First International Workshop on
Human Aspects in Ambient Intelligence

Tibor Bosse, Cristiano Castelfranchi, Mark Neerincx,

Fariba Sadri, Jan Treur (eds.)

Darmstadt, Germany, November 10, 2007

Workshop at the European Conference on Ambient Intelligence

Vrije Universiteit Amsterdam
Agent Systems Research Group
Department of Artificial Intelligence

Coordination Committee

Tibor Bosse	(Vrije Universiteit Amsterdam, Agent Systems Research Group)
Cristiano Castelfranchi	(CNR Rome, Institute of Cognitive Sciences and Technologies)
Mark Neerincx	(TNO Human Factors; Technical University Delft, Man-Machine Interaction)
Fariba Sadri	(Imperial College, Department of Computing)
Jan Treur	(Vrije Universiteit Amsterdam, Agent Systems Research Group)

Program Committee

Gerhard Andersson	(Linköping University, Department of Behavioural Sciences)
Juan Carlos Augusto	(University of Ulster, School of Computing and Mathematics)
Tibor Bosse	(Vrije Universiteit Amsterdam, Agent Systems Research Group)
Antonio Camurri	(University of Genoa, InfoMus Lab)
Nick Cassimatis	(Rensselaer Polytechnic Institute, Cognitive Science Department)
Cristiano Castelfranchi	(CNR Rome, Institute of Cognitive Sciences and Technologies)
James L. Crowley	(INRIA Rhone-Alpes, Perception and Integration for Smart Spaces Group)
Pim Cuijpers	(Vrije Universiteit Amsterdam, Clinical Psychology)
Henk Elffers	(Institute for Criminology and Law; Antwerp University, Faculty of Law)
Rino Falcone	(CNR Rome, Institute of Cognitive Sciences and Technologies)
Dirk Heylen	(University of Twente, Human Media Interaction)
Ingrid Heynderickx	(Philips Research Netherlands)
Anthony Jameson	(DFKI, Human-Computer Interaction)
Paul Lukowicz	(Austrian University for Health Sciences, Medical Informatics and Technology)
Isaac Marks	(King's College London, Institute of Psychiatry/Maudsley Hospital)
Silvia Miksch	(Danube University Krems, Dept of Information and Knowledge Engineering)
Scott Moss	(Manchester Metropolitan University, Centre for Policy Modelling)
Mark Neerincx	(TNO Human Factors; Technical University Delft, Man-Machine Interaction)
Fariba Sadri	(Imperial College, Department of Computing)
Matthias Scheutz	(University of Notre Dame, Artificial Intelligence and Robotics Laboratory)
Elizabeth Sklar	(City University of NY, Brooklyn College, Dept of Comp and Inf Science)
Ron Sun	(Rensselaer Polytechnic Institute, Cognitive Science Department)
Jan Treur	(Vrije Universiteit Amsterdam, Agent Systems Research Group)
Robert L. West	(Carleton University, Department of Cognitive Science)

Table of Contents

Preface p. 4

Jan Treur p. 5
On Human Aspects in Ambient Intelligence

Gangadhar Garipelli, Ferran Galán, Ricardo Chavarriaga, Pierre W. Ferrez, Eileen Lew, José del R. Millán p. 11
IDIAP BCI Research Meets Ambient Intelligence: Designing Intelligent Interaction

Peter-Paul van Maanen, Lisette de Koning, Kees van Dongen p. 29
Design and Validation of HABTA: Human Attention-Based Task Allocator

Jianxin (Roger) Jiao, Qianli Xu, Jun Du p. 45
Affective Human Factors Design with Ambient Intelligence

Anne-mie Sponselee, Ben Schouten, Don Bouwhuis, Charles Willems p. 59
Smart Home Technology for the Elderly: Perceptions of Multidisciplinary Stakeholders

Tina Mioch, Maaike Harbers, Willem A. van Doesburg, Karel van den Bosch p. 73
Enhancing Human Understanding through Intelligent Explanations

José Bravo Rodriguez, Ramón Hervás, Salvador W. Nava, Gabriel Chavira, Carlos Sánchez p. 85
Towards Natural Interaction by Enabling Technologies: A Near Field Communication Approach

Tibor Bosse, Fiemke Both, Charlotte Gerritsen, Mark Hoogendoorn, Jan Treur p. 103
Model-Based Reasoning Methods within an Ambient Intelligent Agent Model

Preface

This workshop addresses multidisciplinary aspects of Ambient Intelligence with human-directed disciplines such as psychology, social science, neuroscience and biomedical sciences. The aim is to get people together from these disciplines or working on cross connections of Ambient Intelligence with these disciplines. The focus is on the use of knowledge from these disciplines in Ambient Intelligence applications, in order to take care of and support in a knowledgeable manner humans in their daily living in medical, psychological and social respects. The workshop can play an important role, for example, to get modellers in the psychological, neurological, social or biomedical disciplines interested in Ambient Intelligence as a high-potential application area for their models, and, for example, get inspiration for problem areas to be addressed for further developments in their disciplines. From the other side, the workshop may make researchers in Computer Science, and Artificial and Ambient Intelligence more aware of the possibilities to incorporate more substantial knowledge from the psychological, neurological social and biomedical disciplines in Ambient Intelligence architectures and applications, and may offer problem specifications that can be addressed by the human-directed sciences.

On Human Aspects in Ambient Intelligence

Jan Treur

Agent Systems Research Group
Department of Artificial Intelligence, Vrije Universiteit Amsterdam
De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands
Email: treur@cs.vu.nl
URL: http://www.few.vu.nl/~treur

Abstract. This paper briefly outlines the scientific area that addresses Ambient Intelligence applications in which not only sensor data, but also knowledge from the human-directed sciences such as biomedical science, neuroscience, and psychological and social sciences is incorporated. This knowledge enables the environment to perform more in-depth, human-like analyses of the functioning of the observed humans, and to come up with better informed actions. It is discussed which ingredients are important to realise this view, and how frameworks can be developed to combine them to obtain the intended type of systems: reflective coupled human-environment systems. Finally, further perspectives are discussed for Ambient Intelligence applications based on these reflective coupled human-environment systems.

1 Introduction

Ambient Intelligence provides possibilities to contribute to more personal care; e.g., (Aarts, Harwig, Schuurmans, 2001; Aarts, Collier, Loenen, Ruyter, 2003; Riva, Vatalaro, Davide, Alcañiz, 2005). Acquisition of sensor information about humans and their functioning is an important factor, but without adequate knowledge for analysis of this information, the scope of such applications is limited. However, devices in the environment possessing such knowledge can show a more human-like understanding and base personal care on this understanding. For example, this may concern elderly people, patients depending on regular medicine usage, surveillance, penitentiary care, psychotherapeutical/self-help communities, but also, for example, humans in highly demanding tasks such as warfare officers, air traffic controllers, crisis and disaster managers, and humans in space missions; e.g., (Green, 2005; Itti and Koch, 2001).

Within human-directed scientific areas, such as cognitive science, psychology, neuroscience and biomedical sciences, models have been and are being developed for a variety of aspects of human functioning. If such models of human processes are represented in a formal and computational format, and incorporated in the human environment in devices that monitor the physical and mental state of the human, then such devices are able to perform a more in-depth analysis of the human's functioning. This can result in an environment that may more effectively affect the state of humans by undertaking in a knowledgeable manner actions that improve their wellbeing and performance. For example, the workspaces of naval officers may include systems that, among others, track their eye movements and characteristics of incoming stimuli (e.g., airplanes on a radar screen), and use this information in a computational model that is able to estimate where their attention is focussed at. When it turns out that an officer neglects parts of a radar screen, such a system can either indicate this to the person, or arrange on the background that another person or computer system takes care of this neglected part. In applications of this type, an ambience is created that has a better understanding of humans, based on computationally formalised knowledge from the human-directed disciplines.

2 Multidisciplinarity: the Ingredients

The area as sketched is essentially multidisciplinary. It combines aspects of Ambient Intelligence with knowledge from human-directed disciplines such as psychology, social science, neuroscience and biomedical sciences. Further development will depend on cooperation between researchers from these disciplines or working on cross connections of Ambient Intelligence with the human-directed disciplines. The focus is on the use of knowledge from these disciplines in Ambient Intelligence applications, in order to take care in a more sophisticated manner of humans in their daily living in medical, psychological and social respects. For example, modellers in the psychological, neurological, social or biomedical disciplines interested in Ambient Intelligence as a high-potential application area for their models, can get inspiration for problem areas to be addressed for further developments in their disciplines. From the other side, researchers in Computer Science, and Artificial and Ambient Intelligence may become more aware of the possibilities to incorporate more substantial knowledge from the psychological, neurological, social and biomedical disciplines in Ambient Intelligence architectures and applications, and may offer problem specifications that can be addressed by the human-directed sciences.

