.10	.32	0	1

Table 2 – Occcurrences of Gestures. Descriptive statistics for number of occurrences all the categories by experimental condition

Cond.	Categories	N	Mean	SD	Min.	Max.
Gestures	Pointing	2	.32	.05	.28	.36
	Interactional	10	1.21	.50	.53	1.98
	Irrelevant	3	1.37	1.09	.6	2.62
	Manipulative	10	3.28	1.31	1.16	5.18
	Side effect expr.	2	.62	.37	.36	.89
	Symbolic	2	1.08	.34	.84	1.32
Gestures and Speech	Pointing	3	.79	.43	.36	1.19
	Interactional	10	1.63	.70	.81	3.30
	Irrelevant	1	.13	-	.13	.13
	Manipulative	10	3.30	.99	2.10	5.00
	Side effect expr.	4	1.38	.64	.85	2.28
	Symbolic	1	.14	-	.14	.14

Table 3 - Duration of Gestural Classes Exhibited by Subjects in the Two Experimental Conditions. Descriptive statistics regarding the duration in seconds of the different categories of gestures

From Table 2 one can see that the frequencies of pointing, irrelevant, side effect of expressive behavior and symbolic gestures for both conditions, Gestures only or Gestures and Speech, are low. A Wilcoxon matched-pairs signed-ranked test was performed for the interactional and manipulative categories. Subjects in the Gestures only condition significantly performed more interactional gestures than when in the Gestures and Speech condition (z = -2.01; p = .045). In relation to the manipulative gestures a similar pattern emerges but only approaching statistical significance (z = -1.89; p = .058).

In relation to the duration of the gestures in seconds, we chose to consider the average duration for each participant as our unit of analysis. Table 3 shows the descriptive statistics.

A similar analysis to the one performed for the frequency of the gestures was run (Wilcoxon matched-pairs signed-ranked test). No statistically significant differences were found, although the results for the interactional category approached significance (z = -1.84, p. = .066).

C. Is it possible that any differences concerning B) can be due to the effect of the order of presentation of the two conditions (gestures only or gestures and speech)?

In order to check possible effects regarding the order of the presentation five different types of variables were created: a) two variables to account the frequencies and duration of the first trial of each type of gesture regardless of whether the experimental condition was gesture only or gesture and speech, b) another two variables similar to the previous one but this time regarding the second trial and c) variable that created two groups corresponding to the two experimental conditions.

To actually investigate the present question we just compared the two groups (the group of subjects that had the gestures only condition first with the group of subjects that had the gestures and speech condition first) regarding the first and second trial for each type of category of gesture. The Mann-Whitney U statistic was used to this effect.

No statistically significant differences in the frequencies of gestures were found in relation to the first trail, meaning that the conditions gesture only or gesture and speech first did not differ in terms of frequency of each type of gestures. However, in relation to the second trial, statistically significant differences were found for the interactional and manipulative gestures. The subjects in the gestures and speech first condition produced more interactional gestures (M *rank* = 8, n = 5) than the subjects on the gestures only first condition (M *rank* = 3, n = 5) U = .000, p = .008. Similarly, the subjects in the gestures and speech first condition produced more manipulative gestures on the second trial (M *rank* = 7.60, n = 5) than the subjects on the gestures only first condition (M *rank* = 3.40, n = 5) U = 2.00, p = .032. This means that the group of subjects that began with the gesture and speech condition when using just gestures produced more interactional and manipulative gestures than the group of subjects of the gestures only first condition when being able to use gestures and speech in their explanation.

Following a similar line of reasoning concerning the statistical testing employed for the frequencies, the analysis

of order effects on the duration of the gestures did not produce statistically significant results. However, the duration of interactional gestures for gestures produced in the second trial approached significance: the subjects in the gestures only first condition produced longer events (*M rank* = 7.3, *n* = 5) than the subjects on the gestures and speech first condition (*M rank* = 3.7, *n* = 5) *U* = 3.50, *p* = .056.

