The 18th IEEE International Symposium on
Robot and Human Interactive Communication
Toyama, Japan, Sept. 27-Oct. 2, 2009

"The first five seconds": Contingent stepwise entry into an interaction as a means to secure sustained engagement in HRI

Karola Pitsch, Hideaki Kuzuoka, Yuya Suzuki, Luise Süssenbach, Paul Luff, Christian Heath

Abstract*— If robot systems are being deployed in real world settings with untrained users who happen to accidentally pass by or could leave at any moment in time, then this places specific demands on the robot system: it needs to secure and maintain the user's engagement. In this, a common and critical problem consists of entering into a 'focused encounter'. It requires each interactional partner to closely react upon the other's actions on a very fine-grained level engaging in a stepwise and dynamic process of mutual adjustments. We report initial findings from a study in which we have developed a preliminary, simple solution to this problem inspired by work from Conversation Analysis [7]. Using this as an instrument to explore the impact of a 'contingent' (CE) vs. 'non-contingent entry' (NCE), we find that users who enter into the interaction in a dynamic and contingent manner show a significantly different way of interacting with the robot than the NCE group.

I. Introduction

In recent years, a range of inititatives have enabled robots and other technical systems to interact with the human user in a more naturalistic way. Quite understandably, attention has mainly focused on investigating how robot systems can interact with humans in laboratory conditions. More recent studies have begun to explore the use of robot systems in the real world – in museums, shopping malls, train stations etc. [1,2,3,] – where users are neither asked to participate in a particular experiment nor receive any prior training. Under such conditions naïve users happen to accidentally pass by a robot located at some place, they have to explore by themselves how the system works and could leave the interaction whenever they would like to. This places specific requirements on the robot system, most importantly to secure and maintain the users' engagement.

Under these conditions, a common and critical problem consists in entering into a 'focused encounter' (Goffman) the robot needs to get the user's attention, the user has to identify that the robot is addressing a recipient, and the robot needs to 'organize' the user into a position where he/she could both orient to the robot and to a common object [4]. For participants to deal with such situations, involves a great deal of interactional work: it requires each interactional partner to closely react upon the other's actions on a very fine-grained level engaging in a stepwise and dynamic process of mutual adjustments. Furthermore, interactional research on human communication has revealed the extent to which "the first five seconds" of an encounter are crucial to how the interaction will continue [5]. Thus, we suggest that it is important – also in human-robot-interaction (HRI) – to take particular care of how to design the opening of the interaction.

In this paper, we will present some initial findings from a study in which we have investigated a rather simple solution for a robot to deal with the practical problem of entering a 'focused encounter', which is inspired by work in interaction analysis. We adapted a Sony Aibo (ESR-7) robot to act as a guide in a Japanese museum. Placed next to a painting, the robot monitors the user's gaze behavior and dynamically adjusts the delivery of its talk: if it loses the visitor's gaze, it stops talking, pauses briefly and restarts its talk. For human conversations, this 'pause and restart' procedure [6] has been shown to be a systematic device for securing a co-participant's attention and alignment. In fact, the need to do so not only happens at the beginning, but is also a frequent task within an interaction. In a previous study, we have shown the 'pause and restart' procedure to be an effective means for a museum guide robot to eliciting a co-participant's gaze during an ongoing explanation [7]. Now, we will explore its effect for the task of entering into a 'focused encounter' and evaluate its impact for sustaining the user's engagement.

II. Background and Related work

In the field of human-robot-interaction, it has become increasingly popular to develop robots that act as museum guides. This is probably because this offers a real world scenario with relatively stable and controlled conditions: the robot has got a clearly defined interactional role as a presenter using some possibly pre-configured explanation. But in doing this, the system needs to be sensitive to the visitors and how their behavior might change over time: how and where they are oriented to and how their conduct is related to objects in the local environment.

While a range of studies have focused on the autonomy of the system designing it to navigate safely through a museum [8], a different line of research has investigated the

Manuscript received March 25, 2009. This work was supported in part by a grant of the British Academy/Japan Society for the Promotion of Science and the FP7 European Project "iTalk. Integration and Transfer of Action and Language Knowledge in Robots" (ICT-214668). Our special thanks go to Keiichi Yamazaki (Saitama University), Akiko Yamazaki (Tokyo University) and Yoshinori Kuno (Saitama University) for their valuable support when conducting the study.

