ICT-2009.7.1
KSERA Project
2010-248085

Deliverable D3.4

The human robot dialog: context awareness for joint attention

31 January 2013
Public Document

The KSERA project (http://www.ksera-project.eu) has received funding from the European Commission under the 7th Framework Programme (FP7) for Research and Technological Development under grant agreement n°2010-248085.

Project acronym:	KSERA
Project full title:	Knowledgeable SErvice Robots for Aging

Work package:	3
Document number:	D3.4
Document title:	The human robot dialog: context awareness for joint attention
Version:	1

Delivery date:	31 January 2013 (M36)
Actual publication date:	06 February 2013
Dissemination level:	Public
Nature:	Report

Editor(s) / lead beneficiary:	Raymond H. Cuijpers (TU/e)
Authors(s):	David O. Johnson Elena Torta Nils Meins Jim Juola
Reviewer(s):	Paul Panek Cornelius Weber

Contents

Executive summary

Deliverable D3.4 addresses the human robot dialog: context awareness for joint attention. In particular, this includes the research that was done to design the KSERA system for recognizing and providing gaze cues and simple gestures, and to be able to judge whether the user is attending to the robot or not. Mutual eye contact is the most straightforward form of joint attention and context awareness.

An important aspect is how to provide gaze gestures and iconic arm gestures in a natural way and to verify their benefit in terms user experience and effectiveness of human-robot interaction. To assess the quality of human-robot interaction methods for evaluating it are developed such as a recall task, a classification task, reaction time measurement and various questionnaires. Results show that human-like gaze behaviour and iconic gestures improve user experience and memory recall, but they hardly affect persuasiveness and effectiveness of HRI. Artificial facial expressions and a dynamic interaction style further improve user experience without affecting the way in which people look at the robot much. Attracting attention is best realised using a waving gesture or through speech.

The health exercise scenario of KSERA is a perfect example of a joint task, which involves showing health exercise gestures, monitoring human gestures and providing feedback. Results show a clear benefit of using a humanoid embodied agent, as opposed to non-embodied methods like video, in terms of user experience. The movement amplitude/frequency ratio is identified as a potential measure for mental effort independent of physical effort. Providing positive feedback is found to be equally effective as negative feedback, but it results in a more positive experience. Serious gaming and playfulness can improve the experience, but older people do not like game elements unless they are functional.

The next stage of joint attention involves objects in the environment and their potential uses. Results show that the head pose of people is sufficiently consistent to be used as a model for inferring which objects are being attended to, and that object recognition is learnable by the Nao robot.

1 Introduction

The aim of WP3 is to make human-robot interaction simple, natural, and engaging in order to ensure that a successful user-robot experience is engendered from the start. Adoption of the system by elderly users will only succeed if the interface promotes a valued, positive interaction. The main WP3 objectives are:

- Improve the user experience by facilitating joint attention between the robot and the user.
- Design and develop a user interface for easy and fun user-robot interaction.
- Promote social connectedness and awareness by static and mobile video services.

The first two objectives are addressed in Task 3.1, which is reported in D3.1 and this deliverable, whereas the latter is associated to task 3.2, which is reported in deliverables D3.2 and D3.3.

Deliverable D3.4 addresses the human robot dialog: context awareness for joint attention. In particular, this includes the research that was done to design the KSERA system for recognizing and providing gaze cues and simple gestures, and to be able to judge whether the user is attending to the robot or not.

An important part of joint attention is to be able to understand human gaze cues and to let the robot provide them to the user in an understandable way. These gestures require a humanoid robot because only then it will be able to provide social gaze cues. The robot must also be able to detect the gaze direction of a person. This resulted in a face tracking and a head pose estimation algorithm for the Nao robot as reported in D3.1. In this deliverable we report the various research studies that were done to improve the HRI in terms of naturalness, usability, user experience and entertainment value. Metrics for measuring the quality of HRI were also developed.

As the KSERA project was used to shape a social robotics research line at the Eindhoven University of Technology, and also the University of Hamburg, many additional resources were recruited through student research projects. The ones that are most relevant for KSERA are reported here.

In section 2 the main outcomes are explained in terms of added value for KSERA. In section 3 the results from the research projects are summarised. Because of the large number of pages the research reports themselves are provided as an Annex to this deliverable. The dissemination level of the Annex is restricted as some of the work is still in the process of being published. Section 4 contains a list of references.

2 Added value of KSERA HRI

2.1 Added value of an anthropomorphic robot as main user interface

KSERA implements a proactive and friendly assistive service in a system that combines ubiquitous monitoring and an anthropomorphic robot with social skills. There is a consensus that humans tend to perceive robots differently than other machines, mainly due to their visible behavior and their anthropomorphism. Humans demonstrate a natural predisposition to attribute some intentional state to artifacts, including high tech gadgets and robots (Giusti & Marti, 2006). Because of its capability to trigger perceptions, attributions, behaviors and emotional reactions it becomes possible to consider the robot as a “special” affordance (Gibson, 1979; Norman, 1988). Especially in assistive robotics, the anthropomorphism is an extra value for the person. These kind of “biologically inspired" robots have been designed to simulate (and by consequence stimulate) the natural social behavior of human beings. These aspects show strong evidence that there is a psychological effect that the system causes in the individual.