In more detail, content from the domain of human-directed sciences, among others, can be taken from areas such as medical physiology, health sciences, neuroscience, cognitive psychology, clinical psychology, psychopathology, sociology, criminology, and exercise and sport sciences. From the domain of Artificial Intelligence, useful contributions can be found in areas such as agent modelling, knowledge and task modelling, and cognitive and social modelling and simulation. Finally, from the Computer Science domain, relevant areas are distributed systems, sensor systems, human-centred software engineering, user modelling, and human-computer interaction.

3 Frameworks to Combine the Ingredients

One of the challenges is to provide frameworks that cover the class of Ambient Intelligence applications showing human-like understanding and supporting behaviour. Here human-like understanding is defined as understanding in the sense of being able to analyse and estimate what is going on in the human's mind (a form of mindreading) and in his or her body (a form of bodyreading). Input for these processes are observed information about the human's state over time, and dynamic models for the human's physical and mental processes. For the mental side such a dynamic model is sometimes called a Theory of Mind (e.g., Baron-Cohen, 1995; Dennett, 1987; Gärdenfors, 2003; Goldman, 2006) and may cover, for example, emotion, attention, intention, and belief. Similarly for the human's physical processes, such a model relates, for example, to skin conditions, heart rates, and levels of blood sugar, insulin, adrenalin, testosterone, serotonin, and specific medicines taken. Note that different types of models are needed: physiological, neurological, cognitive, emotional, social, as well as models of the physical and artificial environment.

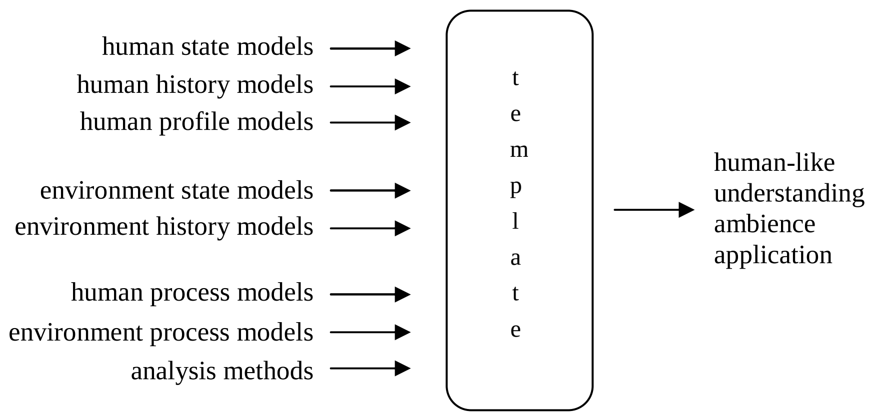


Figure 1 Framework to combine the ingredients

A framework can be used as a template for the specific class of Ambient Intelligence applications as described. The structure of such an ambient software and hardware design can be described in an agent-based manner at a conceptual design level and can be given generic facilities built in to represent the following (see also Figure 1):

- human state and history models
- environment state and history models
- profiles and characteristics models of humans
- ontologies and knowledge from biomedical, neurological, psychological and/or social disciplines
- dynamic process models about human functioning
- dynamic environment process models
- methods for analysis on the basis of such models

Examples of such analysis methods are voice and skin analysis with respect to emotional states, gesture analysis, heart rate analysis. The template can include slots where the application-specific content can be filled to get an executable design for a working system. This specific content together with the generic methods to operate on it, provides a reflective coupled human-environment system, based on a tight cooperation between a human and an ambient system to show human-like understanding of humans and to react from this understanding in a knowledgeable manner.

4 Perspectives of Reflective Coupled Human-Environment Systems

Ambient Intelligence applications in general can be viewed as coupled human-environment systems, where 'coupled' means mutually interacting. For the specific type of applications considered here, however, the coupling takes two different forms; see also Figure 2.

- On the one hand the coupling takes place as interaction between human and environment, as in any Ambient Intelligence application:
  - the environment gets information generated by the human as input, and
  - the human gets information generated by the environment as input.
- In addition, coupling at a more deep, reflective level takes place due to the fact that
  - the environment has and maintains knowledge about the functioning of the human, the environment and their interaction, and
  - the human has and maintains knowledge about functioning of him or herself, the environment, and their interaction

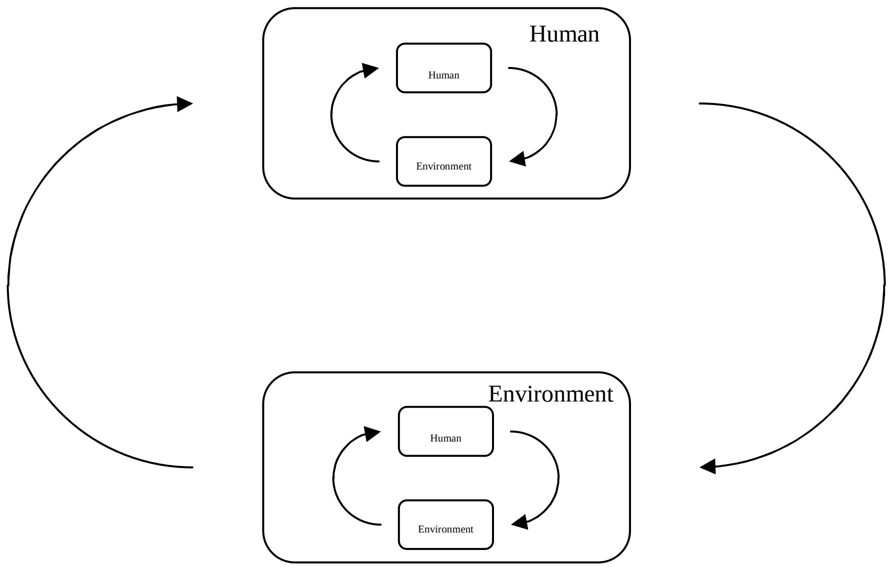


Figure 2 Reflective coupled human-environment systems

So, in such a more specific human-environment system, being coupled does not only mean that the human and its environment interact, but also that they have knowledge, understanding and awareness of each other, themselves and their interaction. This entails two types of awareness:

- *Human awareness*: awareness by the human about the human and environmental processes and their interaction
- *Technological awareness*: awareness by the environment about the human and environmental processes and their interaction

By (human and technological) learning, adaptation and development processes for both the human and the environment, such awareness can also grow over time.

These reflective coupled human-environment systems can have a positive impact at different aggregation levels, from individual via an organisation within society to the society as a whole:

- *Individual level*
    - more effective functioning
    - stimulating healthy functioning and preventing health problems to occur
    - support of learning and development
- *Organisation level*
    - efficient functioning organisation by wellfunctioning members
    - learning and adaptation of the organisation
- *Society level*:
    - limiting costs for illness and inability to work
    - efficient management of environment

Some more specific examples of today's societal challenges, to which reflective coupled human-environment systems can contribute, are elderly care, health management, crime and security.

5 Conclusion

The scientific area that addresses Ambient Intelligence applications in which knowledge from the human-directed sciences is incorporated, has a high potential to provide nontrivial Ambient Intelligence applications based on human-like understanding. Such understanding can result in better informed actions and will feel more natural for humans. Important additional ingredients to realise this view are provided by areas in Computer Science, Artificial Intelligence and Cognitive Science; among others: agent modelling, knowledge and task modelling, user modelling, and cognitive modelling. Furthermore integrative frameworks can be developed to combine the ingredients. The resulting human-environment systems are coupled not only by their mutual interaction, but also in a reflective manner in the sense that both the human and the ambient system have and/or develop a model of the interactive processes of the human and the environment. These reflective coupled human-environment systems are an interesting type of systems to be studied scientifically, and provide a solid foundation for human-like Ambient Intelligence applications with significant benefits for individuals, organisations, and the society as a whole.

Acknowledgements

The author is grateful to a number of colleagues that in discussions have deepened insight in the area, among whom Tibor Bosse, Frank van Harmelen, Mark Hoogendoorn, Johan Hoorn, Michel Klein, Peter-Paul van Maanen, Andre Spijkervet, Maarten van Steen, Gerrit van der Veer, and Chris Verhoef.

References

1. Aarts, E.; Collier, R.; van Loenen, E.; Ruyter, B. de (eds.) (2003). *Ambient Intelligence. Proc. of the First European Symposium, EUSAI 2003*. Lecture Notes in Computer Science, vol. 2875. Springer Verlag, 2003, pp. 432.
2. Aarts, E., Harwig, R. , and Schuurmans, M. (2001), Ambient Intelligence. In: P. Denning (ed.), *The Invisible Future*, McGraw Hill, New York, pp. 235-250.
3. Baron-Cohen, S. (1995). *Mindblindness*. MIT Press
4. Dennett, D.C. (1987). *The Intentional Stance*. MIT Press. Cambridge Mass.
5. Gärdenfors, P. (2003), *How Homo Became Sapiens: On The Evolution Of Thinking*. Oxford University Press, 2003.
6. Goldman, A.I. (2006). *Simulating Minds: The Philosophy, Psychology and Neuroscience of Mind Reading.* Oxford University Press.
7. Green D. J. (2005). Realtime Compliance Management Using a Wireless Realtime Pillbottle – A Report on the Pilot Study of SIMPILL. In: *Proc. of the International Conference for eHealth, Telemedicine and Health, Med-e-Tel'05*, 2005, Luxemburg.
8. Itti, L. and Koch, C., Computational Modeling of Visual Attention, *Nature Reviews Neuroscience*, Vol. 2, No. 3, 2001, pp. 194-203.
9. Riva, G., F. Vatalaro, F. Davide, M. Alcañiz (eds). (2005). *Ambient Intelligence*. IOS Press, 2001.