Karola Pitsch and Luise Süssenbach, Bielefeld University (e-mail: karola.pitsch@uni-bielefeld.de).

Hideaki Kuzuoka and Yuya Suzuki, Tsukuba University (e-mail: kuzuoka@iit.tsukuba.ac.jp).

Paul Luff and Christian Heath, King's College London (e-mail: paul.luff@kcl.ac.uk, christian.heath@kcl.ac.uk).

interaction between robot and user [2, 9, 10], and the role of gaze during talk [11]. Other studies have begun to explore how to best design the robot's explanation of a painting. Based on studies of human interaction, they have explored the precise timing of head and body movement at systematic places in the talk and demonstrated the effect that it can systematically guide the visitor's attention between the guide and the exhibit [1, 10]. Also, the effect of particular communicational devices such as 'pause and restart' to gain a visitor's attention has been shown [7].

These studies (similar to other application areas) assume that the user somehow gets in contact with the system – leaving out the moment of *entering* into a focused encounter. However, research on conversational openings in human-human-interaction (HHI) reveals that participants have to establish mutual awareness, have to recognize and identify each other and check the other's availability before they can actually proceed to deal with content related matters. It turns out that such openings are a highly dynamic, stepwise process during which participants react upon each other on a very fine-grained level using gaze, bodily behavior, spatial repositioning, talk etc. [4, 12, 13]. Thus, it will not be sufficient to implement some pre-configured action script [3], but we will need to find ways of enabling the system to deal with the dynamic and flexible nature of human interaction – i.e. to find new ways of negotiating the tension between 'plans and situated actions' [14]. Although human interaction is highly organized and systematic, the way in which it will unfold in time is not precisely predictable: "an utterance can make a range of sequelae or responses contingently relevant next. Which of alternative contingent actions a next speaker will do, however, is not in principle predictable" [15]. Therefore, we will need to equip systems with (a) ways of monitoring the user's behavior, (b) interpreting this as meaningful events in terms of interaction management, and (c) adjusting its own behavior accordingly.

III. Robot System

In order to explore new ways of enabling autonomous robot systems to dynamically enter into a 'focused encounter' and to secure and maintain the visitor's engagement, we have adapted a Sony Aibo (ESR-7) robot system to act as a museum guide robot in a Japanese museum. The robot was programmed to offer information – by using talk, head movement and gestures – about a painting next to which it was sitting on top of a column in the corner of a large exhibition space (Fig. 1). The robot was set up to work autonomously, engaging by itself with visitors who happened to accidentally pass by during their visit to the museum.

We designed the robot's explanation – in accordance with structural properties of human conversation – to have three distinct phases: (1) an opening sequence, in which the participants could establish mutual awareness, recognize/ identify each other and check the other's availability; (2) the explanation of the painting; and (3) a closing sequence.

Fig.1a and b: Sony AIBO used as a museum guide robot. In order to make the robot's deictic gestures distinct, we attachted a pointing shaped hand to the robot's right arm.

In this paper, we focus on the opening phase and how we could design the system so it would be able to dynamically react upon the visitor's behavior. We enabled the system to detect and monitor the visitor's head orientation, interpret this as an indicator of their attention, and provided a simple mechanism to dynamically break up the pre-designed talk. For this, we have been inspired by work in interactional research: when, in human conversation, a current speaker begins a turn at talk and finds the intended recipient not attending, a speaker often pauses and/or restarts the delivery of the utterance he/she is currently producing. Within an ongoing interaction, this will typically elicit the recipient's attention/gaze [6, 7]. For our system, we have used the 'pause and restart' procedure as a means to explore novel ways of dealing with the 'opening problem' and to investigate its potential effects on the ensuing interaction.[1]