The possibility to have a physical interaction and to use common gestures with a humanoid robot activates the spontaneous emergence of behavioral and emotional response by a person. A human-like embodiment makes the robot dramatically better understandable by humans (Dourish, 2001). The sense of confidence brought by the “shape” of the robot does induce the person to interact with it in the same way they usually do with other persons (Fong, 2003). Taking advantage of the anthropomorphic appearance it becomes possible to exploit the fact that the system and the robot will behave with *transparency*, that is, according to Kim et al. (2006) the ability of the robot to offer explanations of its actions. “Providing explanations of a robot’s actions, particularly ambiguous actions, will lead people to feel that they better understand the robot”. The KSERA system behaves transparently since it explains the actions via the communication channels, voice, mimic and gestures. This permits to reduce the risks of being overwhelming, annoying, or making users feel a lack of control (Norman, 1994).

The behavior of the system and an effective relation with the user is based on its ability to combine in a coherent way different important features, such as the ability to navigate in the room, recognize the human presence, *communicate* with the person using *natural cues* (such as a flexible head and gaze direction, comprehensible gestures) and a pleasant *voice interaction* (good performance in speech recognition, clear speech, effective turn-taking). It is important to consider that all these aspects engage the user more than traditional assistive services. The “mimesis” that the user finds in the morphology and in the behavior of the system can arouse *emotional reactions*. Scientific studies on social robotics show that the emotional arousal is used to facilitate believable human-robot interaction (Canamero & Fredslund, 2001; Ogata & Sugano, 2000) and provide feedback to the user. *Emotions* are an important aspect of the users’ experience with interactive systems (Norman, 2004) and can trigger learning and adaptation, affecting the first reaction and the acceptance of the system.

In the KSERA project the Nao robot was chosen, apart from economical considerations, because of its humanoid shape. We now have extensive evidence that unambiguously demonstrates the huge potential a humanoid robot:

- Without a functional head with eyes, it is impossible to use gaze cues, make eye contact and add functionality of joint attention. All studies demonstrate positive effects of natural gaze cueing of social robots including attitude towards robots, user experience, usability and entertainment value.

- In a similar way it was found that human-like gestures and artificial facial expressions improve memory recall, user experience and entertainment value.

It is safe to say that a key enabler of a social robot is the fact that it has a functional head with eyes and human-like arms. Without it there is no hope of realising natural human-like non-verbal communication, which, when present, has many positive effects on HRI. It was found that head pose is sufficient for mimicking gaze direction. The robot's "eyes" may be immobile in the head and the robot does not have to "see" through them, as long as they can unambiguously portray a gaze direction. Note that pictures of eyes on a flat display can never achieve this goal because of the Mona Lisa effect [1]. Legs are less relevant for conveying social cues.

2.2 Added value of social communication through eye contact, gestures and speech

Many studies addressed the way in which non-verbal social cues should be provided in order to make HRI more natural.
We first studied whether people make eye contact with a robot in a similar way as they do with other people. For that purpose we measured in a psychophysical study region of eye contact (REC) of real people and of the Nao robot. It was shown that the REC was similar (the width and height being about 7 degrees of visual angle). This was reported in D3.1.
Next several studies addressed the role of eye contact and gestures in terms of user experience, quality and effectiveness of HRI. It was found that memory recall and attitude towards robots improved with eye contact and gestures (see section 3.1.1, and 3.1.2), but that reaction times, and classification task performance was not affected (see 3.1.3, and 3.3.1). Nor was the robot considered more persuasive (see section 3.3.3) or did it affect human gaze behaviour (section 3.3.2). Adding more dynamic interactivity of gaze (section 3.1.4) and touch (section 3.3.3), and artificial facial expressions (section 3.1.5, 3.1.6) did improve the quality of HRI.
KSERA has demonstrated that a social robot needs a functional head with eyes, because only then eye contact can be used as an effective social cue. In a similar way hand gestures and artificial facial expressions improve the quality of HRI. If these non-verbal cues are provided in a natural way they do indeed improve user experience and user acceptance.

2.3 Added value of embodiment

Given the availability of cheap wearable/carryable smart systems it begs the question how much benefit there is from an embodied robot. It was already explained above that for people to anthropomorphise with a robot a humanoid body is best, because it is the only way to convey social cues effectively. If the system just needs to provide information this advantage is less clear. It was shown that when people categorize messages delivered by either a smart home or by a robot, the robot is liked more but reaction times are similar (see section 3.1.3).
The added value of embodiment is very evident for doing health exercises. The robot can demonstrate the exercises to the user, stimulate and give feedback. All these things give rise to a positive user experience (section 3.2.1, 3.2.2). Conventional solutions cannot do this.

2.4 Added value of joint attention and context awareness

A health exercise is an example of a highly interactive task or joint task. Both robot and person are attending to each others movements. This requires monitoring the visual attention of a person to objects, limbs or the robot itself. In the latter case it is just monitoring eye contact. It was already argued that eye contact is a very important social cue that improves the quality of HRI. KSERA developed the head pose estimation algorithm (see deliverable D3.1) that allows the robot to monitor eye contact, but is also possible to use it for monitoring attention to other objects/persons in the world. In this context head pose estimation is studied and tested further (see 3.3.4), as well as the interactivity of engaging and disengaging in eye contact (section 3.3.2, 3.3.3). In addition, it was shown that speech and waving with the robot's arms are much better to ways to attract attention than eye contact and facial LED patterns (section 3.3.1). Thus, KSERA has shown very clearly that eye contact is a very powerful social cue for maintaining joint attention, but not for getting attention.

