Talking About Task Progress: Towards Integrating Task Planning and Dialog for Assistive Robotic Services

BROZ, Frank, DI NUOVO, Alessandro <http://orcid.org/0000-0003-2677-2650>, BELPAEME, Tony and CANGELOSI, Angelo

Citation:

Copyright and re-use policy

DE GRUYTER OPEN

Paladyn, J. Behav. Robot. 2015; 6:111–118

Research Article

Open Access

Frank Broz*, Alessandro Di Nuovo, Tony Belpaeme, and Angelo Cangelosi

Talking About Task Progress: Towards Integrating Task Planning and Dialog for Assistive Robotic Services

Abstract: The use of service robots to assist ageing people in their own homes has the potential to allow people to maintain their independence, increasing their health and quality of life. In many assistive applications, robots perform tasks on people's behalf that they are unable or unwilling to monitor directly. It is important that users be given useful and appropriate information about task progress. People being assisted in homes and other real-world environments are likely be engaged in other activities while they wait for a service, so information should also be presented in an appropriate, nonintrusive manner. This paper presents a human-robot interaction experiment investigating what type of feedback people prefer in verbal updates by a service robot about distributed assistive services. People found feedback about time until task completion more useful than feedback about events in task progress or no feedback. We also discuss future research directions that involve giving non-expert users more input into the task planning process when delays or failures occur that necessitate replanning or modifying goals.

Keywords: human-robot interaction, social robotics, task planning and execution, speech, service robots

1 Introduction

In order for assistive robotic services to readily adopted, they must be usable by non-experts with varying levels of experience with technology. This raises new questions about how to inform users about the functionality of these complex systems. As the ageing population of many countries worldwide is increasing, there is increasing interest and investment in using robotics to provide assistance to and support the quality of life of ageing people. Many of these efforts are focussed on types of assistance that enable older people to continue to live independently in their own homes as they age. One such project, ROBOT-ERA, seeks to assist the elderly by providing everyday services such as cleaning and food delivery using a group of collaborating robots [1]. Each robot in the group is designed to act in a specific environment, either in the home, in shared indoor spaces, or outdoors. The group is supported by ambient intelligence and coordinated by a central planner [2]. During the performance of these services, the robot that is with a person in the home is often not the one that is actively carrying out the current stage of a service on the user's behalf. However, through the central planner, this robot has access to the task progress involving the other robots, as well as a potentially accurate estimate of when the service will be completed. In these situations, what information should the robot provide to a person about task progress?

Because assistive robots in the home will be used by non-experts, it is of critical importance that people are able to interact with them in ways that are intuitive and easy to understand. Users interact with the ROBOT-ERA robots using a multimodal speech and tablet-based interface [3, 4]. Focus-groups conducted with ageing people in Italy and Germany in order to guide the design of the ROBOT-ERA services found that people prefer speech-based interactions with a robot over other modalities [5]. Updates given via tablet may also be missed if a person is engaged in another activity and not currently using the tablet. For these reasons, in this work we focus on verbal updates.

In this article, we explore these issues through a human-robot interaction experiment involving a simulated food delivery service. People's impressions of different types of information in verbal feedback about task progress are compared. We evaluate the feedback pre-

*Corresponding Author: Frank Broz: Plymouth University, UK, E-mail: frank.broz@plymouth.ac.uk
Alessandro Di Nuovo: Plymouth University, UK, E-mail: alessandro.dinuovo@plymouth.ac.uk
Tony Belpaeme: Plymouth University, UK, E-mail: tony.belpaeme@plymouth.ac.uk
Angelo Cangelosi: Plymouth University, UK, E-mail: angelo.cangelosi@plymouth.ac.uk

sented based on how useful and informative people found it and whether it met their expectations for the interaction. People preferred information about the time remaining until a task is complete, which supports our hypothesis that users want information that directly effects them and also does not burden them with the details of how a service is performed. While the participants in this exploratory experiment were not elderly, related literature studying task interruptions found that elderly and non-elderly participants have similar responses. We also intend to validate these results through further experimentation with elderly participants. Evaluating the quality of interaction between elderly people and service robots in realistic scenarios and environments is a key goal of the ROBOT-ERA project, and focussed laboratory experiments such as this one help to inform the design of the system. Finally, we discuss how future work on more sophisticated integration between a dialog system and planner could enable non-expert users to help to modify or repair plans according to their preferences when task execution fails.