Implementing this solution in our Sony Aibo (ESR-7) robot, we have created the following set up: the robot sits in waiting position (head down, arms in 90 degree angle next to its body) on its column next to the painting. Once it detects somebody gazing either at itself or the painting, it leaves the resting position – lifting the head, lighting the eyes and turning its head towards the detected person. After 1.0 second, the robot then starts to talk: "Excuse me" – (0.5) pause – "Would you like to hear a brief explanation about this work by Cézanne?" During this, the system monitors the visitor's gaze, and depending on him/ her looking towards the robot (❶) or not (⓪) either produces the question as a whole, or interrupts the talk according to the 'pause and restart' procedure presented in Fig. 2. This way, the robot does not simply deliver a predefined speech sequence, but it does this dynamically, with regard to the visitor's current attention. Given that the 'opening' is the moment to organize the entry into the conversation and prepare for the ensuing content related talk, the visitors

[1] In doing this, we make a first attempt to use particular insights from Conversation Analysis as an *inspiration* for resolving particular problems encountered when designing sociable robots. Whilst we draw upon systematic interactional procedures found in human interaction, we do not aim at replicating human behavior: Our concern is to find appropriate ways *for a robot* to deal with the flexibility of natural interaction.

should be – at its end – in a relevant position for the robot to procede with the explanation of the painting. [2]

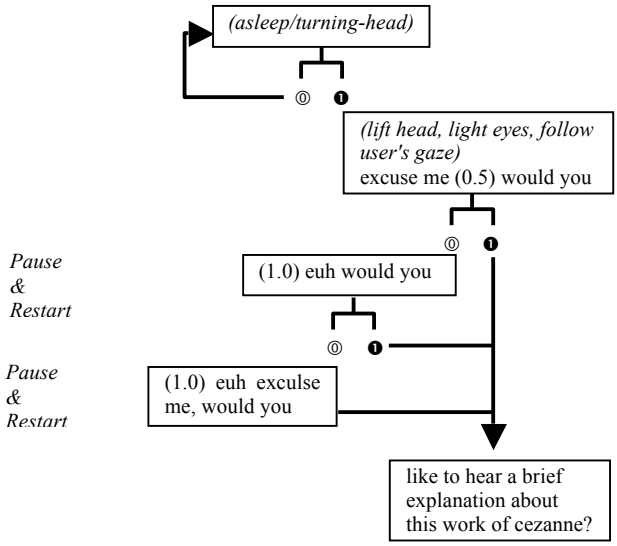

Fig.2. Opening sequence: Model of stepwise adjustment of robot's talk depending on the user's attention (measured as gaze: ❶ user gazes to robot, ⓪ user gazes away)

While running autonomously, the robot was controlled by an external laptop, connected via a wireless network. It was equipped with an external two-camera unit placed behind the robot and providing a 100 degree field of view. In addition, there was another camera beneath the painting. These visual inputs are analysed using the Intel Open Source Computer Vision Library (OpenCV), specifically used the face recognition algorithm. We assumed that if the camera unit behind the robot detected a visitor's face, he/she was looking at the robot and if the camera beneath the painting detected a visitor's face, he/she was looking at the painting. In order to give the appearance of the robot speaking, we used pre-recorded voice fragments that were generated using the free speech synthesizer AquesTalk. Based on predefined sequences and the current sensory data, the program chooses appropriate voice fragments and gestures and controls the robot accordingly.

IV. Experiment

In November 2007, we have conducted a one day field trial with our robot system at the Ohara Museum of Art, Kurashiki, Japan, recognized for its European master pieces and attracting about a million visitors per year. The Sony Aibo robot was placed on top of a column in the corner of a large exhibition space next to a landscape painting by Cézanne (64.5 x 81.0 cm) which it was set up to explain to visitors who happened to pass by. When entering the room, visitors had been informed by signs about an ongoing experiment and being video-taped, but they did not receive any guidance in what exactly the robot could do, how it would behave and how they were supposed to interact with it. Thus, the robot was faced with naïve users and the initial practical problem of getting in contact with them, to potentially organize them into a relevant location where they could both listen to the robot and inspect the painting. In contrast to laboratory studies, the visitors could disengage at any moment in time and walk away.

The experiment took place for about 5 hours during one afternoon with each interactional episode taking about 2:30 minutes. The interactions were videotaped with two cameras, one from behind the robot capturing details of the visitor's upper body postures, their head movements and facial expressions (fig.1a); a second camera captured the entire scene including visitors from behind, robot and the painting (Fig. 1b). In general, visitors had no problems using the system and found it engaging. Only some smaller issues arose, such as the volume of Aibo's internal speakers which appeared low at times when the museum was busy.