2.5 Added value of entertainment value and fun

The physical appearance of the Nao robot and the natural human-robot interaction already contribute for a large part to the entertainment value of the robot. This is evident from all studies because of the consistently high ratings in terms of likeability of the robot. Given the recent developments in 'serious gaming', where game elements are used to enhance 'serious' functionalities, such as a virtual coach for improving medication adherence. In the KSERA context the robot is already embodied, but the entertainment value of a health exercise or similar interaction may still benefit from 'serious game' elements. A study with focus groups and domain experts revealed that adding game elements is not automatically a good thing (section 3.4.2) for applications of a serious nature. On the other hand, the Nao robot is ideally suited to mimic human behviour including emotions when playing games, which provide high entertainment value (section 3.4.1)

3 KSERA research results on HRI

3.1 Natural Human Robot Interaction

The overall objective of WP3 is to make human-robot interaction simple, natural, and engaging in order to insure that a successful user-robot experience is engendered from the start. It is generally accepted that adoption of robotic systems by elderly users will only succeed if the interface promotes a valued, positive interaction from the beginning.

Here we summarize research papers on how natural human robot interaction should be designed. In particular the role of gestures, gaze behaviour, facial LED colours, interactivity and embodiment are investigated. In addition performance metrics are developed such as memory recall performance, human processing speed and attitude towards robots.

3.1.1 The effect of natural head and iconic hand gestures on message recall in human-robot interaction

Relevance to KSERA

For humans gestures and eye contact are very common social cues that aid communication between humans. For robots such an assertion is much less obvious as social cues are typically underdeveloped in robots or not present at all. This study addresses the benefit of iconic hand gestures and gaze behaviour of the robot on message recall and attitude towards robots. Message recall and attitude towards robots are two novel metrics for addressing the quality of human-robot interaction.

Research Questions

- Does the implementation of human-like head gesture result in better recall of speech messages?
- Does the implementation of iconic hand gestures result in better recall of speech messages?
- Does the implementation of both natural gaze behaviour and iconic gestures result in an improvement of recall of speech messages larger than head or hand gestures solitaire?
- A more 'natural' communication improves the overall evaluation of the robot.

Results

It was found that sentences paired with a gesture were on average better recalled than sentences without a gesture. This effect was not significant, but the results implied a trend. No main effect of ‘gaze’ (eye-contact) was found, but an interaction effect is found between the effect of gaze and gesture. The effect of gesture was found to be larger in the 'gaze' (eye-contact) condition compared to the 'no gaze' condition, and recall was best for the condition were there was a directed gaze and an iconic gesture. These results are similar to what was hypothesized.
The Godspeed questionnaire (Bartneck, 2009) was used to measure differences in attitude towards the Nao robot before and after the experiment. Overall, the Nao robot was rated positive. The robot was liked, was perceived as reasonable intelligent, and not harmful. There were only small changes in attitude caused by the exposure during the experiment. The robot was perceived on average as being a bit less animal-like after the experiment, less safe (both not significant), and significantly less intelligent. This decrease in perceived

intelligence means that the expectation of the participants towards the robots intelligence was high relative to the current abilities of the robot.

D3.4 Annex
Section 1.1. The effect of natural head and iconic hand gestures on message recall in human-robot interaction, pp. 4-36.

3.1.2 Effects of Eye Contact and Iconic Gestures on Message Retention in Human-Robot Interaction

Relevance to KSERA
This is a follow-up study based on 3.1.1. A possible confound of the content of the messages was removed, more subjects were measured and the methodological approach is much refined. The grammatical structure of the messages is taken into account and systematically analysed. This work was presented as a poster on the AAL Forum 2012 in Eindhoven, The Netherlands, and is submitted as a paper to the International Journal of Social Robotics [1].

Research Questions

- humanlike gaze behavior has a positive influence on message recall
- iconic body gestures have a positive effect on message recall and
- the gaze and gesture effects are additive in nature, meaning that the combined effect of body gestures and human-like gaze behavior is larger than either effect individually
- gestures are expected to facilitate retention by making it easier for subjects to store the content of the messages in the form of an image. In that case the positive influence of gestures on message retention is present in all parts of the presented sentences.

Results
This study investigated the effects of iconic gestures and eye contact on message retention in human-robot interaction. A humanoid robot gave short verbal messages to participants (N=24), accompanied either by iconic gestures or no gestures while making eye contact with the participant or looking away.
Results show that the use of iconic gestures indeed aids retention, but only of the verb to which the action-depicting gestures pertain. This indicates that gestures are not merged with the verbal content at encoding, but rather are stored separately and used as a mnemonic aid at retrieval. This suggests that gestures presented by robots are not necessarily processed in the same way that non-verbal cues presented by humans are. In both HHI and HRI, a message accompanied by gestures is easier to remember. In HHI, however, non-verbal cues are generally perceived subconsciously and merged with the verbal content. This does not seem to be the case in HRI. Literature suggests that participants remember a verbal message presented by a robot better when the robot looks at them more, but this result could not be replicated in our research. This indicates that the description “looking at the addressee" is an oversimplification of the relevant non-verbal cue; the setting(environment, task) largely determines how the robot's gaze behavior is interpreted. In conclusion, our research adds to the growing evidence that much is to be gained in taking an inspiration from humans' non-verbal communication when designing interactions between robots and humans.