Figure 1: A participant and the robot in the experimental setup.

2 Related Work

In most research on speech-based interaction with service robots, dialog concerns tasks that the interacting robot is performing itself, typically co-located with the person it is speaking to. In the Team Talk dialog system, users instruct a team of robots in a treasure hunting task[6]. CoBot

is a robot that uses dialog to support its tasks of guiding visitors to meetings and providing them with local information [7]. In these example applications, the robot uses speech interaction to elicit help from the user in either task planning or localisation in addition to providing information about the task. Given the assistive nature of the ROBOT-ERA services and the target user group (elderly people who may not be experienced technology users and who may have minor physical and cognitive impairments), the system should be capable of delivering services autonomously without taxing the user by making them a critical part of task performance. People may want information about the service being performed on their behalf, but directly observing the actions of the remote robots would be unnecessarily time-consuming and inconvenient. There is no need for the user to continuously monitor the robots' progress, and they are likely to focus their attention on other activities while they wait for the service to complete. People need information that provides important context about task progress with minimal distraction.

In this sense, the information updates given by the robot are related to the types of updates given by reminding agents, and the literature on task interruption is relevant to designing these updates. There is research on task interruption that focusses specifically on elderly users, but this work involves interactions with ambient intelligence or virtual agents rather than physical robots [8]. In interactions with a virtual agent, people preferred interruptions that contained social and empathetic content, which may also be the case for robotic agents [9]. A study on multimodal interruptions up by Warnock et al. found that elderly users reacted to the interruption modalities similarly to non-elderly users and that the notification modalities evaluated all had similar effects in terms of task interruption [10]. This suggests that the best modality for an interruption may be context dependent rather than a certain modality always being preferable.

These interactions are also related to research in human-in-the-loop planning. However, most work on human-in-the-loop planning for robots is focussed on supporting technically experienced expert users. The planning tasks may require the user to provide low-level control to the robot during difficult parts of a task [11]. Or the user may be called on to assign tasks to robots themselves [12]. As we discuss above, the assistive nature of our application makes offloading the difficult portions of a task onto the user inadvisable. Ideally, the dialog system should guide users to change or modify plans if necessary in a way which is intuitive to understand and which increases the user's control over the service's outcome with-

out involving them in details of how it will be accomplished. But research that involves humans interacting with autonomously generated plans highlights the difficulty of interpreting these plans for even experienced technical users [13]. Therefore, how to involve non-experts in re-planning in cases where plans fail or are delayed is a difficult open issue.

3 System

Coro, one of the indoor mobile robots designed for the ROBOT-ERA project, was used for the experiment. The control software is the system developed for all of the ROBOT-ERA robots [14]. A few modifications were made for the simplified, single-robot simulated food delivery scenario that was used. The planner was replaced by a script that executed the steps of the task (the robot's motion and speech) at specified times in order to ensure consistent timings among all experimental trials.

During the experiment, participants ordered the service using the web-based interface (see Di Nuovo et al for a detailed description) [4]. In the full ROBOT-ERA services, the interface to the robots is multimodal, and users may also give speech commands to the robot via a dialog manager. The dialog manager was not used for this experiment, and the robot produced speech only. As in the full system, Acapela's Voice As A Service was used to synthesise the robot's speech [15].

4 Experiment

4.1 Design and Procedure

The experiment was conducted on the campus of Plymouth University in a classroom set up to simulate a home environment. The example service used for the experiment is meal delivery. Participants were instructed to use the tablet interface to select a menu option and order a meal for delivery. Once they ordered a meal, the robot executed a pre-defined script of speech acts and accompanying behaviour. The experimental setup is shown in Figure 1.