IDIAP BCI Research Meets Ambient Intelligence: Designing Intelligent Interaction

Gangadhar Garipelli[12*], Ferran Galán[13], Ricardo Chavarriaga[1] Pierre W. Ferrez[12], Eileen Lew[12], and José del R. Millán[12]

[1] IDIAP Research Institute, Martigny, Switzerland.
[2] Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland.
[3] University of Barcelona, Barcelona, Spain.

Abstract. This paper is aimed to introduce IDIAP Brain Computer Interface (IBCI) research that successfully applied Ambience Intelligence (AmI) principles in designing intelligent brain-machine interactions. We proceed through IBCI applications describing how machines can decode and react to the human mental commands, cognitive and emotive states. We show how effective human-machine interaction for brain computer interfacing (BCI) can be achieved through, 1) asynchronous and spontaneous BCI, 2) shared control between the human and machine, 3) online learning and 4) the use of cognitive state recognition. Identifying common principles in BCI research and ambiance intelligence (AmI) research, we discuss IBCI applications. With the current studies on recognition of human cognitive states, we argue for the possibility of designing empathic environments or devices that have a better human like understanding directly from brain signals.

1 Motivation

Brain Computer Interfacing (BCI) or Brain Machine Interfacing (BMI) refers to interaction with devices, where user's intentions represented as several brain states are deciphered and translated into actions without requiring any physical action [44] [25] [21]. There is a growing interest in the use of brain signals for communicating and operating devices, which is facilitated by the advances in the the measurement technologies in the past decades. As BCI bypasses the classical neuromuscular communication channels, this technology is intended to use for rehabilitation of tetraplegic or paraplegic patients to improve their communication, mobility and independence. The BCI research also opens up new possibilities in natural interaction for able-bodied people (e.g., for space applications, where environment is inherently hostile and dangerous for astronauts, who could greatly benefit from direct mental teleoperation of external semi-automatic manipulators [26], and for entertainment applications like multimedia gaming [20]). Typical applications of BCI are communication aids such as spelling devices [5] [31] [25] and mobility aids such as wheelchair [41]. In the current paper, we

* Gangadhar.Garipelli@idiap.ch

introduce the design of IDIAP BCI (IBCI) towards intelligent interaction and empathic devices and show how its key features are consistant with Ambient Intelligence (AmI) design principles.

The vision of AmI is designing digital environments in which the electronics devices are sensitive to people's needs, personalized to their requirements, anticipatory of their behavior and responsive to their presence. The main goals of AmI based interaction are, 1) understanding human function and sensory information, 2) analysis of sensory information, 3) design of human-like-reasoning devices and intelligent interaction. The BCI research brings a new sensing modality for tracking neurophysiological information related human's cognitive and emotive sates. The brain computer interaction can be seen also as a new mode of interaction similar to speech and vision based interaction, but with mental commands directly from the brain signals. So the general principles of interaction for destining AmI are also valid for the design of interaction with BCI.

Since we are interested in designing interactions between intelligent systems (human user and intelligent machine), it is natural to derive ideas from human-human communication. These ideas offer a starting point for a quest for new forms of interaction. According to Schmidt [39], for interaction, the context information is important. In particular, the key components are,

- *Shared knowledge*: In communication between intelligent systems, a common knowledge is essential for understanding each other. This common knowledge is extensive and is usually not explicitly mentioned. In most cases, this common knowledge includes world or environment model.
- *Communication error recovery*: Communication between intelligent systems may not be error free, many conversations include short term misunderstandings and ambiguities. But the misunderstandings are often detected by monitoring the response of the communication partner. In case of misinterpretation the dialogs are repeated and corrected.
- *Surrounding situation and context*: Communication and interaction between intelligent systems will happen in a specific situation. Inclusion of contextual information (e.g., a model of environment) provides common ground with implicit conventions.

In the current paper, we review IBCI applications that incorporated the principles for interaction. In particular, the interaction of IBCI-system is designed with the following components, *1) shared knowledge between the robot and user*. The ongoing work on recognition of human anticipatory behavior described in section 4 is based on this principle. For example, consider a scenario of an intelligent robotic wheelchair facing a dining table in a hall of several tables. From the robot-controller's point of view, the table is an obstacle and it can't decide by it self whether to dock to it or to avoid it. But it is the user who decides to dock to it if he wants to take brakefast. The user *anticipates* for the docking event to happen if he wishes to dock. The shared knowledge allows the robot to make corresponding actions (e.g., docking, or avoiding the obstacle) upon the recognition of anticipation related brain activity of the user. The shared knowledge, i.e., robot's detection of a table and user's anticipation of events allows to achieve the desired goal. *2) communication error recovery through feedback and the detection of error related brain activity.* We have implemented these two mechanisms in our applications that allow the user to correct his commands from the feedback of recognized commands by classifiers (described in 3) as well as the robot to change its commands up on the recognition of error related brain activity (described in 4) and *3) context filtering of illogical mental commands inferred by the interface.* For a brain actuated robot application (described in section 3.1), the filtering is achieved by using a finate state machine that translated the mental commands into device commands according to the environmental context. In the case of a brain actuated wheelchair application (described in section 3.2), the filtering is achieved by combining the probabilities inferred by the classifier for mental commands with that of context-based-filter of the robotic wheel chair.

In the next section we review the state of art of BCI research along with the methods that lead to the success of IBCI. In section 3, we review IBCI applications that implement the key principles introduced above. In section 4, we show the possibility of designing empathic devices with the recognition of user's cognitive states directly from brain signals. Finally in section 5 we discuss conclusions and future directions of research.

2 BCI research and IBCI system

A schematic of a BCI system is shown in Figure 1. Brain electrical activity is acquired using electrodes (either implanted inside the brain or externally on the scalp). From the recorded signals, features (e.g., amplitudes of evoked potentials, or sensory motor cortex rhythms) that reflect user's intent, are extracted using signal processing methods. These features are then translated into device commands (e.g., using neural networks) which are then issued to systems like, virtual-keyboards, mobile robots, robotic wheelchairs and computer games. *Feedback* from these systems is given to the user using various modalities (e.g., visual, auditory etc.).

BCI is broadly classified into three categories based on invasiveness of the recording technique as 1) invasive, 2) partially invasive and 3) non-invasive BCI [22]. For an invasive BCI, the electrodes are implanted directly into the grey matter of the brain during neurosurgery. As they rest in the grey matter, it can produce the highest quality signals of BCI devices but are prone to scar-tissue build-up, causing the signal to become weaker or even lost as the body reacts to a foreign object in the brain [19]. Partially invasive BCI [10] uses electrodes implanted inside the skull but resting outside the brain rather than amidst the grey matter (e.g., Electrocorticography, ECoG). They produce better resolution than non-invasive electrodes and have lower risk of forming scar-tissue in the brain than fully invasive electrodes. Finally, Electroencephalograph (EEG), Magnetoencephalography (MEG) and functional magnetic resonance imaging (fMRI) have both been used successfully in non-invasive BCI. Among all the EEG is the most used signal acquisition method mainly due to its fine temporal resolution, ease of use, portability and low set-up cost. Since the current paper is based on

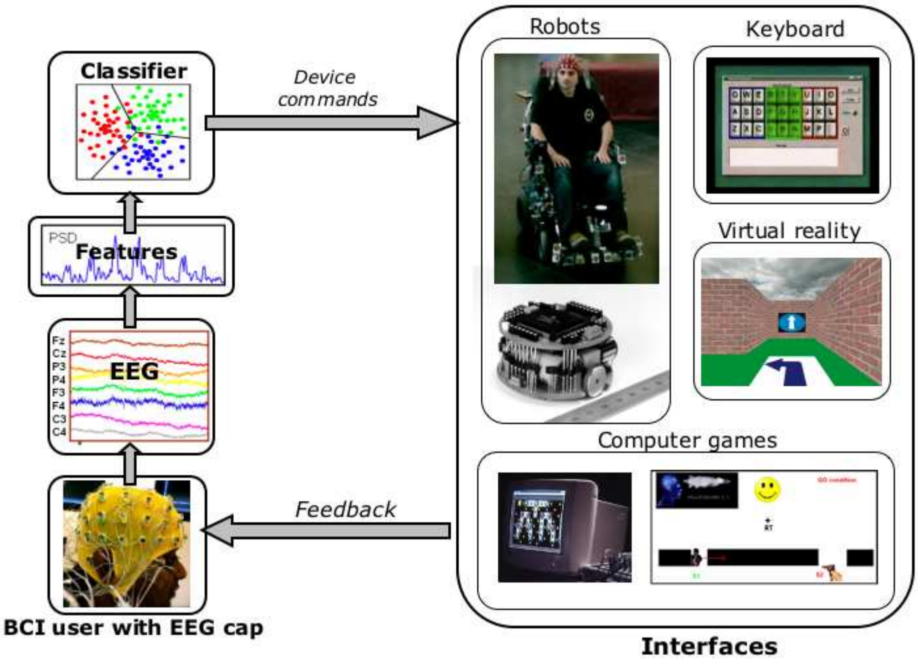


Fig. 1. Operation of a brain computer interfacing (BCI) system.

non-invasive IBCI system, we review feature extraction and classification stages of EEG based BCI.