D3.4 Annex
Section 1.2. Effects of Eye Contact and Iconic Gestures on Message Retention in Human-Robot Interaction, pp. 37-48.

3.1.3 The effect of embodiment on the intelligibility of robot messages

Relevance to KSERA
In this study a comparison is made between messages delivered by a robot and those delivered by a loudspeaker system of a smart home environment. The robotic agent is embodied: it walks and uses gestures to accompany speech, whereas the smart home is ubiquitous and omni-present. Several performance measures were used including reaction time, intelligibility, perceived usefulness and the Godspeed questionnaire. This work is partly published in [2].

Research Questions
- What is the best way to transfer information to elderly users? It is expected that the user has a more positive attitude towards the Nao robot than to the smart-home system, and that the Nao robot is experienced as more friendly and likeable.
- Does the user perform a classification task more successfully when interacting with a robot than with a smart home system?

Results
Overall the user has a more positive attitude towards the Nao robot than towards the smart home system. The robot is perceived as more animate and likeable. The intelligibility and perceived usefulness were the same for both systems. Performance of the categorization task was also similar, but people were slower to respond to the robot that to the smart home. These results confirm the positive attitude towards robots, but this does not necessarily lead to better performance of a classification task. Possibly, this will change in less ideal circumstances where noise reduces the intelligibility of messages.

D3.4 Annex
Section 1.3. The effect of embodiment on the intelligibility of robot messages, pp. 49-93.

3.1.4 Getting a better attitude towards robots by petting

Relevance to KSERA
This study examines how touch and giving rewards can further improve HRI. Participants showed picture cards to the robot and touched the robot’s head either when the answer was correct (positive feedback) or incorrect (negative feedback).

Research Questions
- Does more physical contact improve user experience and attitude towards robot
- Do people prefer positive or negative feedback and with which interaction type

Results
It was expected that participant’s attitude would improve if they had a lot of physical contact with Nao. This was true for Likeability and Perceived safety and Perceived intelligence.

Reaction times were slightly faster for positive feedback than for negative. Participants preferred speech and touching as interaction type. Petting and body language were not preferred.

D3.4 Annex
Section 1.4. Getting a better attitude towards robots by petting, pp. 94-114.

3.1.5 Imitating Human Emotions with Artificial Facial Expressions

Relevance to KSERA
The Nao robot cannot show human facial expressions, but does have the possibility to show coloured LED patterns in its eyes. This work investigates how these LED patterns can be used to mimic emotions. It is submitted as journal paper [3].

Research Questions
- What eye LED colors and patterns do users associate with the emotions: anger, disgust, fear, happiness, sadness, and surprise?
- Do users recognize emotions imitated by the robot with eye LED colors and patterns?

Results
A robot can imitate human emotions with flashing patterns of color around its eyes.

The results show that the Nao robot can use LED patterns to imitate emotions. However, the LED patterns that are recognized as an emotion were not necessarily the ones we expected. For example, LED patterns for Happiness, Sadness, and Fear were most often recognized as Surprise. This does not mean a robot cannot use flashing patterns of color around its eyes to imitate human emotions, it just means we were not very good at predicting which emotions the patterns would imitate.

Cartoon patterns appear to be more recognizable than flashing lights with varying periods and rise/fall times.

Our method of using the distance in a ROC graph can be used to determine the best one of a set of gestures or artificial facial expressions.

We used the distance from a pattern to the most recognizable point on a ROC (Receiver Operating Characteristic) graph (true-positive rate (TPR) =1, false-positive rate (FPR) = 0) to determine the best LED pattern to imitate an emotion. The same technique could be used to determine the best of any set of artificial facial expressions (e.g., avatar eyes) or gestures (e.g., shrugging shoulders) to imitate any arbitrary set of emotions, not just the ones we imitated.

None of the LED patterns we used in the experiments were recognized as Disgust. One possible explanation is gaze. In all of our experiments, the robot was looking directly at the participant (i.e., direct gaze). Research has shown that humans tend to express approach-oriented emotions, such as anger, joy, and love, more with a direct gaze, while they express avoidance-oriented emotions, such as disgust, embarrassment, and sorrow, more with an averted gaze. Combining LED patterns with an averted gaze might lead to one that was recognized as Disgust. Additionally, we might find combinations for the other emotions that are better recognized (i.e., the distance measure on the ROC graph is shorter).

D3.4 Annex

Section 1.5. Imitating Human Emotions with Artificial Facial Expressions, pp. 115-131.

3.1.6 Artificial Facial Expressions

Relevance to KSERA
This is a follow-up study to 3.1.5 Imitating Human Emotions with Artificial Facial Expressions. It addresses possible confounds and extends emotional expressions to head movements, eye shapes and mouth shapes. This work is still in progress [4].