The experiment has a within subjects design with three conditions: no feedback, event-based feedback, and time-based feedback. In all conditions, the participants orders a meal using the tablet interface. Five minutes later, the robot goes to the door of the room as if to collect the delivery (no meal delivery was actually performed during this experiment). While they waited for the delivery, participants were able to read magazines provided for them. This was done to simulate the conditions of ordering and waiting for a service in the home, where a person's attention may be focussed on a leisure activity. After experiencing each condition, the participant filled out a short questionnaire consisting of Likert scale responses to four statements. At the end of the experiment, the participant was asked to make a forced choice selection of their preferred condition. They were also invited to write comments about their experience of the experiment.

The scripts which the robot followed, along with timing information, are given in Table 1. The speech acts by the robot varied by condition as follows:

- **No feedback:** The robot speaks to inform the participant the order has been placed. At the end of the trial, the robot informs the participant that their order has arrived.
- **Event feedback:** In addition to the same statements as the no feedback condition, the robot informs the participant when their meal order has been prepared and is being dispatched.
- **Time feedback:** In addition to the same statements as the no feedback condition, the robot gives the participant an estimate of when their order will arrive. Later, the robot informs the participant that the estimated time until delivery has changed.

Participants were non-elderly adults recruited from the Plymouth University campus. The participants were students and staff from a variety of departments in the university. None of the participants had prior experience with the ROBOT-ERA robots or system. Twenty-one people participated in the study (M=10, F=11). The data for one user was excluded because they failed to fully complete the questionnaires.

4.2 Results

We hypothesize that users would prefer the time-based condition over the other two conditions. This hypothesis is motivated by the requirements of our application area. In assistive services, people are likely to be concerned about when a task is achieved, but uninterested in the detailed steps to achievement. The user should to be free to concentrate on other things while the service is carried out with minimal guidance. We also expect that feedback that allows someone to predict when a service will complete will be preferred to a lack of feedback. Even though the greater number of speech acts are potentially more distracting or intrusive, we expect that users will not mind distractions

Table 1: Timeline of robot speech acts for each experimental condition.

Time	No feedback	Event feedback	Time feedback
Order placed	"Your meal has been ordered for delivery."	"Your meal has been ordered for delivery."	"Your meal has been ordered for delivery."
30 seconds			"Your meal will be delivered in 7 minutes."
3 minutes		"Your meal has been prepared. It is being dispatched for delivery."	"The delivery time has changed. Your meal will be delivered in 2 minutes."
5 minutes	""Your meal delivery is here."	"Your meal delivery is here."	"Your meal delivery is here."

as long as they provide useful information. Finally, we are interested in whether the type of feedback would effect people's expectations about how long the service would take to perform. We assume that people would prefer to be given an estimate (even if that estimate must be corrected later) over an absence of information about when a task would complete.

We designed statements to evaluate participants' opinions of the robot's speech acts in each condition based on their usefulness, informativeness, appropriateness, and how well they matched their expectations. The statements on appropriateness and expectations were negatively keyed, so they were reverse scored for analysis. While the responses to individual statements are not combined in our analysis, this allows for more intuitive interpretation of results as higher scores are always positive. The questionnaire statements are:

– **Usefulness**: "I found the robot's statements about the delivery useful."
– **Informativeness**: "The robot gave me enough information about when the delivery would arrive."
– **Appropriateness**: "The robot spoke too often." (reverse scored)
– **Expectation**: "The delivery took longer than I thought it would." (reverse scored)

To test our hypothesis, we make planned comparisons between the time condition and the other two conditions. The measure used is the median response for each statement. We treat the Likert scale responses as ordinal data and use a Wilcoxon signed rank test (exact) to test for statistical significance in the cases where the medians differ (see Table 2). The distributions of responses are shown in Figure 2.

4.2.1 Usefulness

The time feedback condition was preferred to both other conditions with statistical significance. Participants were neutral about the usefulness of the robot's speech in the no feedback condition. They somewhat agreed that event-based feedback and strongly agreed that time-based feedback were useful. In their comments, (see Section 4.2.5) some participants suggested that they would like to receive both event and time feedback, while others indicated that they did not find event feedback useful.

It is possible that the usefulness of event feedback might be greater in situations in which task execution is delayed or fails because it provides more information about the current state of the system. Our example service scenario did not explore these types of planning and execution difficulties, though we expect this to be an interesting direction for future research as we will address in Section 5.