V. Data and Method

During the experiment, 117 episodes of human-robot-interaction with 231 visitors were recorded. For analysis, we combined qualitative and quantitative approaches, single case analysis and work on a large corpus base. This mixed approach enables us to start with explorative, in-depth qualitative analysis of a small collection of cases drawing on Ethnomethodology and Conversation Analysis to detect analytical issues and phenomena "from the data themselves" [16] and to link this with systematically studying their effects over a collection of similar cases.

In a first step, an explorative qualitative analysis of a collection of 15 cases has been carried out and has resulted in a range of observed interactional phenomena. We then produced a generic transcript of Aibo's actions and at the places of 'pauses and restart', we specified for each episode its concrete realisation. For each single case, we annotated the participants' reactions (nodding, speech, stepping forward, moving upper body, gaze towards Aibo/painting/ visitors), and their temporal relation to Aibo's activities.

For the analysis presented in this paper, we then discarded all episodes (although potentially relevant for other research issues) in which users being familiar with the system came along as visitors (6), in which journalists interrupted the interaction (4), in which the system was not activated although participants explicitly attempted to do so (16), and those of large visitor groups, such as school classes etc., where the interaction among the participants highly influenced their behaviour (11). For the remaining 80 episodes, with 148 visitors, we transferred the annotations of the user behavior into an Excel spreadsheet to gain an overview of the corpus. From this overview, we

[2] We have also used this similar 'pause and restart' procedure at later stages during the explanation phase – but we will not report on this here. As they occur later these uses do not effect the analysis presented here.

discarded another set of cases in which it was obvious that visitors dropped out because they engaged in a conversation *about* Aibo, turned towards inspecting the (partially hidden) laptop running the system, were using an auditory museum guide when they arrived, or some other unexpected events. After this procedure, a subset of 87 visitors with relatively comparable interactional constraints remained, which forms the corpus for the present analysis.

VI. Stepwise Entry into an Interaction

As a first step, we present results from the qualitative analysis taking a close look at the interaction in the way in which it unfolds between the system and some visitors.

Let us consider the following fragment[3], in which a pair of visitors approaches the corner where the Cézanne painting and Aibo are located. At a distance of about three meters, they come to a halt, and the female visitor to the left (V1) begins to look at the Cézanne painting (Fig.3a). At this stage, the system detects a face, which triggers Aibo to lift its head, light and flash the eyes and to gaze in the direction of the detected face (Fig.3b). This, in turn, engenders a shift in the visitors' orientation: V1 and V2 (the male visitor to the right) fully turn their heads to look at the robot (Fig.3c). At this moment of mutual awareness, Aibo begins to talk: "sumimasen (0.5) kochira no" / "excuse me (0.5) this" (line 02).

Fragment 1:

```
01 A: (gaze down)          |(lift head) |
   V1:(gaze at paintig)    |            |(gaze at A)
   V2:                                  (gaze at A)
       *Fig.3a              *Fig.3b       *Fig.3c

02 A: sumimasen (0.5) kochira no
      excuse me       this
                              ↑no gaze
```

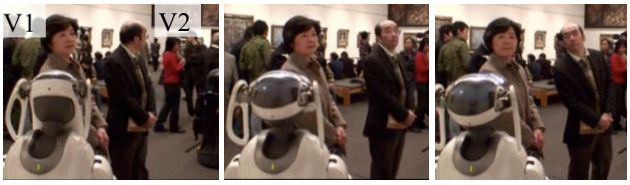

Fig. 3a Fig. 3b Fig. 3c

However, as the visitors are standing quite a distance away, the system fails to continue to detect their faces, so that 'no gaze' is detected. This triggers 'Pause & Restart 1':

```
03 A: (1.5) eeh kochira no
             euh this
      pause restart      ↑no gaze
```

After "eeh kochira no" (line 03) the system again cannot detect the visitors' faces, so that 'Pause & Restart 2' is triggered (line 04-05):