Research Questions
- What eye LED colors do users associate with the emotions: anger, disgust, fear, happiness, sadness, and surprise?
- What eye LED patterns do users associate with the emotions: anger, disgust, fear, happiness, sadness, and surprise?
- What gaze direction do users associate with the emotions: anger, disgust, fear, happiness, sadness, and surprise?
- What mouth shape do users associate with the emotions: anger, disgust, fear, happiness, sadness, and surprise?

Results
To be provided after completion of experiment

Publications
1. Colin Lambrechts, C. and Stelma, J. (2013). Eye Color Report. Student Project Report.
2. Houben, M. (2013). Eye Pattern Report. Student Project Report.
3. Trninic, K. (2013). Eye Gaze Report. Student Project Report.
4. de Graaf, M. (2013). Mouth Shape Report. Student Project Report.

3.1.7 Towards Robust Speech Recognition for Human-Robot Interaction

Relevance to KSERA
Commanding a robot, or just communicating to it, is non-trivial for non-technicians. The humanoid appearance of the Nao robot, however, suggests that one may communicate with it like with a person. Natural speech thus becomes an important - if not the most important - communication channel between the human and the robot. While speech generation by today's robots is functional, for example, Nao's speech output is easy to understand by most people, this is not entirely true in the other direction, i.e. automatic speech recogntion (ASR) is not reliable in general. The conditions under which the user's speech arrives at the robot, such as the distance between the robot and the person, the self-generated noise by the robot, the positioning of the microphones, reverberances in an indoor room, other noise sources, etc. cannot be fully controlled in the KSERA setup and other household environments.

Research Questions
- How well does ASR perform when comparing different physical setups?
- Which ASR systems are most effective, and which are their constraints?

Results

It was found that the recognition rate was highest by a headset followed by a ceiling boundary microphone and then the microphones of the Nao robot. The latter were ineffective mainly due to the noise generated by the Nao cooling fan, which is placed too close to the microphones. Fan and the 4 microphones, reside inside the Nao head shell, which is not an optimal design. On the ASR side, finite-state grammars and statistical language models (bi-gram and tri-gram) were compared. The overall recognition rates with the Nao were insufficient, while the number of produced false positives could be kept in reasonable bounds using a multi-pass decoder. As a consequence of this study, and considering that a headset worn regularly by the user would not be acceptable, it was decided to use a bluetooth microphone either mounted on the outside of the Nao head, or placed on a table close to the user.

D3.4 Annex
Section 1.6. Towards Robust Speech Recognition for Human-Robot Interaction, pp.132-138.

3.2 Embodied Health Exercise

An important part of KSERA is the health exercise scenario. The following studies address questions on how to monitor human performance, how to improve attitude towards robots and how to provide feedback to users. As such they address the main WP3 objectives of improving user experience and designing a user interface for easy and fun user-robot interaction.

3.2.1 Movement exaggeration of an embodied agent as health exercise performance enhancer

Relevance to KSERA
When performing a health exercise the robot's embodied nature make it ideally suited to instruct the user by showing the exercise to user. This study addresses how movement amplitude and frequency can be used to manipulate user experience.

Research Questions

- Is increasing the robot's amplitude and/or speed a good performance enhancer? Performance in this context means maintaining arms in a desired range of amplitude or speed and improved attitudes towards robots.
- Is there an inverse relationship between frequency and amplitude of the participants' movements

Results
Participants liked robots more after the experiment, indicating that exercising with a robot improves attitudes towards it. However, participants perceived robots as less intelligent after the experiment in comparison to before. The results also confirmed that participants had no trouble in following the robot's varying movement amplitudes and frequencies.

For one exercise a negative correlation between amplitude and frequency was found. Interestingly, participants rated every exercise and all robot speed-amplitude stimuli as not tiring. This means that physical effort can be varied without affecting perceived tiredness.

Publications

D3.4 Annex
Section 2.1. Movement exaggeration of an embodied agent as health exercise performance enhancer, pp. 140-180.

3.2.2 Improving Health Exercise Performance With An Embodied Agent Using Verbal Feedback

Relevance to KSERA
When performing a health exercise the robot also needs to monitor performance and give feedback. This study addresses how positive and negative feedback affect performance and user experience.

Research Questions

- What is the effect of positive and negative feedback of an embodied humanoid robot during a health exercise on the performance and user experience of a user.
- Is the amplitude-frequency ratio a good measure for perceived effort?

Results
Positive feedback resulted in better physical performance than negative feedback for the low exhausting movements. Participants performed equally well for both exhausting and non-exhausting movements. For the most exhausting movements giving no feedback resulted in best performance. Thus, when people have energy to spare, an embodied agent should motivate them with positive feedback to make them go that extra mile. However, in order to achieve the best physical performance when people are already exercising at their physical maximum, the best way to go is to not let the robot disturb them with any kind of verbal feedback.
In terms of user experience it was found that positive feedback makes exercising somewhat more interesting than negative feedback. Positive feedback also made participants perceive the exercise as less tiring than negative feedback.
Finally, it was found that user experience was enhanced by the verbal feedback given by the humanoid robot. The Godspeed questionnaire suggested that people perceived the robot as more lifelike, likeable and safe after they exercised with it.

D3.4 Annex
Section 2.2. Improving Health Exercise Performance With An Embodied Agent Using Verbal Feedback, pp. 181-215.