4.2.2 Informativeness

The time feedback condition was preferred to both other conditions with statistical significance. Participants strongly agreed that the time condition gave them adequate information about when the service would complete. They were in slight disagreement to neutral in the other conditions. In the event feedback condition, while participants found the robot's speech somewhat useful, it did not help them to form an estimate of when the service would finish.
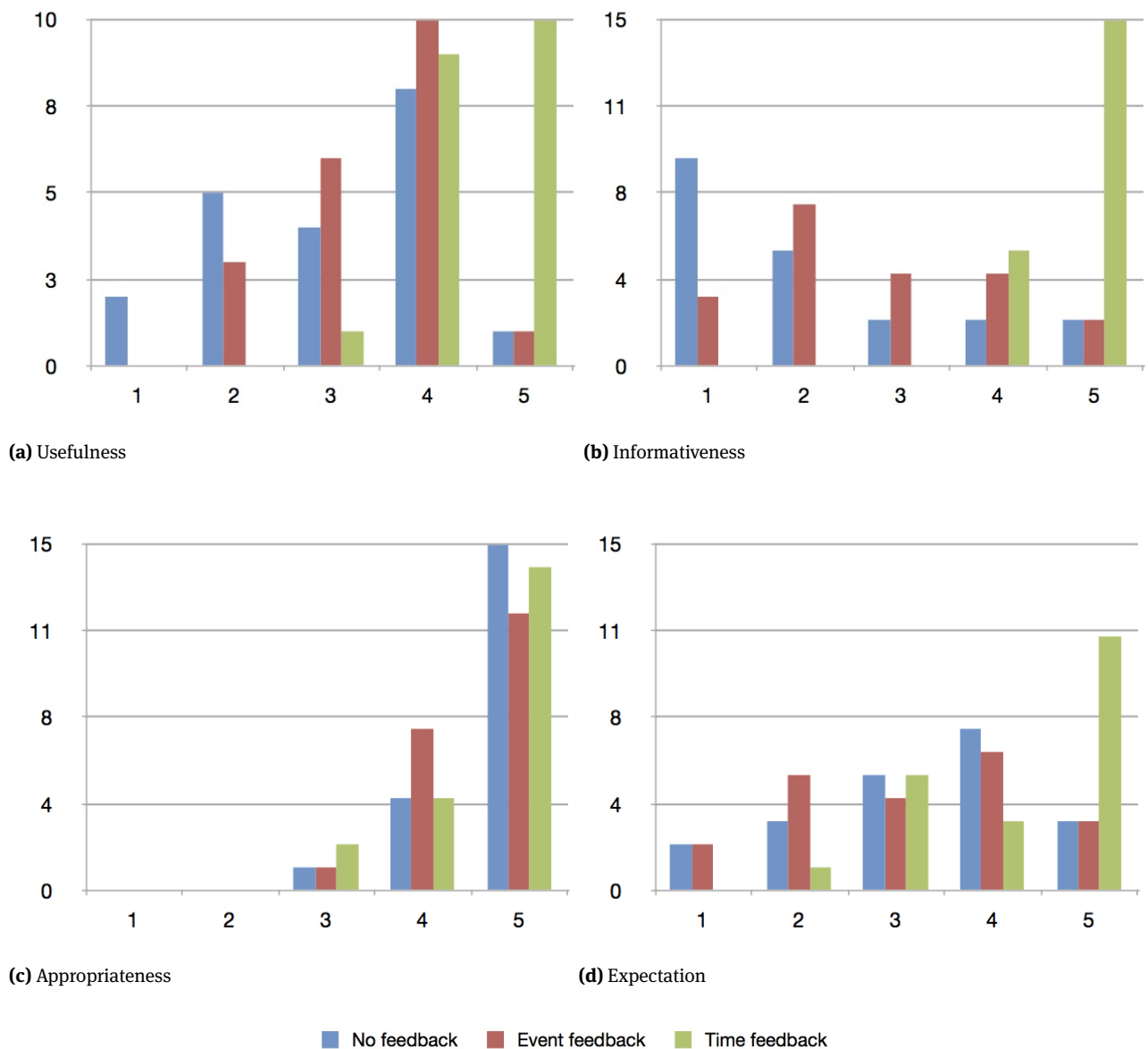
(a) Usefulness

(b) Informativeness

(c) Appropriateness

(d) Expectation

■ No feedback　■ Event feedback　■ Time feedback

Figure 2: Likert-scale questionnaire responses (1 = strongly disagree - 5 = strongly agree).