[3] For verbal utterances, the transcript gives the Japanese original with a litteral English translation below; visible actions are written in brackets.

```
04 A: (1.0)|(1.0)      |ano |(.)        |sumimasen
                        euh              excuse me
      pause             restart
   V1:     |(step fwd)|     |(gaze to A)|
              *Fig.3d           *Fig.3e

05 A: (.)  |kochira no sezan|nu no sakuhin ni
            this       cézanne's   work
   V1:(nod)|
```

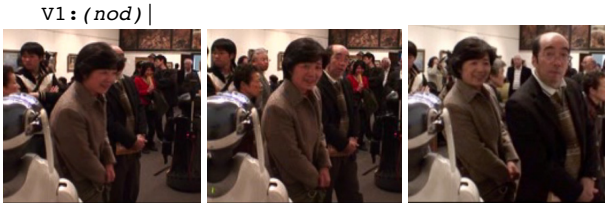

Fig.3d Fig.3e Fig.3f

While V1 and V2 remained in their positions after 'Pauses & Restart 1', now – after 1.0 seconds of silence (line 04) – V1 makes a step forward towards Aibo, slightly bending her torso and looking down (Fig.3d). This takes about another second, so that Aibo's restart "ano" (line 04) happens to follow precisely in next turn position. This – in turn – is answered by V1 turning her face to the robot (Fig.3e), and Aibo then utters again "sumimasen / excuse me". Not only does this analysis of the sequential structure of the interaction reveal that – apparently (i.e. with regard to visible interaction) – Aibo's and V1's actions seem to systematically respond to each other, but also that the visitor herself shows her impression of the system's responsive abilities: she answers Aibo's "sumimasen / excuse me" by nodding at the robot (line 05). Thus, structurally speaking, she delivers a 'go-ahead' [17], inviting the robot to continue. The system then indeed does produce the initial question:

```
05 A: (.)  |kochira no sezan|nu no sakuhin ni
            this       cézanne's   work
   V1:(nod)|

              nod↓    ↑      ↓ ↑     ↓ ↑
06 A: kansuru katan na se|tsu|mei wo |okiki ni  |
      about brief explanation        hear
   V1:                               |(step fwd)|
         *Fig.3f                            *Fig.3g

07 A: nari|masu ka, (2.0)|kochira no e    wa, …
      you would like to ? this       painting is
   V2:    |(step fwd)  |
            *Fig.3h                    *Fig.3i
```

Fig.3g Fig.3h Fig.3i

During this period of asking the question, also the second visitor V2 approaches step by step, and V1 takes another little step towards Aibo. This way, at the end of Aibo's

question – i.e. before the actual explanation begins – both visitors find themselves in the relevant location for being able to both listen to Aibo's talk and inspect the painting.

This analysis reveals:

1) Similar to HHI, entering an interaction is a practical task also in HRI that involves several steps of mutual adjustment between the participants and to establish certain pre-requisites before the actual topic talk starts.
2) The implemented procedure of making the system monitor the user's gaze and breaking up turn-units by using the 'pause and restart' procedure turns out to be an adequate means to help create a contingent, stepwise interaction between robot and human user.
3) Whilst using the 'pause and restart' procedure for entering the interaction is highly dynamic, reacting upon the user's gaze at specific moments in time, other features, such as the duration of the pauses, are pre-programmed. Apparently, due to the design of the pause length it appears to the user as being responsive to his/her actions.

VII. Contingent vs. Non-contingent Openings

While in fragment 1, the simple mechanism of combining face detection with the 'pause and restart' procedure performs particularly well to enable the robot to engage into what seems a sequentially unfolding opening of a 'focused encounter', this is, however, not always the case in our data. In fact, from the total of 87 cases examined, we find about 46 cases (52.9 %), in which human and robot manage to produce a stepwise entry into the interaction, in which one participants' actions appear to be reacting correctly upon the other's. We call this a "contingent entry" (CE), refering to Yamaoka et al.'s [18] understanding of this term as "a correspondence of one's behaviour to another's behaviour". In 41 cases (47.1 %) our mechanism appears to fail in the sense that it is not able to make the system produce the right reactions upon the user's previous actions and/or with the correct timing. For example, under certain conditions the system performs the 'pause and restart' procedure *although* the user is showing attention and has already stepped into a relevant position for both inspecting the painting and listening to the robot. We call this "non-contingent entry" (NCE).