Based on the operation, non-invasive BCI systems can be classified into two types, 1) system driven, 2) user driven. The system driven BCI uses EEG waveforms that are generated automatically in response to external stimulus (e.g., visual, auditory stimulus from the interfacing machine), called evoked potentials (EP). One example is the P300 signal, which is a potential evoked by an awaited infrequent event that appears at centro-parietal locations along the mid line of the scalp. It is a positive wave peaking around 300 ms after task-relevant stimuli [36]. Traditionally, P300 has been used in BCI research to develop virtual keyboards [2] [9] with a typing speed of five letters per minute, but recently this same potential has also been the basis for brain-actuated control of a virtual reality system [4] and of a wheelchair [36]. Steady-state visual evoked potentials (SSVEP) are another example of evoked potentials that are induced by a visual stimulus repeated at a rate higher than 6 Hz [40]. Most SSVEP-based BCI system depend on muscular control of gaze direction for their operation, whereas all other kinds of BCI systems do not depend on the brain's normal output pathways of peripheral nerves and muscles [15] [24] [40]. The main drawback of system driven BCI is that, since the subject's response is locked to the stimulus, he cannot generate mental commands at any time he wants.

On contrast the user-driven BCI is based on self modulation of EEG rhythms by the user, i.e. spontaneous brain activity. for instance, self modulation by imagination of movements can result in changes in EEG rhythm in central region of the scalp overlying the sensorimotor cortex [3] [8] [33] [45]. These rhythms are the basis of several BCI systems [3] [8] in which imagination of hand movement gives rise to an amplitude suppression in the α-band (8-12 Hz) and β-band (13-28 Hz) [4] rhythms over the contralateral primary hand motor cortical area [33]. Wolpaw and co-workers [43] [45] used continuous changes in the amplitudes of these rhythms to move a cursor in a computer screen. Alternatively, some researchers measure slow cortical potentials (SCP) whose negative amplitudes are related to the overall preparatory excitation level of a given cortical network, the more negative the more active over the top of the scalp at electrode-Cz [5] [18]. Attentional modulation seems to constitute the cognitive strategy in the physiological regulation of SCP. The team lead by Birbaumer has widely shown that healthy subjects as well as severely paralyzed patients can learn to self-control their SCPs through operant conditioning to move an object on a computer screen in a BCI referred to as *Thought Translation Device* (TTD) [17].

EEG-based BCIs are limited by a low channel capacity. Most of the current systems have a channel capacity below 0.5 bits/s [43]. One of the main reasons for such a low bandwidth is that they are based on *synchronous protocols* , where EEG is time-locked to externally paced cues repeated every 4-10 s and the response of the BCI is the average decision over this period (system driven BCI) [5] [31] [34] [37] [43]. The system dribven BCI is not natural for the user since his response is always time-locked to externally placed cues generated by the system. The user can't not decide by him self whenever he want to make a decision. On the contrary, the IBCI group utilizes more flexible *asynchronous protocols* where the subject makes self-paced decisions (user-driven) on when to stop performing a mental task and start immediately the next one [27] [30] [29]. In such asynchronous protocols, the subject can voluntarily change the mental task (e.g., imagination hand movement. See figure 2(b) for scalp topographies of EEG activity during these mental tasks in α band) being executed at any moment without waiting for external cues (this approach is grounded in a number of neurocognitive studies that have found that different mental tasks such as mental rotation of geometric figures [46], arithmetic operations [7], or language [32] activate local cortical areas to a different extent). The time of response of an asynchronous IBCI can be below 1 second (responds every 0.5 second) [30]. The rapid responses of asynchronous BCIs, together with their performance, give a theoretical channel capacity between 1 and 1.5 bits/s.

Coming to the feature extraction, IBCI team analyzes continuous variations of EEG rhythms over several frequency bands. The user specific EEG-features

[4] EEG activity is typically described in terms of rhythmic activity. The rhythmic activity is divided into several frequency bands (e.g., α band from 8 to 12 Hz. Suppression in this band power is usually observed in sensory motor areas while the user performing mental imagination.)

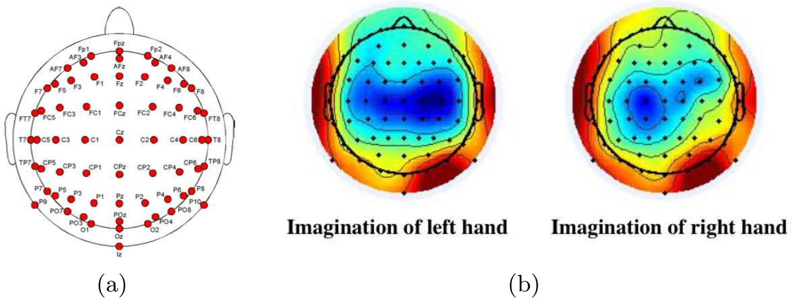


Fig. 2. (a) Top view of electrode positions according to the international 10-20 system. (c) Event related de-synchronization: decreased α-band (12Hz) power contralaterally to the imagination of hand. Dark regions correspond to lower power.

extracted using Canonical Variate Analysis (CVA) for multi-class problems [13]. This technique maximizes the separability between the patterns generated by executing the different mental tasks. For the classification of these features, IBCI team typically uses Gaussian classifiers which are a modified Gaussian Mixture Model (GMM) [28] [29]. The output of the classifier is posterior probability that can be used to label specific classes and an 'unknown' class. It is worth noting that the use of statistical rejection criteria helps to deal with an important aspect of a BCI, namely idle states where the user is not involved in any particular mental task. These classifiers has been shown to perform better than support vector machines (SVMs), temporal processing neural networks (NNs) [16], committees of multi later perceptrons (MLPs), learning vector quantization and incremental radial basis networks [28] (for more details on the Gaussian classifier, refer [28], for a review of classification engines for BCI in general, refer Lotte et al [23]).

To sum up, the key principles behind the feature extraction and classification parts of IBCI system are, 1) using task induced EEG rhythms over cortical areas as features of mental commands, 2) canonical feature extraction for multi-class problems, 3) statistical classifier with rejection criteria and 4) asynchronous and spontaneous operation. Current research of IBCI group also focuses on adding 'cognitive states recognition' to 'mental command recognition' to improve its performance. The recognition of cognitive states can be used for implementing the principles of intelligent interaction like, shared knowledge and error recovery as described in section 1. The details of recognition of the cognitive states are described in section 4. In the next section, we discuss the design of brain actuated interaction to drive a mobile robot and a robotic wheelchair in natural environments.

3 IBCI applications

In this section we present the current applications of IBCI system for controlling a mobile robot and a robotic wheelchair through intelligent interaction in the light of AmI with the following key components, 1) shared control, 2) error control through detection of error related potentials, 3) inclusion of contextual information.

3.1 Brain-actuated robots

Recently, IBCI group has shown for the first time that asynchronous analysis of EEG signals is sufficient for humans to continuously control a mobile robot (i.e. Khepera) along non-trivial trajectories requiring fast and frequent switches between mental tasks [30]. Human users learned to mentally drive the robot between rooms in a house-like environment (see Figure 3(a)). Furthermore, mental control was only marginally worse than manual control on the same task. A key element of this brain-actuated robot is shared control between two intelligent agents (i.e., the human user and the robot). The user only gives high-level mental commands (e.g., *turn left*, *turn right*, *forward*) that the robot performs autonomously. Another critical feature is that a BCI asynchronous operation, allowing the user to issue mental commands at any moment.

In order to endow the system with flexible, robust control, there is no one-to-one mapping of user's mental commands and robot's actions. Instead, we combine environmental information gathered from robot's on-board sensors with mental commands to take appropriate actions according to the context (i.e. shared control). This combination is implemented by a Finite State Automation (FSA) [30]. The transitions between different behaviors are determined by the 3 mental commands, 6 perceptual states of the environment (based on the robots sensory readings: *left wall*, *right wall*, *wall or obstacle in front*, *left obstacle*, *right obstacle*, and *free space*) and a few internal memory variables. These perceptual states are determined by using a neural network classifier that takes input from the sensory readings [30]. The memory variables keep contextual information required to implement correctly the different behaviors. Figure 3(b) shows a fragment of the FSA (for full description, see [30]). As shown in the figure, if the robot is performing the behavior forward and perceives a wall to the left, it switches automatically to the behavior follow left wall. The transition to the behavior-forward is necessary, for example, in the case where the robot is approaching an open door and the user wants the robot not to enter into the room.