4.2.3 Appropriateness

No statistically significant differences were found between the conditions. This statement was included to investigate whether increasing the number of speech acts would negatively effect participants' opinion of the robot. Because participants were focussed on an enjoyable activity while waiting (reading magazines), it is possible that frequent interruptions by the robot could be seen as socially inappropriate, annoying and disruptive. But we expected to find that the robot's speech would not be seen as inappropriate as long as the feedback given was perceived to be useful.

The robot spoke four times in five minutes in the time feedback condition, three times in the event feedback condition, and twice in the no feedback condition. In no condition was the amount of speech judged as excessive by the participants.

4.2.4 Expectation

The time feedback condition was preferred to both other conditions with statistical significance. This statement was included to measure how the different types of information presented in the conditions would effect participants' prior expectations about how long the task would take to complete (participants were not given any information about how long each trial would take before the start of the experiment). Users strongly disagreed that the delivery took longer than expected in the time condition and were neutral in the other conditions.

It is possible that giving people an accurate estimate of how long they have to wait could lead to them feeling as if they are waiting longer by calling their attention to the passage of time. However, that seems not to have been the case in this scenario.

4.2.5 Free Comments and Forced Choice Question

Participants were invited to optionally record free comments about their impressions of the experiment upon completion. These unstructured responses where useful to gain insight into their preferences and their impressions of the interactions they had with the robot.

One aspect of the interaction that participants commented on was the robot's motion. The robot turned its body towards participants when speaking and slightly away from them between notifications. A number of participants commented that the motion of the robot made the interaction feel natural. This suggests that physical embodiment and motion had an impact on peoples' impressions of the robot's social presence during these interactions. This is one way in which giving notifications using a robot differs from presenting the same type of information using a tablet or a computer.

In addition to the questionnaire statements, we also asked participants to make a forced choice about their preferred condition in order to confirm the results of the statement responses. In the forced choice decision, 79% of users preferred the time feedback, 21% preferred event feedback, and no users preferred no feedback (see Figure 3). One user selected both time feedback and event feedback. Their response was excluded from the results because of failure to follow directions.

Participants' comments gave further insight into why the time feedback condition was preferred. Comments from several participants suggested they would prefer both forms of feedback together (which was not a condition that was evaluated in this experiment) though other

users disagreed. For example, one participant wrote, "I preferred the trial with the time updates as I didn't really care about when my food was ready for delivery, but only when it would actually be delivered." One user who preferred the event-based condition expressed a desire for the robot to inform the user about what future event would trigger the next notification (for example, having the robot say that it will inform them when the delivery is dispatched). How and when to combine time and event feedback is a topic for further study. We hypothesise that event feedback will become more useful to users in situations in which plan execution fails or is delayed. We will discuss this further in the next section.
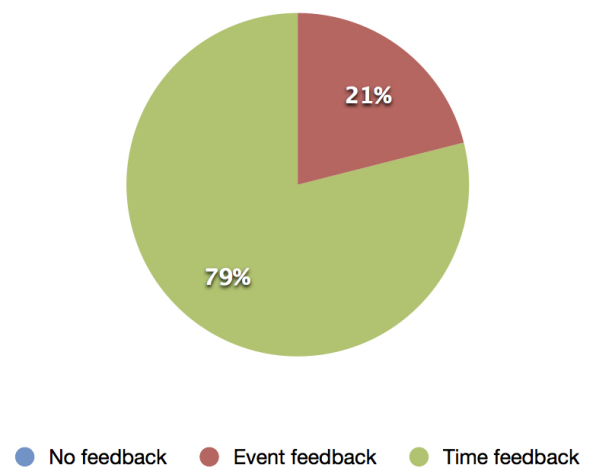

Figure 3: Participant responses to forced choice question in which they were asked to select there preferred interaction.