Fragment 2 shows a NCE-example: Here, we find a visitor approaching the system (Fig.4b, the system then detects the visitor's gaze, lights its eyes (Fig.4b) and says "sumimasen" / "excuse me". Normally, at this stage, the robot is programmed to turn its head into the direction where it has detected the gaze. But in this particular case its head remains oriented to the right side (it seems that the system has been confused with another person arriving at the picture in a little distance). As in the first fragment, the robot's talk invites the visitor to approach (Fig.4b). However, as the visitor brings his ear close to the robot (Fig.4c), this appears to the system as having lost the visitor's gaze, and accordingly triggers the restart. For the visitor, who did follow the robot's initial initiative, this behaviour necessarily appears to be strange: he backs off a bit (Fig.4d). Again, Aibo and the visitor don't make gaze contact, so that Aibo performs another restart, which lets the visitor back off further and attempt to walk away (Fig.4e). Here, the sequential relationship between the robot's and the visitor's actions falls apart, and no further interaction ensues.

Fragment 2:

Fig.4a

Fig.4b

Fig.4c

Fig.4d

Fig.4e

This result can be explained to a large extent by the simplicity of the mechanism used. The system had been set up to only detect gaze direction, but the video data suggests that other aspects of human conduct are involved as well.

VIII. The Impact of Openings for the User's Further Engagement with the System

At the outset, we did not aim at developing – at the first attempt – a perfect mechanism that could handle all possible cases of entering into a 'focused encounter', but rather we were looking to have a preliminary, simple instrument that would allow us to study the effects of breaking up turn units at particular places and its impact on the following course of action. In this sense, the mechanism we used performed well as our data enables us to investigate the further *implications* of contingent vs. non-contingent entry into the interaction for HRI. Combined qualitative and quantitative analysis reveals that, in the two conditions (CE vs. NCE), users react differently upon the robot's further actions. This suggests that – similar to HHI – also in HRI the way in which the opening of an interaction is designed is consequential for what is going to follow. In the two conditions, a different interactional situation has been created between the robot and the user.

Firstly, those visitors who experience a contingent entry, tend to remain until the very end of the robot's explanation of the painting, whereas those visitors who do not experience contingency and responsiveness of the system leave the interaction before the closing section begins. In the CE-condition, 8 participants out of 46 leave early (17.4 %), whereas in the NCE-condition 16 participants out of 41 leave early (39.0 %). The difference is significant at the 5% level (chi-square test, χ^2=5.1, p=0.024).

TABLE 1
NUMBER OF VISITORS LEAVING EARLY (DURING AIBO'S EXPLANATION BEFORE THE CLOSING SECTION STARTS)

	CE	NCE	∑
Total number of cases	46	41	87
No. of participants leaving early	8	16	24
Percentage of particpants leaving early	17.4%	39.0%	27.6%

Secondly, in the two groups we can identify different ways of carrying out the interaction with the robot. As part of the opening sequence, Aibo asks the visitors whether they would like to hear a brief explanation about Cézanne's work. In the two conditions, the visitors have a tendency (chi-square test, χ^2=3.14, p=0.076) to react differently: In the CE group 80 % of visitors produce a response, while only 63 % do so in the NCE condition.

TABLE 2
NUMBER OF VISITORS RESPONDING TO AIBO'S QUESTION AT THE END OF THE OPENING SEQUENCE

	CE		NCE		∑
Total no. of cases	46		41		87
No. of visitors answering	37 (80%)		26 (63%)		62 (71%)
- Hai/speech	9		5		
- Nod	5	27	5	13	
- Hai + nod	12		3		
- Whistle	1				
- Change body position (upper body; feet stable)	10	10	13	13	
No. of visitors not answering:	9 (20%)		15 (37%)		25 (29%)
- Change in gaze direction	4		4		
- Smile	-	9	2	15	
- Talk to others about Aibo	2		-		
- No action	3		9		

Thirdly, in the closing section of the interaction, we can again find a significant difference between the two conditions. In order to bring the interaction to an end, Aibo suggests to visitors to explore the exhibition by themselves, and in doing so performs a series of small head nods. Then it says "arigato" / "thank you" and bows. How do visitors respond to Aibo's closing of the interaction?