D. Are there any differences concerning B) due to the effect of fatigue on the production of gestures?

In order to test possible effects of fatigue we just picked the two variables created in the previous section regarding the first and second trial regardless of the condition and compared the frequencies and durations. No statistical significances were found for any type of gesture.

E. What is the distribution of co-occurrence of different types of gestures?

In this case we wanted to know how many times and when the coders did classified more than one gesture in the same time window[1]. Table 4 gives the frequencies and type of the occurrences for the gestures only first condition while Table 5 gives the frequencies for the gestures and speech first condition.

Tables 4 and 5 suggest that co-occurrence happened more frequently when the subjects started by demonstrating using gestures and speech first. However, a closer look at the data also tells us that 3 subjects did not show any co-occurring gestures (curiously all in the gestures only first condition). Thus, only seven subjects really contributed to the tables. Furthermore, one of the subjects on the gestures and speech first condition did produce a large amount of co-occurrences (13 co-occurrences of Interactional+Manipulative in the first trial and 2 on the second trial). So, these results have to be carefully interpreted.

Trial	Type of co-occurrence	Freq.
1	Interactional+ Manipulative	3
	Manipulative+Irrelevant	1
2	Interactional+Side effect express. behavior	1

Table 4 - Frequencies of co-occurrence per type and trial for the gestures only first condition

Trial	Type of co-ocurence	Freq.
1	Interactional+Side effect exprss. behavior	1
	Interactional+Manipulative	16
	Manipulative+Irrelevant	1
	Manipulative+Deictic	1
2	Manipulative+Symbolic	1
	Interactional+Manipulative	6
	Interactional+Irrelevant	1
	Interactional+Side effect exprss. behavior	1

Table 5 - Frequencies of co-occurrence per type and trial for the gestures and speech first condition

[1] In the coding of the data presented here a distinction was not made between coding one behavior in more than one category or coding more than one behavior in the exactly the same time window.

F. What method of explanation (gestures only or gestures and speech) did the participants' prefer?

In the post-sessions questionnaire, the participants were asked which method of explanation they preferred. Eight out of the ten participants stated preferring the use of gestures and speech. Only one person considered gestures only while the remaining participant stated no preference. The reasons stated by the eight participants to prefer speech and gesture were: "it felt more natural", "it conveys more information" and "is closer to the way humans usually communicate". For the participant that chose gestures the reason was that it was just easier to do it that way.

Another related question was if they could see themselves teaching a robot using similar methods. Again eight out of ten participants considered that they could envision using gestures and speech to teach a robot. Two participants answered that both methods were plausible. Interestingly, the participant that stated to prefer using gestures to explain the task (previous referred to question) did answer gestures and speech in this question. The reasons the eight participants gave to the plausibility of using gestures and speech to teach the robot were related to: personal preference or ability and belief on "being able to produce better explanations".

G. Which method (gestures only or gestures and speech) did they think produced the better explanation?

The last comment in the previous sub-section already gives a hint on which method people thought produced the better explanation. In fact, eight out of the ten participants chose gestures and speech. The reasons invoked were: gesture and speech are complimentary, using gestures and speech allows conveying more information, using speech helps focus on the task and speech allows the clear identification of objects.

We also asked the participants if they could give us their opinion regarding which method of explaining made them reflect more on how to demonstrate the task. Six participants pointed to the gestures only method while the remaining four considered that gestures and speech made them reflect more on the task. The main reasons the six participants that chose gestures gave for their opinion were: not being a natural way of explaining and more difficult to demonstrate the task. The four participants that named gestures and speech considered that this method of explaining "forced" them to verbalize aspects of the task.

IV. Discussion and Future Work

The experiments presented here serve to validate the general coding scheme for gesture in human-robot interaction (developed following Nehaniv et al. [4]), and provide us with some first information on the distribution and duration of gestures in the various classes (Tables 2 & 3). Moreover, a detailed corpus annotated according to the

classification has been obtained to allow HRI researchers to examine example naturally occurring gestures that will hopefully be more indicative in their types and characteristics that on-board HRI systems will need to recognize (for a discussion of the technological challenges behind gesture recognition see, for example, [7, 29]).