3.3 Improve the user experience by facilitating joint attention between the robot and the user

Human gaze behaviour is a very important social cue as it signals the person's visual attention. Robots typically do not exhibit this behaviour because their visual acuity is the same everywhere in their visual field. However, if a robot does show similar gaze behaviour, people automatically interpret this as a social cue for visual attention. In addition, this is a crucial enabler for joint attention and eye contact. The research reports in this section are investigating how to monitor, attract and keep attention. In addition human gaze behaviour is studied so that it can be used to design a mental model for the robot that enables it to infer which objects a person is attending to. These studies mainly address the WP3 objective of improving the user experience by facilitating joint attention between the robot and the user.

3.3.1 How can a robot attract the attention of its human partner? A comparative study over different modalities for attracting attention

Relevance to KSERA
This study addresses the question how the robot should attract attention in terms of efficiency and user experience. This work was presented as a talk at the International Conference of Social Robotics (ICSR 2012) in Chengdu, China, and is published as a conference paper at the ICSR 2012 [5]. The work won the ‘Best paper (finalist)’ award, and it was selected for the special issue of the IJSR due in 2013.

Research Questions

- What is the best way for a robot to attract the attention of its human partner?
- Is cognitive load an important factor for the way in which to attract attention?

Results
Participants watched news items on a TV, some of which were marked as important, and had to answer a few questions about them afterwards. During some news items the robot tried to attract the attention. Reaction time and user experience were measured.
It was found that reaction times for speech were shortest, followed by the waving gesture and eye LEDs blinking gesture. Trying to establish eye-contact was slowest. It is interesting to note that speaking and waving both involve an auditory component either in the form of speech utterance or in the form of noise produced by the robot's actuators, whereas blinking the eye LEDs and establishing eye contact are of a purely visual nature. Thus, it seems that sound, which is an omni-directional cue, is more salient than the visual channel.
No significant difference was found in terms of reaction time when participants were presented with important and not-important news items.
Finally, waving was the most clear, present and most-liked action closely followed by speech, while eye contact was least appreciated. This suggests that communicative cues for attracting attention should be least ambiguous and as specific as possible.

D3.4 Annex
Section 3.1. How can a robot attract the attention of its human partner? A comparative study over different modalities for attracting attention, pp. 217-227.

3.3.2 Influence of the gazing behaviour of a robot on the gazing behaviour of a human

Relevance to KSERA
This study addresses the effect of the robot’s gaze behaviour on that of the human partner and its influence on persuasiveness of a robot. An eye tracker was used to measure human gaze behaviour.

Research Hypotheses

- The more the robot gazes at the eyes of the participant, the more the participant will gaze at the eyes of the robot.

- The more the robot gazes at the eyes of the participant, the more details the test-subject will remember about the story the robot is telling, and the stronger the persuasive power of the story becomes.

Results
The results provided no evidence in support of our hypotheses. The results show that there was no influence of robot gazing on participant gazing, nor on the persuasive power of the robot. The head movements of participants had to be constrained because of the eye tracker, and the participants' task was to listen to stories. This could explain why no effect of robot's gaze behaviour was found. The artificial speech synthesis was rated as unnatural, which may explain the absence of persuasive power.

D3.4 Annex
Section 3.2. Influence of the gazing behaviour of a robot on the gazing behaviour of a human, pp. 228-251.

3.3.3 The effect of gaze behavior on the attitude towards humanoid robots

Relevance to KSERA
Making a robot's gaze behaviour more human-like should have a positive influence on the mental well-being of the user. Therefore, this study compares different gaze behaviours of the Nao robot in terms of its interactivity. The 'desert survival task' was used for measuring persuasiveness.

Research Questions
How do different types of gaze behaviour influence the attitude towards robots? It is expected that:

- When a humanoid robot applies looking-while-listening gaze behaviour, positive attitude towards the robot and persuasion by the robot will be higher.
- When a humanoid robot reacts to gaze behaviour of a person (dynamic condition), positive attitude towards the robot and persuasion by the robot will be higher.
- Looking-while-listening and dynamic gaze behaviour will have an additive effect on the positive attitude towards the robot and persuasion by the robot.

Results
The results indicated that when the robot applied looking-while-listening gaze behavior, positive attitude towards the robot and persuasion by the robot (first hypothesis) were not significantly higher. This is in contradiction with findings in human-human interaction by (Argyle, Lefebvre, & Cook, 1974).
When the robot dynamically reacted to gaze behavior of a person, the attitude towards the robot was more positive, but not significantly so. However it did show a trend.
No effects were found on the persuasiveness of the robot.

D3.4 Annex
Section 3.3. The effect of gaze behavior on the attitude towards humanoid robots, pp. 252-262.

3.3.4 Inferring attended objects from estimated headpose

Relevance to KSERA
This study measures human head poses when looking at various objects in the environment. This information is necessary for constructing a model human visual attention and joint attention. A publication is in preparation [6].

Research Question
- Can head pose estimation be used to recognize gaze direction?

Results
Lab experiments showed that the yaw and pitch could be used to predict the user's gaze direction with a first order linear equation.

D3.4 Annex
Section 3.4. Inferring attended objects from estimated headpose, pp. 263-274.