TABLE 3
NUMBER OF VISITORS RESPONDING TO AIBO DURING THE CLOSING SEQUENCE

	CE	NCE	∑
Total no. of cases	46	41	87
No. of participants remaining until end	38 (100%)	25 (100%)	63 (100%)
Appropriate response	31 (81%)	7 (28%)	38 (60%)
- Nod/bow	19	4	
- Talk (no nod/bow)	12	3	
No response:	7 (19%)	18 (72%)	25 (40%)

In the CE group, in 81 % of the cases visitors respond to the robot's farewell by nodding/bowing and/or some talk saying "arigato" / "thank you". In the NCE condition, we find similar behaviour only in 28 % of the cases. This difference is significant at the 1% level (chi-square test, χ^2=17, p=0.000029).

In sum, the participants who experience a contingent, stepwise entry into the interaction react – during the following interaction – differently towards the robot than those who do not experience the system as being directly responsive towards their own actions. In the CE-condition, a situation of mutual responsiveness has been established – which, as the data reveal – not only leads users to answer to the robot's question, but even invites them to engage in an activity such as bowing, with a technical system. Thus, the way in which the opening of a 'focused encounter' is interactionally organized has – similar to what is know from HHI – a crucial and distinct impact on what is going to follow in the ensuing interaction.

IX. SUMMARY OF RESULTS

In this paper, we have presented some first findings from a study in which we have focused on one common and critical practical problem that robot systems have to deal with, once they take the step from the laboratory to the real world: how to secure and sustain a user's engagement. Focussing particularly on the practical problem of entering an interaction, we have developed and investigated a first, simple solution derived from human interactional practices: a combination of monitoring the user's face orientation and – with regard to this – breaking up the robot's pre-configured talk by applying the 'pause and restart' procedure [6, 7]. Our analysis reveals:

1) Our solution performed well in 52.9 % of the cases, enabling the robot system to engage in a contingent, stepwise entry into a focused encounter with the user. Whereas in the remaining 47.1 % of the cases, the user's and robot's actions did not contingently react upon each other. Thus, while being a first step in the right direction, the mechanism chosen is too simple to handle the complexity of natural interaction.
2) More importantly, our approach provides a way for us to systematically explore the effects of dynamically breaking up turn units at particular places and to study the effects of a contingent vs. non-contingent entry into an interaction. We have been able to show that – similar to HHI – the way in which the "first five seconds" of an interaction with a robot emerge has a significant effect on the user's further engagement with the system (leaving/staying, responsiveness, exchanging rituals).

X. Discussion

While the particular topic examined in our study – entering into a 'focused encounter' – concerns a particular interactional problem, it also addresses a general issue: how to enable technical systems to deal with the dynamic nature of social interaction. As the case of 'openings' reveals, natural interaction requires each participant to closely react upon the other's actions on a micro-level engaging in a stepwise process of mutual adjustments. In this process, it is, in principle, not precisely predictable how the interaction might continue. After a given action, some relevant next actions might be highly expectable ("conditional relevant"), but there is no guarantee that a certain structural provision might indeed be responded to as supposed. This is what Schegloff [15] refers to when he uses the term "contingency". From this, further questions arise: How could a system be enabled to recognize relevant next actions? How could it identify non-contingent responses to its own actions? Which kind of mechanisms (top-down vs. bottom-up) would a system need to be able to engage in the dynamically unfolding structure of social interaction? – To help answering these questions, we suggest, that interactional approaches, such as Conversation Analysis, which investigate the sequential organisation of human interaction, seem to be a particularly insightful resource.

Building on the findings presented in this paper, our next steps forward will be to make the stepwise procedure more sensitive to the user's behavior and to develop some kind of back up if the robot does not secure the user's engagement. For this, we will need (i) to develop a more complexe technical framework and (ii) to undertake more empirical analysis about the ways in which humans precisely organize the entry into a focused encounter. We will need to look at different ways in which pauses and restarts can be deployed, e.g. varying the timing, organization of these; looking into a more systematic progression or upgrading to and from these devices. Also, we are aware that such devices could be over-used so we need to investigate ways of transforming these to make them appear to be of the moment. Further comparisons will be required to evaluate the effectiveness of such devices and to examine more closely those cases, in which users – beyond the introductary part – leave the interaction. Could this potentially be related to issues of conting vs. non-contingent conduct as well?