The inter-rater agreement reached seems satisfactory, especially taking into consideration the small number of occurrences for the irrelevant, side effect of expressive behavior, symbolic and pointing gestures (Table 1). Nevertheless, we intend to investigate further the disagreements obtained for the irrelevant and side effect of expressive behavior categories since it might be the case that the two are not clear and distinctive enough.

Another issue that clearly emerged from the analysis was the low frequency of pointing and symbolic gestures. In fact, we were expecting that the constraint of not being allowed to use speech would make people resort to pointing and symbolic gestures to supplement their manipulative gestures. However, it seems that people, in the gestures only condition, chose to be more detailed in their manipulation of objects and sometimes use a special type of manipulation to make their explanation more salient: they would grasp the object, transport it to the front of the video camera, turn it a bit to exhibit the object and then place it on the table. However, even the subjects who showed this sequence were not consistent throughout. Nevertheless, this example clearly suggests the need to investigate further sequences of activity.

The frequencies of interactional and manipulative gestures were higher for the gestures only condition. However, when testing the effect of the order of presentation, we saw that starting with gestures and speech made subjects produce more gestures in the following condition, gestures only, than the other way around. A possible explanation for this effect of order is that people felt less comfortable when starting with the gestures only condition and that constrained their following demonstration. The subjects' answers to the post-session questionnaire support this view. Subjects preferred using gestures and speech to only gestures in their explanations. They also thought their explanation was better when using gestures and speech. The reason they invoked more frequently was the degree of naturalness of this demonstration method.

More surprising was the subjects' answers to the question concerning which method of demonstrating made them reflect more about the task. The opinions were almost split: some subjects considered that using gestures only made them think more about how to demonstrate due to the novelty of the situation. However, some participants pointed out that gestures and speech made them think harder because they had to verbalize about what they were doing. This issue suggests a certain tension between the nature of the method used to demonstrate and the task itself. In relation to the task itself it might be the case that the reason to the extra work related to the verbalization is connected to the degree of automaticity of the task. People are just not used to verbally explaining how to lay a table, they just do it.

The line of reasoning considered in this section also highlights the differences and similarities of our study and its results from Lozano and Tversky's study [21]: the participants that used gestures and speech (*Speakers*) in their explanation exhibited the cart pieces to be assembled and pointed to objects less than people just using gestures (*Gesturers*). In our study that was not the case. In relation to gestures that convey information about action, the *Gesturers*, in Lozano and Tversky study, produced more gestures than *Speakers*. If we consider the results regarding our manipulative gestures, it seems we do have a similar result. So, in terms of general results, it seems that the main difference is the frequency of pointing and exhibiting manipulative gestures when people are asked to demonstrate how to perform a task using gestures and speech. Lozano and Tversky's experimental task involved the explanation of actions to accomplish the assembly of an object while in our case the task had more to do with the structural layout of a particular setting. Furthermore, in their experimental task the participants were faced with a novel challenge thus the degree of automaticity of the actions to the demonstrated was perhaps lower. The degree of automaticity might have influenced the way people chose to demonstrate the laying of the table when asked to use gestures and speech: instead of using pointing and exhibiting they just manipulated the objects to their corresponding places. Thus, it might be the case that the types of gestures produced are closely linked not only to the presence or absence of speech but also to the nature of the task itself.