A final element is the use of an appropriate feedback indicating the current mental state recognized by the embedded classifier. This is done by means of three lights (red, blue, green) on top of the robot that corresponds to the three mental commands (turn right, turn left, move forward). Thus, if the robot is following the left wall and is approaching an open door, a blue feedback indicates that the robot will turn left to continue following the left wall (and, so, it will enter into the room). On the contrary, a green feedback indicates that the robot will move forward along the corridor when facing the doorway and will not enter into the room. This simple feedback allows users to correct the robot trajectory in case of errors in the recognition of the mental states or errors in the execution of the desired behavior (due to the limitations of the robot

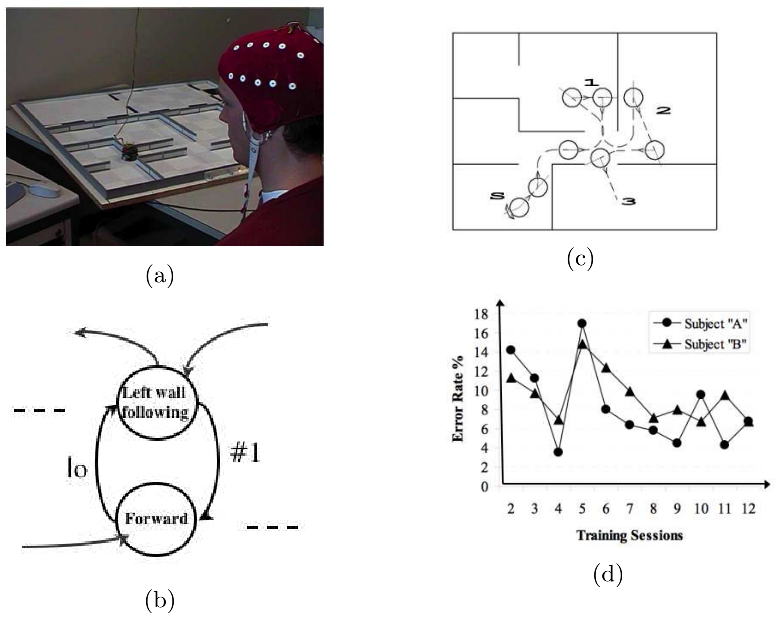


Fig. 3. Brain actuated robot control: (a) a user driving brain actuated robot (b) a fragment of Finate State Automaton (FSA) used for the shared control. Transitions between the 6 behaviors were determined by 3 mental states (#1: turn left, #2: turn right, #3: go forward), 6 perceptual states (lo: leftwall, ol:right wall, ö: wall or obstacle in front), and some memory variables.) (c) Experimental platform and a typical trajectory followed by the robot under the mental control. The robot started in bottom left room and then visited three other rooms, (d) error rate (percentage of false positives)(adopted from [30]).

Table 1. Time in seconds for three different trials in controlling first mentally and then manually by two users.

Trial	User-A		User-B	
	Mental-control	manual-control	Mental-control	Manual-control
1	149	124	219	156
2	183	135	189	155
3	191	129	175	117
Average	174	129	194	143

sensors). The figure 3(c) shows a typical trajectory of brain actuated robot. After 5 and 3 days of initial training with the interface, the users achieved a satisfactory level of performance (correct recognition was above 65% while the errors were below 7% - the remaining were 'unknown' response) (see Figure 3(d)). Table 3.1 gives the time in seconds necessary to generate the desired trajectories in three different trials for the two participants comparing mental control and

Table 2. Comparison of bit-rate of online classification with static initial classifier.

Session #	Static initial classifier	Online classification
1	0.29	1.44
2	0.20	1.41
3	0.14	1.34
4	0.18	1.34
Average	0.20 ± .06	1.38±0.05

mannual control. Remarkably, trial duration for mental control was comparable with mannual control. On average, brain-actuated control of the robot takes only 35% longer than manual control for both the participants. The figure 3(d) shows the performance curve two users. First, a clear improvement can be observed during the first day (sessions 2 to 4), with an excellent performance. Second, the performance degrades at the beginning of second day (session 5) but recovers at the end. This shows difficulty of generalizing from one day to the next due to natural variability of brain signals. This variability can be compensated by incorporating online-learning as discussed in following paragraphs.

The variability of EEG signal within a session and from session to session is due to several factors including the background activity, fatigue and concentration levels, and intentional change of subject's strategy. This means that the classifier designed with past data might not perform well for the present or future data. To deal with this problem, IBCI applies adaptive algorithms that are constantly tuned to the subject. These techniques improve the performance of the BCI system allowing the subject to learn to use BCI more effectively. We first build classifier with the past data and then, as new EEG is obtained during the use of the BCI, we use it for updating the classifier (for more details, refer [1] [6]).

The studies on online learning are performed on offline data collected during brain-actuated robot control. The improvements in terms of the bit-rate comparing static initial and adaptive classifier are shown in table 3.1 (bit rate is channel capacity as explained in [29]). The online classification rates are much higher than the static classifiers. Moreover, the classifiers obtained at the end of each session (i.e. that were modified online throughout the experiment) outperforms the initial classifier.

From the above brain-actuated robot application, and consistent with the AmI principles, we conclude that, 1) fusing of knowledge of the human user and intelligent robot allows for effective human-computer interaction; 2) apart from using shared knowledge and contextual information, the error recovery (achieved by using feedback in the present case) is also important for successful control; 3) online adaption of the intelligent system will improve the interaction performance.

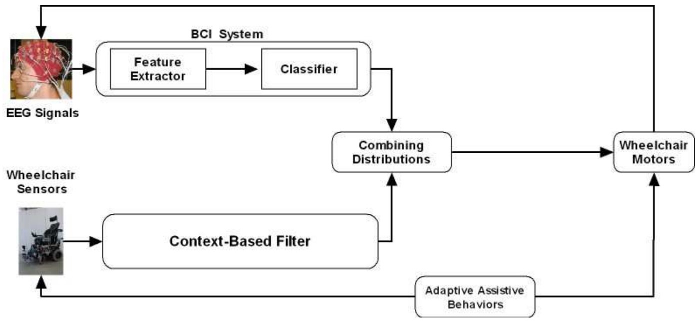


Fig. 4. A schematic of shared control system

3.2 Brain actuated wheelchair

The recent studies of the IBCI in collaboration with KV Leuven under the framework of the European project MAIA (http://www.maia-project.org) aim at the development of a brain-actuated wheelchair that can provide mobility directly through mental control. An autonomous controller embedded in the wheelchair could serve help paralyzed patient for navigation. However, the subject might lose the feeling of continuous control with such a controller. The loss of independence is undesirable and therefore, shared control between the user and the controller is more suitable [35]. IBCI's design for such a system has three basic elements [35] [41], 1) adaptive shared controller that fuses the human and wheelchair decisions in a Bayesian way for better steering commands , 2) context information from the model of environment for filtering out unlikely decisions taken by the classifier and 3) assistive behaviors (collision avoidance (A0) obstacle avoidance(A1), and orientation recovery (A2)) based on the model of environment (e.g., openings in a corridor). See Figure 4 for the architecture of shared control of brain-actuated wheelchair.

Similar to the brain-actuated robot control, user can steer the wheelchair by issuing three discrete mental commands. The induced EEG rhythms (power spectrum density computed over one second in a selected subset of electrodes and frequency bands [14]) due to these mental commands are classified by a statistical Gaussian classifier whose outputs are posterior probabilities for the device commands (*move forward*, *turn left* and *turn right*). The asynchronous BCI system responds every 0.5 seconds by sending these probability distribution of the three mental commands to the shared controller which are then translated into steering commands (i.e., translational (ν) and rotational (ω) velocity). Instead of directly executing the user's steering commands, the shared control system takes environmental situation into account which is registered through a laser scanner. With this knowledge, the controller triggers one of the assistive behaviors using a *winner-takes-all* method (e.g., if the user steers too close to an obstacle, an avoidance behavior of the shared control is activated to prevent collision). Studies with the adaptive shared control are illustrated in figure 5(b). Without A2, the subject makes unnecessary loops while navigating (refer to [35]). The