3.3.5 Hybrid Ensembles Using Hopfield Neural Networks for Robust Face Detection

Relevance to KSERA
An important task for human-robot-interaction is to recognize the awareness of the user, i.e. to assess whether the human looks to the robot. This perception of the user's attention awareness is solved through face detection and head-pose estimation. Robust face decion is important because it is the requirement for the following head-pose estimation. The pre-trained and state of the art, open source (OpenCV) implementation of Viola and Jones Adaboost algorithm fails in some cases, e.g. in certain light conditions. Therefore we created a hybrid solution using ensembles of Hopfield Neural Networks to improve the accuracy of the face detector.

Research Questions
- In how far can we improve the accuracy of a face detection system by increasing the diversity through using additional Hopfield Neural Networks within an Adaboost ensemble?
- Could the combination of different image preprocessing methods improve the accuracy of a face detection system?

Results
By comparing with ensembles that are trained within the same constraints and conditions, we have shown in the first study that a hybrid solution using different classification methods has the ability to improve the classification accuracy. This hybrid consists of the threshold classifier used in the original work of Viola and Jones and of Hopfield Neural Networks using Haar-like features. In the second study, we have shown that accuracy can by improved by using different, but fast and easy to use, image pre-processing methods on grey-valued images. These filters were applied to the original image and then combined for the parallel and simultaneous usage in Hopfield Neural Networks, but also as an average combining for the usage in threshold classifiers. Our experiments show an improvement to the state of the art in face detection.

D3.4 Annex
Section 3.5. Hybrid Ensembles Using Hopfield Neural Networks, pp. 275-283.
Section 3.6. Adaboost and Hopfield Neural Networks on Different Image Representations for Robust Face Detection, pp. 284-290.

3.3.6 Object learning with natural language

Relevance to KSERA
A robot that interacts with a person should ideally not only be able to recognise the person, but also objects in the surround. In KSERA, it was studied to recognise the pulseoximeter, which is specifically important for COPD patients to monitor their health status. In a more general situation, however, important objects may not be known in advance and cannot be hard-coded into the system. Hence, the (primary or secondary) user will have to teach the robot to recognise the object. Since the user is not trained to use a programming interface, the same natural language dialog system used also for the KSERA interaction with the primary user has been used in this study. Together with head pose estimation and attended object inference, learning object recognition provides the basis for context aware joint attention. This work is published as conference paper of CSIP 2012 [7]

Research Question
- How can we bridge the gap between the needs of a natural human robot interaction and the capabilities of current humanoid platforms?

Results

A working system that learns objects in real time by dialog with a user was presented. It was tested with students and at a University Open Day (EXPO). First, observations about the implementation and robustness of visual object recognition were made. Hybrid ensembles of neural learning methods have been shown effective, and their effectiveness was increased with the diversity of features. Colour- and texture-based features for object recognition methods are effective for discrimination between objects. Plain objects with little features, such as a drinking glass, are hard to recognise with a single camera. It could be concluded that objects that are integral parts in a smart home should be designed to be easily recognisable by robotic vision. Currently, this leads to only subtle design constraints, such as high visual contrasts on the object.

Second, observations about the speech interface were made. The running system was easy to use. A "teaching" dialog allowed to show an object to the robot and say the object's name so that the robot learnt it; a "classification" and a "find" dialog allowed easy testing, at which the robot told about a shown object's identity ("classification"), or found an object in its visual field and then pointed to it ("find"). However, to unexperienced laymen, the system was not easy to use. First, participants needed to be instructed to use the respective dialogs correctly. Second, voice recognition was unreliable with certain speakers, such as children."

D3.4 Annex
Section 3.7 Object Learning with Natural Language in a Distributed Intelligent System - A Case Study of Human-Robot Interaction, pp. 291-299.

3.4 Design and develop a user interface for easy and fun user-robot interaction

One of the main WP3 objectives is to Design and develop a user interface for easy and fun user-robot interaction. For a large part this involves making the robot's behaviour more natural and interactive (see sections 3.1 – 3.3). However, what constitutes entertainment value remains elusive. Recent developments in the domain of 'serious gaming' show the potential of games to improve the playfulness and entertainment value of a wide variety of services for elderly. The studies reported here address the entertainment value of the humanoid Nao robot.

3.4.1 Exploring the Entertainment Value of the Nao Robot

Relevance to KSERA

This study explores ways to improve the entertainment value of HRI. This work is being prepared for publication [7].

Research Questions

- Do users recognize emotions imitated by the robot through speech, gestures, and artificial facial expressions?
- Is the user-robot interaction fun?

Results

- Findings of a preliminary study to determine whether designed behavioral patterns of speech, gesture, and LED eye patterns were adequate to express the intended levels of confidence and surprise provided strong evidence that the happy and sad gestures were indeed perceived as a positive and negative emotion respectively and that adding a happy and sad LED eye pattern to the gesture even amplified the emotional expression of the robot.
- Subjective measurements of the player experience indicate that positive affect ratings of the entertainment value of the game on both the PANAS and GEQ questionnaires were comparatively high for all measurement blocks. Negative affect, in contrast, received very low ratings on both questionnaires.
- As an objective measure of the entertainment experienced by participants, the participants' laughs, in reaction to the interactive behavior the robot was expressing, were counted.
  - 14 out of 19 participants (74%) laughed at least twice during the whole experiment.
  - Participants laughed primarily when the robot was expressing one of the emotional behavioural patterns (i.e., surprise, happiness, displeased, sad, and the little victory dance which was performed by the robot whenever it won a game).
- After finishing the experiment, eight out of the 19 participants (42%) informed the experimenter on their own initiative that playing the game with the robot was fun. Among these eight people were even three who belonged to the group of participants that didn't laugh once during the whole experiment.