What were the lessons learned relevant for HRI? Three issues seem particularly important. The first one concerns the subjects' preference for the gestures and speech method of demonstrating and its implications. This choice is not surprising but it definitely supports the perspective that people might prefer to interact with robots in a "natural" way [5]. The second issue is the infrequent occurrence of pointing and symbolic gestures. It seems that in routine daily tasks people are not naturally likely to give detailed accounts of the way the tasks should be performed beyond the actual simple demonstration of how to accomplish it. So the questions are: (a) what specific strategies can be used in robotic systems to accommodate this? (b) can people accommodate to the need of being more explicit regarding their explanations? (Related to this issue is a general question for technological systems: to what degree should HRI designers expect them to adapt to technology rather than the other way around?) Finally, the third issue is the possible interaction between the type of task and the type of gestures produced. This point stresses the importance of knowing the context in which gestures are produced and their interaction history [4]. A possible shortcoming of the

experiment is the lack of any explicit feedback from the system to the human subjects, as there may be reason to suspect that the occurrence and character of gestures in an interaction may well be shaped by such factors as recipient design and communicative negotiation [30]. Further studies should indicate to what extent the design of the feedback as well as robot's appearance affect the distribution of gesture. Preliminary studies we have carried out suggest that this is indeed the case [28].

In terms of future work with the data collected in this study, we will now turn our attention to the classification of the speech produced in relation to the gestures that co-occur. We intend to find out to that extent the speech disambiguates the gestures and activity being pursued or if time lags between the production of speech and closely related gestures might provoke misconceptions. In terms of next studies, we find particularly important to investigate the effect of feedback from the system in people's production of interactional gestures. In fact, we believe that the introduction of feedback not only might alter the frequency of interactional gestures but also of the other types.

REFERENCES

[1] S. Kiesler and P. Hinds, "Introduction to This Special Issue on Human-Robot Interaction," *Human-Computer Interaction*, vol. 19, pp. 1-8, 2004.

[2] K. Dautenhahn, "Robots we Like to Live With? - A Developmental Perspective on a Personalized, Life-Long Robot Companion," in *Proceedings Of The IEEE Ro-man, 13th International workshop on Robot and Human Interactive Communication*. Kurashiki, Okayama, Japan: IEEE Press, 2004, pp. 17-22.

[3] B. Robins, K. Dautenhahn, C. Nehaniv, N. A. Mirza, D. Francois, and L. Olsson, "Sustaining interaction dynamics and engagement in dyadic child-robot interaction kinesics: Lessons learnt from an exploratory study," in *Proc. IEEE Ro-man 2005*, 2005, pp. 716-722.

[4] C. Nehaniv, K. Dautenhahn, J. Kubacki, M. Haegele, C. Parlitz, and R. Alami, "A methodological approach relating the calssification of gesture to identification of human intent in the context of human-robot interaction," in *Proc. IEEE Ro-Man*, 2005, pp. 371-377.

[5] K. Dautenhahn, "The Art of Designing Socially Intelligent Agents: Science, Fiction, and the Human in the Loop," *Applied Artificial Intelligence*, vol. 12, pp. 573-617, 1998.

[6] T. Fong, I. Nourbakhsh, and K. Dautenhahn, "A survey of socially interactive robots," *Robotics and Autonomous Systems*, vol. 42, pp. 143, 2003.

[7] N. Otero, S. Knoop, C. Nehaniv, D. S. Syrdal, K. Dautenhahn, and R. Dillman, "Distribution and Recognition of Gestures in Human-Robot Interaction," in *Proceedings of Ro-man 2006 (this volume)*, 2006.

[8] J. Cassell, "Nudge Nudge Wink Wink: Elements of Face-to-Face Conversation for Embodied Conversational Agents," in *Embodied Conversational Agents*, J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, Eds. Cambridge, Massachusetts: The MIT Press, 2000, pp. 1-27.

[9] D. McNeill, *Hand and Mind*. Chicago: Chicago University Press, 1992.

[10] A. Kendon, "Gesture," *Annual Review Of Anthropology*, vol. 26, pp. 109-128, 1997.

[11] M. Kipp, *Gesture Generation by Imitation: From Human Behaviour to Computer Character Animation*. Boca Raton, Florida: Dissertation.com, 2004.

[12] S. Ozcaliskan and S. Goldin-Meadow, "Do parents lead their children by the hand?" *Journal Of Child Language*, vol. 32, pp. 481-505, 2005.

[13] J. M. Iverson and S. Goldin-Meadow, "Gesture paves the way for language development," *Psychological Science*, vol. 16, pp. 367-371, 2005.