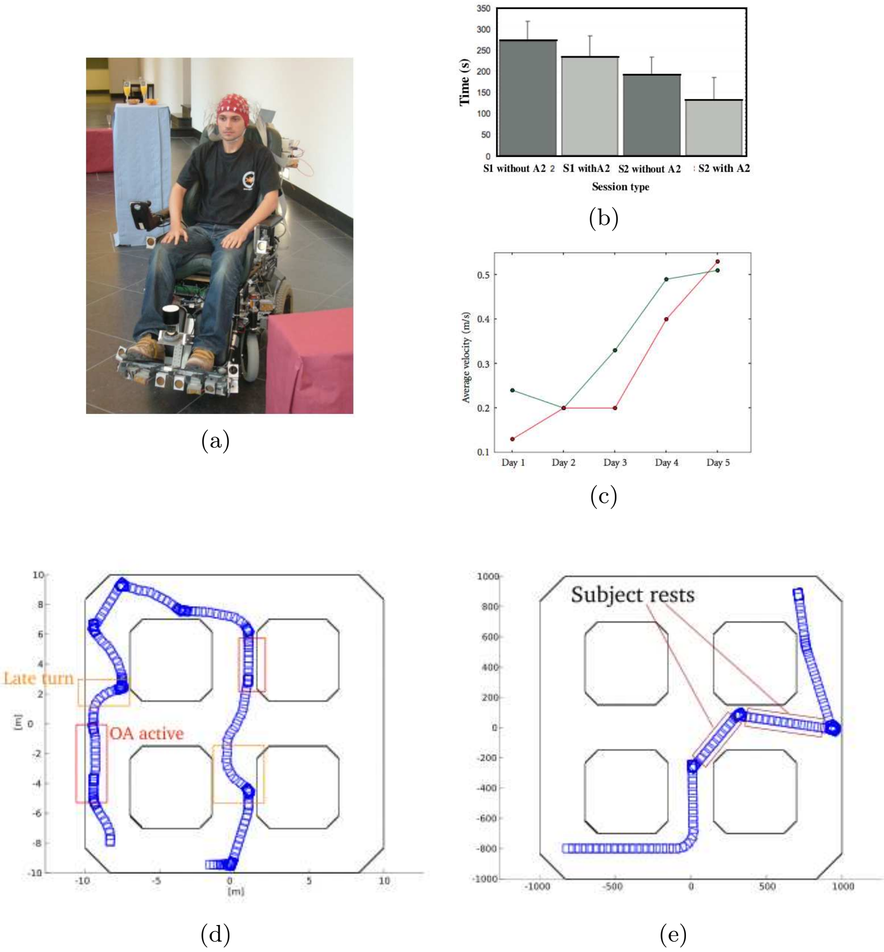


Fig. 5. (a) The subject wearing an electroencephalogram (EEG) sensor cap maneuvering the robotic wheelchair through a natural indoor environment. The visible are the sensors of the robotic platform: a laser range scanner in front and sonar sensors are visible at the lower part of the wheelchair, (b) Average navigation time with and without orientation recovery (A2) for sessions 1 and 2. (c) Average of velocity during five training days. The lower line represents the performance without context filter and the upper one represents the average velocity when the context filter is active. (d) Wheelchair trajectory in a different environment without context filtering. (e) A trajectory of a session with context filtering (the figure reproduced from [41]).

average elapsed time and average distance travelled (refer to [35] and [41]) also reduced significantly in navigating towards the goal.

Since the mental command recognition is not perfect we need to correct them using environmental context. This is archived by adding a context filter to adaptive shared controller. The controller estimates the environmental context for detecting illogical steering signals. The context filtering in three steps. First, the context estimation is performed by defining a general, a-priory known model of the user intention (e.g., smooth and efficient forward navigation through the environment) on one hand and a constant automatic estimation of the environmental situation on the other hand. The situations were modelled as the number and location of *openings* (i.e. wide, open spaces through which user might safely navigate). Second, each opening represents a general direction in which the user might opt to continue his travel. With the knowledge of current situation, a probability distribution concerning the possible user actions were built. Third, the intelligent controller combines the probability distributions of the statistical classifier of BCI system and probability distribution generated from the environmental knowledge, so as to get a better estimation of the user's local steering intent.

The performance of the context filtering is also tested in terms of velocity of maneuvering (see figure 5(c)). Inspire of the fact that the user's driving skills improve gradually, both with and without context filtering, activation of this feature allow the user to steer faster in early stages of training. Figure 5(d) shows a trajectory performed without context filtering. We can see that there are many-nearby collisions (collisions won't happen due to A1), resulting in a rather jaggered path. Enabling context filter results in smoother trajectories, although near-collisions still occur (see figure 5(e)). More results on context filtering are described in [41].

4 Cognitive state recognition: Towards empathic devices

Cognition is a mental process of knowing, including aspects such as awareness, perception, reasoning, and judgment. An *empathic agent* is a cognizant that comprehends needs, feelings, problems and views of humans and responds to them. Recently, IBCI group started investigating on the use of brain signals related to cognitive process for boosting the IBCI performance. By implementing recognition of cognitive states, the IBCI becomes a basic empathic agent. In particular the group is investigating on using "user's awareness of machine's error" and "human anticipatory states".

4.1 Decoding human awareness of machine error

BCIs are prone to errors in the recognition of user's intent from his mental commands. As in the human-human interaction, an elegant way of improving the BCI performance is to use a verification procedure directly based on presence of error related potentials (ErrP) in the brain activity. ErrP is a potential elicited after presenting the feedback of an error as response which is clearly detected in FCz and Cz electrode (see figures 6(a) and 6(b)). Several studies show the presence of this potential in typical choice reaction tasks when the subject makes mistake by himself. At IDIAP, in the context of BCI, we have shown that ErrPs are elicited even when error is made by the interface during the recognition of subject's intension. Ferrez and Millán termed this type of ErrP as *interaction-ErrP* [11] [12], as it is elicited by the presentation of feedback indicating incorrect response of simulated BCI.

Furthermore, we are interested in how ErrP can be used to improve the performance of a BCI. As shown in the Figure 6(c), after translating the subject's intention form his mental command, into control command for the robot, the BCI provides a feedback of it, which will be executed only if no ErrP follows the feedback (see figure 6(c)). In this new interaction method, the challenge is to recognize the ErrP on single trials. After characterization of these potentials, we have developed classification technique that archive successful recognition of these potentials (up to 80% correct classification of ErrP and up to 83% of correct trials [11]). In addition, this type of interaction improves the bit-rate of the BCI system by 28% for three-class problem and by 72% for two-class problem(see [11] for more detailed results).

4.2 Decoding human anticipation

Animals have the ability to anticipate to upcoming events given a predictive model. In particular, in humans, the EEG correlates of anticipation are well known, and one of such signal is Contingent Negative Variation (CNV). CNV is an increasing negative shift of the cortical electrical potential associated with anticipated response to an external stimulus. It is therefore interpreted as both an expectation related potential and anticipation related potential [38] [42]. Recognition of CNV can be used for implementing *shared knowledge* of the human user and a semi-autonomous system in making final decisions. For example, a robotic wheelchair facing a dining-table can not decide by itself whether to dock to it or to avoid it (i.e., obstacle avoidance behavior). But, the presence or absence of anticipation related potentials in the subject's EEG will allow the wheelchair to make a final decision. The question that we are addressing in this section is that, "Is it possible for machines to predict human anticipation to particular events?".

We study changes in the CNV depending on the task-dependent relevance of external stimulus (S1) in a classical Go/NoGo CNV paradigm. In "Go" condition, the subject is instructed to anticipate to imperative stimulus (S2) and press a key on its arrival and in "NoGo" condition, the subject instructed to do nothing. On-line recognition of such changes provides information that can be used by the semi-autonomous system in situations when it is not able to make reliable decisions. Grand averages of potentials recorded in CNV Go/NoGo paradigm is shown in figure 6(d). These potentials are classified using a simple Linear Discriminant Analysis (LDA) classifier. The results show that the anticipatory potentials can be classified up to an accuracy of 70% at least 0.5 sec before the subject presses a key.

The recognition of ErrP and anticipatory signals from the EEG introduces empathic capabilities in the BCI system. Thus, we show the feasibility of designing empathic devices that can predict human actions, judgements and needs directly brain signals. Further, the implementation of other emotive states recognition, such as attention, alarms, frustration and confusion will improve empathic capabilities of our BCI system.

5 Conclusions and future work

In this paper we have introduced several applications of the IDIAP BCI (IBCI). In particular this paper shows how these applications successfully integrate design principles from human-centered approaches for intelligent interaction in the domain of Brain Computer interfaces. Namely, we have shown how the described IBCI systems are endowed with 1) Shared knowledge, 2) communication error recovery and 3) contextual information. These principles are consistant with from AmI design criteria and allows for the robust performance of the IBCI systems by showing strong evidence of the potential synergy between AmI and BCI research.

Sharing the knowledge between the human user and robot perception of environment is achieved by using FSA for the brain-actuated control of a robot. In the case of brain-actuated wheelchair, it is achieved by combining the probability distributions inferred by the Gaussian classifier from the user's mental commands with those inferred from the environment by the wheelchair sensors. We have shown two possible methods for communication recovery, 1) by giving a feedback of the recognized mental commands to the user so that he can change his mental commands in case of error, and 2) with the use of interaction-ErrP. We have also shown that context filtering of illogical mental commands inferred by the interface improves the driving performance of the wheelchair. The recent work of IBCI team shows a way to improve empathic capabilities of a machine by using human cognitive state recognition (e.g., recognition of ErrPs and anticipation related potentials). These capabilities can be improved by recognizing other cognitive and emotive states such as attentional level, frustration, alarm, and confusion.