D3.4 Annex

Section 4.1. Exploring the Entertainment Value of the Nao Robot, pp. 301-349.

3.4.2 Games and playful interaction to support Senior-robot interaction

Relevance to KSERA
This project addressed playful interaction of socially assistive robots using focus groups and interviews with robot experts, care professionals and seniors.

Research Questions
- To learn what seniors think about robots, what they associate with the concept "robot" and what their current multimedia or technology usage is, some focus groups with seniors were conducted. Additionally, a discussion was held with some caregivers to learn about their view of seniors and the usage of robotics in care centres or the seniors own living environment.
- The goal of interviews was to consult the experts about the needs of seniors in regard to their projects, tasks that seniors experienced difficulties with, the caregivers (formal or informal) who assist these seniors, the most striking facts noticed in senior-robot interaction affecting the robot acceptance and finally in what way a playful interaction can improve the acceptance of a robot in a seniors environment.

Results
From the focus groups with seniors it can be said that there is some interest in a playful robot instead of a more serious one, while at the same time the three interviewed experts agree that playful interaction would help raise acceptance. However, research has shown that playful interaction has its costs on the cooperation between humans and robots and more specifically, a much lower cooperation to the robot's requests was observed (Goetz & Kiesler, 2002). Taking this into consideration, and the fact that a robot deployed as health assistant for seniors should achieve cooperation during serious tasks, we can assume that playful interaction level should be adjusted based on how serious the task is.

D3.4 Annex
Section 4.2. Games and playful interaction to support Senior-robot interaction, pp. 350-382.

4 References

[1] Todorović, D. (2006). Geometrical basis of perception of gaze direction. Vision Research, 46(21), 3549–3562. doi:10.1016/j.visres.2006.04.011
[2] Dijk, E.T. Van, Cuijpers, Raymond H., and Torta, Elena (submitted). Effects of Eye Contact and Iconic Gestures on Message Retention in Human-Robot Interaction. International Journal of Social Robotics.
[3] Torta, E., Oberzaucher, J., Werner, F., Cuijpers, R.H., Juola, J.F. (2012). The attitude toward socially assistive robots in intelligent homes: results from laboratory studies and field trials. Journal of Human-Robot Interaction 1(2), 76-99. DOI: 10.5898/JHRI.1.2.Torta.
[4] Johnson, D. O., Cuijpers. R.H., & van der Pol, D. (2013). Imitating Human Emotions with the Nao Robot's Eye LEDs. International Journal of Social Robotics (in review).
[5] Johnson, D. O. and Cuijpers, R.H. (2013). Imitating Human Emotions with a Robot's Eye LED Color, Eye LED Pattern, Gaze Direction, and Mouth Shape. (in preparation).
[6] Torta, E., Heumen, J. van, Cuijpers, R. H., & Juola, J. F. (2012). How Can a Robot Attract the Attention of Its Human Partner? A Comparative Study over Different Modalities for Attracting Attention. In S. S. Ge, O. Khatib, J.-J. Cabibihan, R. Simmons, &

M.-A. Williams (Eds.), Social Robotics (pp. 288–297). Springer Berlin Heidelberg. DOI: 10.1007/978-3-642-34103-8_29

[7] Johnson, D. O. and Cuijpers, R.H. (2013). Predicting Gaze Direction from Head Pose Yaw and Pitch. (in preparation).

[8] Heinrich, S., Folleher, P., Springstübe, P., Strahl, E., Twiefel, J., Weber, C., and Wermter, S. (2012) Object Learning with Natural Language in a Distributed Intelligent System - A Case Study of Human-Robot Interaction. In: IEEE First International Conference on Cognitive Systems and Information Processing (CSIP 2012), in press (Springer).

[9] Johnson, D. O., Cuijpers, R.H., and Pollmann, K. (2013). Exploring the Entertainment Value of the Nao Robot. (in preparation).

[10] Heinrich, S., Wermter, S. (2011). Towards Robust Speech Recognition for Human-Robot Interaction. Proceedings of the IROS2011 Workshop on Cognitive Neuroscience Robotics (CNR), pp. 29-34, San Francisco, CA, USA, September 2011.

[11] Meins, N., Jirak, D., Weber, C., Wermter, S. (2012). Adaboost and Hopfield Neural Networks on Different Image Representations for Robust Face Detection. Proceedings of the 12th International Conference on Hybrid Intelligent Systems (HIS 2012), pp. 531-536, IEEE. Pune, IN, December 2012.

[12] Meins, N., Wermter, S., Weber, C. (2012). Hybrid Ensembles Using Hopfield Neural Networks. Proceedings of the 22nd International Conference on Artificial Neural Networks (ICANN 2012), LNCS 7552, pp. 403-410, Springer Heidelberg. Lausanne, CH, September 2012.

5 Annex: Research Reports

The annex is provided as a separate document because of its large size and restricted dissemination level.