[14] M. A. Singer and S. Goldin-Meadow, "Children learn when their teacher's gestures and speech differ," *Psychological Science*, vol. 16, pp. 85-89, 2005.

[15] W. M. Roth and D. Lawless, "Scientific investigations, metaphorical gestures, and the emergence of abstract scientific concepts," *Learning And Instruction*, vol. 12, pp. 285-304, 2002.

[16] S. D. Kelly, M. Singer, J. Hicks, and S. Goldin-Meadow, "A helping hand in assessing children's knowledge: Instructing adults to attend to gesture," *Cognition And Instruction*, vol. 20, pp. 1-26, 2002.

[17] M. W. Alibali, M. Bassok, K. O. Solomon, S. E. Syc, and S. Goldin-Meadow, "Illuminating mental representations through speech and gesture," *Psychological Science*, vol. 10, pp. 327-333, 1999.

[18] P. Garber and S. Goldin-Meadow, "Gesture offers insight into problem-solving in adults and children," *Cognitive Science*, vol. 26, pp. 817-831, 2002.

[19] J. Cassell, D. McNeill, and K. E. McCullough, "Speech-Gesture Mismatches: Evidence for One Underlying Representation of Linguistic and Non-Linguistic Information," *Pragmatics and Cognition*, vol. 7, pp. 1-33, 1999.

[20] R. M. Krauss, R. A. Dushay, Y. Chen, and F. Rauscher, "The Communicative Value of Conversational Hand Gestures," *Journal of Experimental Social Psychology*, vol. 31, pp. 533-552, 1995.

[21] S. Lozano and B. Tversky, "Communicative gestures facilitate problem solving for both communicators and recipients," *Journal Of Memory And Language*, in press.

[22] S. Ghidary, Y. Nakata, H. Saito, M. Hattori, and T. Takamori, " Multi-Modal Interaction of Human and Home Robot in the Context of Room Map Generation," *Autonomous Robots*, vol. 13, pp. 169-184, 2002.

[23] J. Y. Oh, C. W. Lee, and B. J. You, "Gesture recognition by attention control method for intelligent humanoid robot," in *Knowledge-Based Intelligent Information And Engineering Systems, Pt 1, Proceedings*, vol. 3681, *Lecture Notes In Artificial Intelligence*, 2005, pp. 1139-1145.

[24] K. Severinson-Eklundh, A. Green, and H. Huttenrauch, "Social and collaborative aspects of interaction with a service robot," *Robotics and Autonomous Systems*, vol. 42, pp. 223, 2003.

[25] J. Cassell, "A Framework For Gesture Generation And Interpretation," in *Computer Vision in Human-Machine Interaction*, R. Cipolla and A. Pentland, Eds. New York: Cambridge University Press, 1998, pp. 191-215.

[26] J. Cassell and K. R. Thorisson, "The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents," *Applied Artificial Intelligence*, vol. 13, pp. 519-538, 1999.

[27] S. Koop, P. Tepper, K. Ferriman, and J. Cassell, "Trading Spaces: How Humans and Humanoids use Speech and Gesture to Give Directions," *Spatial Cognition and Computation*, in press.

[28] N. Otero, C. Nehaniv, K. Dautenhahn, J. Saunders, and A. Alissandrakis, "Naturally Occurring Gestures in a Human-Robot Teaching Scenario: An exploratory study," School of Computer Science, Faculty of Engineering and Information Sciences, University of Hertfordshire, Technical Report 443, 2005.

[29] S. Knoop, S. Vacek, and R. Dillman, "Sensor fusion for 3D human body tracking with an articulated 3D body model," in *Proceedings IEEE Intl Conference Robotics and Automation (ICRA)*. Orlando, Florida, 2006.

[30] P. G. T. Healey, N. I. K. Swoboda, I. Umata, and Y. Katagiri, "Graphical representation in graphical dialogue," *International Journal of Human-Computer Studies*, vol. 57, pp. 375, 2002.