References

[1] Yamazaki, K., Yamazaki, A., Okada, M., Kuno, Y., Kobayashi, Y., Hoshi, Y., Pitsch, K., Luff, P., Heath, C., & Vom Lehn, D. (2009). Revealing Gauguin: Engaging Visitors in Robot Guide's Explanation in an Art Museum, CHI 2009.

[2] Shiomi, M., Kanda, T., Koizumi, S., Ishiguro, H., and Norihiro H. (2007): 'Group Attention Control for Communication Robots with Wizard of OZ Approach', in Proc. of HRI '07, 2007, pp. 121-128.

[3] Shiomi, M., Sakamoto, D., Kanda, T., Ishi, C. T., Ishiguro, H., & Hagita, N. (2008). A Semi-autonomous Communication Robot - A Field Trial at a Train Station, HRI'08. March 12-15, 2008, Amsterdam, Netherlands.

[4] Schegloff, E. A. (2002). Opening Sequencing. In J. E. Katz & M. Aakhus (Eds.), Perpetual Contact: Mobile communication, private talk, public performance (pp. 326-385). Cambridge: Cambridge University Press.

[5] Schegloff, E. A. (1967). The First Five Seconds. The Order of Conversational Openings (Unpublished Ph.D. Dissertation), University of California.

[6] Goodwin, C. (1980). Restarts, Pauses, and the Achievement of Mutual Gaze at Turn-Beginning. Sociological Inquiry, 50, 272-302.

[7] Kuzuoka, H., Pitsch, K., Suzuki, Y., Kawaguchi, I., Yamazaki, K., Kuno, Y., Yamazaki, A., Heath, C., & Luff, P. (2008). Effects of Restarts and Pauses on Achieving a State of Mutual Gaze between a Human and a Robot, CSCW 2008 (pp. 201-204).

[8] Bennewitz, M., Faber, F., Joho, D., Schreiber, M., & Behnke, S. (2005). Towards a Humanoid Museum Guide Robot that Interacts with Multiple Persons, Proc. HUMANOIDS 2005 (pp. 418-423).

[9] Nomura, T., Tasaki, T., Kanda, T., Shiomi, M., Ishiguro, H., & Hagita, N. (2006). Questionnaire-based social research on opinions of Japanese visitors for communication robots at an exhibit. AI & Society, 21, 167-183.

[10] Sidner, C. L., Lee, C., Kidd, C. D., & Rich, C. (2005). Explorations in engagement for humans and robots. Artificial Intelligence, 166, 140-164.

[11] Mutlu, B., Hodgins, J. K., & Forlizzi, J. (2006). A storytelling robot: modeling and evaluation of human-like gaze behavior, HUMANOIDS 2006 (pp. 518-523).

[12] Kendon, A. (1990). Spatial Organization in Social Encouters: The F-Formation System (pp. 209-238). Cambridge, UK: Cambridge University Press.

[13] Mondada, L. (2008). Emergent focused interactions in public places. A systematic analysis of the multimodal achievement of a common interactional space. Journal of Pragmatics.

[14] Suchman, L. (1987). Plans and Situated Actions. The problem of human machine communication. Cambridge: Cambridge University Press.

[15] Schegloff, E. A. (1996). Issues of Relevance for Discourse Analysis: Contingency in Action, Interaction, and Co-Participant Context. In E. H. Hovy & D. R. Scott (Eds.), Computational and Conversational Discourse: Burning Issues - An Interdisciplinary Account (pp. 3-38): Springer.

[16] Sacks, H. (1992). Lectures on Conversation. Oxford: Blackwell.

[17] Schegloff, E. A. (2007). Sequence Organisation in Interaction. A Primer in Conversation Analysis: Cambridge University Press.

[18] Yamaoka, F., Kanda, T., Ishiguro, H., & Hagita, N. (2007). How contingent should a lifelike robot be? The relationship between contingency and complexity. Connection Science, 19, 143-162.