5 Future Work

This was an exploratory study into how to provide notifications about task progress for distributed robotic services in assistive domains, and there are a number of directions for further research. Most importantly, this study only investigated the case where the service was executed more or less "as planned" without failures or delays (in fact, in our time-based feedback condition, the service completed earlier than the robot's initial estimate). People may want more or different types of information in cases where task execution fails or a service may become severely delayed. Because the planner used in ROBOT-ERA is able to repair plans when problems arise during execution, it would be possible to inform users about these changes. A planner

Table 2: Median Likert-scale responses for each condition with 95% confidence intervals on the median of the difference for the planned comparisons. Statistical significance levels: * p < 0.05, ** p < 0.01, *** p < 0.001.

	No feedback	Event feedback	Time feedback
Usefulness	3 [-3 -1] ***	4 [-2 -1]***	4.5
Informativeness	2 [-4 -2] ***	2.5 [-3.5 -1.5] ***	5
Appropriateness	5	5	5
Expectation	3.5 [-2.5 -0.5] *	3 [-2.5 -0.5] **	5

that is aware of a user's information needs could potentially include dialog acts about task progress as part of its plans.

Delays and failures are common occurrences in real-world autonomous systems, and how to communicate with people about them is an important part of creating acceptable and interpretable robot behaviour. The constraint-based configuration planner used for ROBOT-ERA can deal with task deadlines, limited resources, and concurrent goals and can support giving accurate estimates of task completion times [16]. It also performs execution monitoring, allowing the system to detect if and when the current plan becomes unachievable.

Because the services provided effect the people being given assistance directly, it is reasonable to expect that they would want to have feedback about why plans have failed and input on what constraints should be relaxed to find an achievable plan. But finding an explanation for why a plan has become unachievable is not straightforward. Coming up with a concise, reasonable, and interpretable set of possible relaxations to present to a user as options for re-planning is an even more complex problem, in particular because the users of assistive technologies are non-experts who may not (and should not need to) understand the internal workings of the system.

Allowing users of assistive robotic services greater control over replanning could make these robotic services more acceptable by allowing a person to select alternatives according to their preferences. For example, a user might want to cancel a service or reschedule it for another day if a delay would lead to a conflict with another activity. The planner can't be sure to have complete knowledge of a user's schedule, so requesting user input is a way of preventing these conflicts.

A second level of complexity arises from the medium of communication. In our motivating example, people interact with the system's planner through a domestic robot in their home. In addition to what information should be provided, care should be taken in determining when and how often the user should be notified about task progress. Determining how long of a delay necessitates a notification is likely to depend on a variety of factors including: the service underway, the user's current activity, and potential impact on the user's future plans.

Because this experiment was conducted with non-elderly participants, it is not certain that elderly users will have the same information preferences. Related work on task interruption suggests that elderly and non-elderly people react to interruptions in a similar manner. However, as this related work did not involve robots, the preferences found in this experiment should be validated with our intended end user population. We see this study as a tool to identify promising interaction design directions that we will evaluate through interactions between elderly users and the ROBOT-ERA robotic services.

6 Conclusions

This was an exploratory study into how to provide feedback about task progress in distributed assistive robotic services. The purpose was to find out what type of task progress information people found useful and informative and to test whether speech-based notifications by the robot were seen as appropriate and led to satisfactory expectations about how long a service would take to complete. Participants were found to prefer time-based statements about task progress, though the results suggest that they also found event-based feedback to be somewhat useful. They also found the amount of speech notifications appropriate for the example scenario tested. This study only investigated the case where the service's execution did not experience delays or failures. People may want more or different types of information in cases where task execution fails or a service becomes severely delayed. How to involve a user in re-planning decisions when problems arise and when and in which cases users should be notified about changes to a planned service are also issues for further investigation. These issues will be explored through interactions between elderly users and the ROBOT-ERA robots in which the system performs assistive services in realistic environments.

References

[1] F. Cavallo, M. Aquilano, M. C. Carrozza, and P. Dario, "Robot-era project: The vision of 3d service robotics," in *Proceedings of the International Society for Gerontechnology 8th World Conference on Gerontechnology*, ser. ISG*ISARC '12, 2012.