In summary, IBCI research shows the feasibility of developing systems that have an enhanced comprehension of human's cognitive and emotive states establishing boosted intelligent human-machine interactions. The synergy between AmI and BCI research will permit to develop empathic systems and environments, providing tools for making human-machine interaction more resemblance to human-human interactions.

6 Acknowledgements

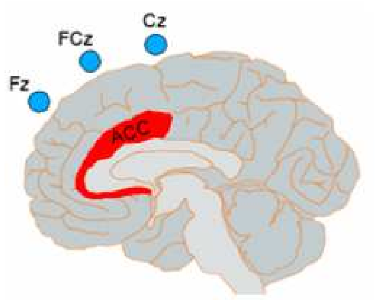
This work was supported in part by the Swiss National Science Foundation through the National Centre of Research on "Interactive Multimodal Information Management (IM2)", European IST Programme FET Project FP6-003758, and FP6-IST-027140. This paper only reflects the author's views and founding agencies are not liable for any use that may be made of the information contained herein. We thank 'Division PMA team' of Katholieke Universiteit, Leuven, for it's collaboration on brain-actuated-wheelchair application.

References

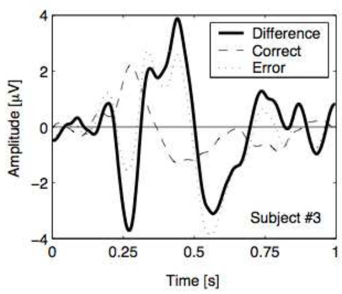
1. Buttfield A., Ferrez P.W., and Millán J. del R. Towards a robust BCI: Error potentials and online learning. *IEEE Trans. on Neural Systems and Rehabilitation Engineering*, 14(2):164–168, 2006.
2. B. Z. Allison and J. A Pineda. ERPs evoked by different matrix sizes: implications for a brain computer interface (bci) system. *IEEE Trans Neural Syst Rehabil Eng*, 11(2):110–113, Jun 2003.
3. F. Babiloni, F. Cincotti, L. Lazzarini, J. del. R Millán, J. Mouriño, M. Varsta, J. Heikkonen, L. Bianchi, and M. G. Marciani. Linear classification of low-resolution EEG patterns produced by imagined hand movements. *IEEE Trans Rehabil Eng*, 8(2):186–188, Jun 2000.
4. J. D. Bayliss. Use of the evoked potential P300 component for control in a virtual apartment. *IEEE Trans Neural Syst Rehabil Eng*, 11(2):113–6, June 2003.
5. N. Birbaumer, N. Ghanayim, T. Hinterberger, I. Iversen, B. Kotchoubey, A. Kübler, J. Perelmouter, E. Taub, and H. Flor. A spelling device for the paralysed. *Nature*, 398(6725):297–8, Mar 25 1999.
6. Anna Buttfield and José del R. Millán. Online classifier adaptation in brain-computer interfaces. IDIAP-RR 16, IDIAP, 2006.
7. F. Chochon, L. Cohen, P. F. van de Moortele, and S. Dehaene. Differential contributions of the left and right inferior parietal lobules to number processing. *J Cogn Neurosci*, 11(6):617–630, Nov 1999.
8. G. Dornhege, B. Blankertz, G. Curio, and K. R. Müller. Boosting bit rates in noninvasive eeg single-trial classifications by feature combination and multiclass paradigms. *IEEE Trans Biomed Eng*, 51(6):993–1002, Jun 2004.
9. L. A. Farwell and E. Donchin. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalogr Clin Neurophysiol*, 70(6):510–523, Dec 1988.
10. E. A. Felton, J. A. Wilson, J. C. Williams, and P. C. Garell. Electrocorticographically controlled brain-computer interfaces using motor and sensory imagery in patients with temporary subdural electrode implants. report of four cases. *J Neurosurg*, 106(3):495–500, Mar 2007.
11. P.W. Ferrez and J. del R. Millán. You are wrong!—automatic detection of interaction errors from brain waves. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence*, Edinburgh, UK, August 2005.
12. P.W. Ferrez and J. del R. Millán. Error-related EEG potentials in brain-computer interfaces. In G. Dornhege, J. d. R. Millán, T. Hinterberger, D. McFarland, and K.-R. Müller, editors, *Towards Brain-Computer Interfacing*. The MIT Press, 2007.
13. F. Galán, P. W. Ferrez, F. Oliva, J. Guàrdia, and J. del R. Millán. Feature extraction for multiclass BCI using canonical variates analysis. In *IEEE International Symposium on Intelligent Singnal Processing, Alcalá de Henares*, 2007.
14. F. Galán, M. Nuttin, E. Lew, P. W. Ferrez, G. Vanacker, J. Philips, H. V. Brussel, and J. del. R. Millán. An asynchronous and non-invasive brain-actuated wheelchair. In *In 13th International Symposium on Robotics Research, Hiroshima, Japan.*, 2007.

15. Xiaorong Gao, Dingfeng Xu, Ming Cheng, and Shangkai Gao. A BCI-based environmental controller for the motion-disabled. *IEEE Trans Neural Syst Rehabil Eng*, 11(2):137–140, Jun 2003.
16. A. Hauser, P. E. Sottas, and J. d. R. Millán. Temporal processing of brain activity foe the recognition of EEG patterns. In *Proceedings 12th Int. Conf. on Artificial Neural Nettworks*, pages 1125–1130, 2002.
17. T. Hinterberger, N. Neumann, M. Pham, A. Kübler, A. Grether, N. Hofmayer, B. Wilhelm, H. Flor, and N. Birbaumer. A multimodal brain-based feedback and communication system. *Exp Brain Res*, 154(4):521–526, Feb 2004.
18. T. Hinterberger, S. Schmidt, N. Neumann, J. Mellinger, B. Blankertz, G. Curio, and N. Birbaumer. Brain-computer communication and slow cortical potentials. *IEEE Trans Biomed Eng*, 51(6):1011–1018, Jun 2004.
19. P. R. Kennedy and R. A. Bakay. Restoration of neural output from a paralysed patient by a direct brain connection. *Neuroreport*, 9:1707–1711, 1998.
20. R. Krepki, B. Blankert, G. Curio, and K. R. Müller. The Berlin brain-computer interface (BBCI) towards a new communication channel for online control in gaming applications. *Multimedia Tools and Applications*, pages 73–90, February 2007.
21. A. Kübler and K. R. Müller. *An introduction to brain computer interfacing in Towards Brain-Computer Interfacing.* MIT Press, 2007.
22. M. A. Lebedev and M. A. Nicolelis. Brain-machine interfaces: past, present and future. *Trends Neurosci*, 29:536–546, 2006.
23. F. Lotte, M. Congedo, A. Lcuyer, F. Lamarche, and B. Arnaldi. A review of classification algorithms for eeg-based brain-computer interfaces. *J Neural Eng*, 4(2):R1–R13, Jun 2007.
24. M. Middendorf, G. McMillan, G. Calhoun, and K. S. Jones. Brain-computer interfaces based on the steady-state visual-evoked response. *IEEE Trans Rehabil Eng*, 8(2):211–214, Jun 2000.
25. J. del R. Millán. Brain-computer interfaces. In Michael A. Arbib, editor, *The Handbook of Brain Theory and Neural Networks: The Second Edition*. The MIT Press, 2002.
26. J. del R. Millán, P. W. Ferrez, and A. Buttfield. Non-invasive brain-machine interfaces. *ESA report on studies on feasibility of non-invasive brain-machine interfaces (www.esa.int/gsp/ACT/bio/BMI.htm)*, 2006.
27. J. del. R. Millán and J. Mouriño. Asynchronous BCI and local neural classifiers: An overview of the Adaptive Brain Interface project. *IEEE Trans. on Neural Systems and Rehabilitation Engineering, Special Issue on Brain-Computer Interface Technology*, 11(2):159–161, 2003.
28. J. del. R. Millán, J. Mouriño, M. Franzé, F. Cincotti, M. Varsta, J. Heikkonen, and F. Babiloni. A local neural classifier for the recognition of EEG patterns associated to mental tasks. *IEEE Trans Neural Networks*, 13(3):678–686, 2002.
29. J. del. R. Millán, F. Renkens, J. Mouriño, and W. Gerstner. Non-invasive brain-actuated control of a mobile robot by human EEG. *IEEE Trans Biomed Eng*, 51(6):1026–1033, Jun 2004.
30. J. del. R. Millán, F. Renkens, J. Mourino, and G. Wulfram. Brain-actuated interaction. *Artif. Intell.*, 159(1-2):241–259, 2004.
31. B. Obermaier, G. R. Müller, and G. Pfurtscheller. “Virtual keyboard” controlled by spontaneous EEG activity. *IEEE Trans Neural Syst Rehabil Eng*, 11(4):422–426, Dec 2003.
32. S. E. Petersen, P. T. Fox, M. I. Posner, M. Mintun, and M. E. Raichle. Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature*, 331(6157):585–589, Feb 1988.