[2] M. Di Rocco, S. Sathyakeerthy, J. Grosinger, F. Pecora, A. Saffiotti, M. Bonaccorsi, F. Cavallo, R. Limosani, A. Manzi, G. Teti, and P. Dario, "A planner for ambient assisted living: From high-level reasoning to low-level robot execution and back." in *AAAI Spring Symposium: Qualitative Representations for Robots*, ser. AAAI Technical Report, vol. SS-14-06. AAAI, 2014.

[3] F. Broz, A. Di Nuovo, T. Belpaeme, and A. Cangelosi, "Multimodal robot feedback for eldercare," in *Workshop on Robot Feedback in Human-Robot Interaction: How to make a Robot Readable for a Human Interaction Partner at IEEE International Symposium on Robot and Human Interactive Communication*, ser. Ro-Man '12, 2012.

[4] A. Di Nuovo, F. Broz, T. Belpaeme, and A. Cangelosi, "A web based multi-modal interface for elderly users of the robot-era multi-robot services," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, ser. SMC'14, 2014.

[5] F. Marcellini, R. Bevilacqua, S. Klemcke, F. Pecora, M. Aquilano, and F. Cavallo. (2012) Report on the robotic service analysis with respect to elderly users, ROBOT-ERA deliverable D2.1.

[6] M. Marge, A. Pappu, B. Frisch, T. K. Harris, and A. I. Rudnicky, "Exploring spoken dialog interaction in human-robot teams," in *Robots, Games, and Research: Success stories in USARSim Workshop at IROS*, ser. IROS '09, 2009.

[7] S. Rosenthal and M. M. Veloso, "Mixed-initiative long-term interactions with an all-day-companion robot." in *AAAI Fall Symposium: Dialog with Robots*, ser. AAAI Technical Report, vol. FS-10-05. AAAI, 2010.

[8] M. R. McGee-Lennon, M. K. Wolters, and S. Brewster, "User-centred multimodal reminders for assistive living," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '11. New York, NY, USA: ACM, 2011, pp. 2105–2114. [Online]. Available: http://doi.acm.org/10.1145/1978942.1979248

[9] T. Bickmore, D. Mauer, F. Crespo, and T. Brown, "Negotiating task interruptions with virtual agents for health behavior change," in *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems - Volume 3*, ser. AAMAS '08. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, 2008, pp. 1241–1244. [Online]. Available: http://dl.acm.org/citation.cfm?id=1402821.1402841

[10] D. Warnock, M. McGee-Lennon, and S. Brewster, "Multiple notification modalities and older users," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '13. New York, NY, USA: ACM, 2013, pp. 1091–1094. [Online]. Available: http://doi.acm.org/10.1145/2470654.2466139

[11] A. Leeper, K. Hsiao, M. Ciocarlie, L. Takayama, and D. Gossow, "Strategies for human-in-the-loop robotic grasping," in *Human-Robot Interaction (HRI), 2012 7th ACM/IEEE International Conference on*, March 2012, pp. 1–8.

[12] C. A. Miller and R. Parasuraman, "Designing for flexible interaction between humans and automation: delegation interfaces for supervisory control," *Human Factors*, vol. 49, no. 1, pp. 57–75, 2007.

[13] E. M. Roth, M. L. Hanson, C. Hopkins, V. Mancuso, and G. L. Zacharias, "Human in the loop evaluation of a mixed-initiative system for planning and control of multiple uav teams," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 48, no. 3, pp. 280–284, 2004.

[14] F. Cavallo, R. Limosani, A. Manzi, M. Bonaccorsi, R. Esposito, M. Di Rocco, F. Pecora, G. Teti, A. Saffiotti, and P. Dario, "Development of a socially believable multi-robot solution from town to home," *Cognitive Computation*, pp. 1–14, 2014. [Online]. Available: http://dx.doi.org/10.1007/s12559-014-9290-z

[15] "Acapela voice as a service," http://www.acapela-vaas.com.

[16] M. Di Rocco, F. Pecora, and A. Saffiotti, "When robots are late: Configuration planning for multiple robots with dynamic goals," in *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, Nov 2013, pp. 9515–5922.