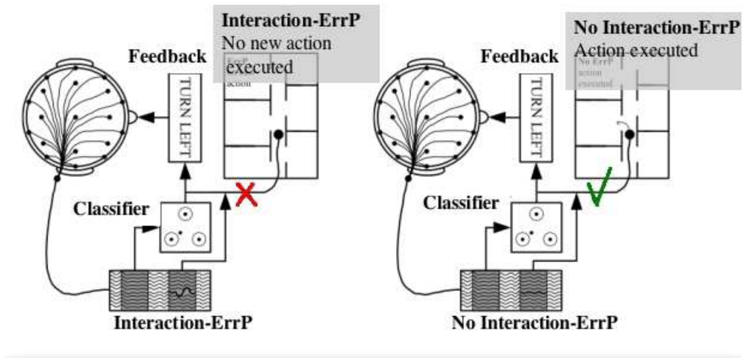
33. G. Pfurtscheller and C. Neuper. Motor imagery activates primary sensorimotor area in humans. *Neurosci Lett*, 239(2-3):65–68, Dec 1997.
34. G. Pfurtscheller and C. Neuper. Motor imagery and direct brain-computer communication. In *IEEE*, volume 89, pages 1123–1134, 2001.
35. J. Philips, J. del. R. Millán, G. Vanacker, E. Lew, and F. Galán. Adaptive shared control of a brain-actuated simulated wheelchair. In *IEEE 10th International Conference on Rehabilitation Robotics*, 2007.
36. B. Rebsamen, E. Burdet, Teo C.L., Q. Zeng, C. Guan, M. Ang, and C. Laugier. A brain control wheelchair with a P300 based bci and a path following controller. In *Proc. 1st IEEE/RAS-EMBS Int. Conf. Biomedical Robotics and Biomechatronics.*, 2006.
37. S. J. Roberts and W. D. Penny. Real-time brain-computer interfacing: a preliminary study using bayesian learning. *Med Biol Eng Comput*, 38(1):56–61, Jan 2000.
38. R. Rosen. *Anticipatory systems.* Pergamon press, 1985.
39. Albrecht Schmidt. *Interactive Context-Aware Systems Interacting with Ambient Intelligence*. IOS Press, 2006.
40. E. E. Sutter. The brain response interface: communication through visually-induced electrical brain responses. *J. Microcomput. Appl.*, 15:31–45, 1992.
41. G. Vanacker, J. del. R. Millán, E. Lew, P. Ferrez, F. Galán, J. Philips, H. V. Brussel, and M. Nuttin. Context-based filtering for assisted brain-actuated wheelchair driving. *Computational Intelligence and Neuroscience*, 2007, 2007.
42. W. G. Walter, R. Cooper, V. J. Aldridge, W. C. McCallum, and A. L. Winter. Contingent negative variation: An electric sign of sensorimotor association and expectancy in the human brain. *Nature*, 203:380–4, Jul 25 1964.
43. J. R. Wolpaw. Brain-computer interfaces (BCIs) for communication and control: a mini-review. *Suppl Clin Neurophysiol*, 57:607–613, 2004.
44. J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan. Brain-computer interfaces for communication and control. *Clin Neurophysiol*, 113(6):767–791, Jun 2002.
45. J. R. Wolpaw, D. L. McFarland, and T. M. Vaughan. Brain-Computer Interface Research at the Wadsworth Center. *IEEE Trans. Rehab. Eng*, 8:222–226, 2000.
46. A. Yoshino, M. Inoue, and A. Suzuki. A topographic electrophysiologic study of mental rotation. *Brain Res Cogn Brain Res*, 9(2):121–124, Mar 2000.



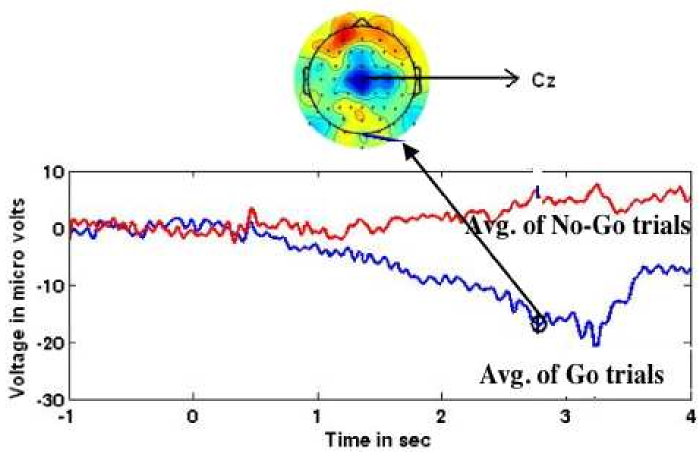
(a)



(b)



(c)



(d)

Fig. 6. (a) Position of Fz, FCz and Cz electrodes with respect to Anterior Cingulate Cortex (ACC) which is the origin of ErrP (b) Average of the difference (error-minus correct) between the potentials of error and correct trials at Cz electrode. (c) BCI based on 'Interaction ErrP'. The BCI user receives visual feedback of indicating the output of the classification engine before the actual execution of the associated command(e.g., "turn-left"). If the feedback elicits an ErrP, this command is simply ignored and the robot will stay executing the previous command (right). Otherwise, the command is sent to the robot (left). (d) Contingent Negative Variation with a classical Go/NoGo task: external stimulus (S1) appears at time "0" secs, and imperative stimulus (S2) appears at "3.25" secs. The upper line is average of NoGo trials and lower line is average of Go trials. On the top, the scalp topography of CNV is shown at time 2.75 secs.

Design and Validation of HABTA: Human Attention-Based Task Allocator

Peter-Paul van Maanen[1,2], Lisette de Koning[1], and Kees van Dongen[1]

[1] TNO Human Factors, P.O. Box 23, 3769 ZG Soesterberg, The Netherlands
{peter-paul.vanmaanen, lisette.dekoning, kees.vandongen}@tno.nl
[2] Vrije Universiteit Amsterdam, De Boelelaan 1081a,
1081 HV Amsterdam, The Netherlands

Abstract. This paper addresses the development of an adaptive cooperative agent in a domain that suffers from human error in the allocation of attention. The design is discussed of a component of this adaptive agent, called Human Attention-Based Task Allocator (HABTA), capable of managing agent and human attention. The HABTA-component reallocates the human's and agent's focus of attention to tasks or objects based on an estimation of the current human allocation of attention and by comparison of this estimation with certain normative rules. The main contribution of the present paper is the description of the combined approach of design and validation for the development of such components. Two complementary experiments of validation of HABTA are described. The first experiment validates the model of human attention that is incorporated in HABTA, comparing estimations of the model with those of humans. The second experiment validates the HABTA-component itself, measuring its effect in terms of human-agent team performance, trust, and reliance. Finally, some intermediary results of the first experiment are shown, using human data in the domain of naval warfare.

1 Introduction

Several challenges can be identified for work on future naval platforms. Information volumes for navigation, system monitoring, and tactical tasks will increase as the complexity of the internal and external environment also increases [1]. The trend of reduced manning is expected to continue as a result of economic pressures and humans will be responsible for more tasks, tasks with increased load, and tasks with which they will have less experience. Problems with attention allocation are more likely to occur when more has to be done with less. To avoid these attention allocation problems, in this paper it is proposed that humans are supported by cooperative agents capable of managing their own and the human's allocation of attention. It is expected that these attention managers have a significant positive impact: when attentional switches between tasks or objects are often solicited, where the human's lack of experience with the environment makes it harder for them to select the appropriate attentional focus, or where an inappropriate selection of attentional focus may cause serious damage.

In domains like air traffic control (ATC) or naval tactical picture compilation these properties are found, even when the people involved are experienced.

The present study discusses the design and validation of a component of an adaptive agent, called *Human Attention-Based Task Allocator (HABTA)*, capable of managing agent and human attention. This component is based on two cognitive models: one that describes the current allocation of a humans attention and one that prescribes the way his attention should be allocated. If there is a discrepancy between the output of the two models, HABTA reallocates the tasks between the human and the agent, for instance depending on certain rules the human and agent agreed upon. Models of attention or situation awareness have already been developed and used to predict faults in attention allocation (e.g., the SEEV model [2]), but less is known about how they can be used to initiate agent adaptation, or automatic task reallocation more specifically. Furthermore, since in many domains (like ATC) it is the tasks altogether rather than mere visual stimuli that eventually require allocation of attention, the design and validation discussed in this paper is more focused on cognitive rather than visual attention. Off course the mentioned tasks also require visual attention, but all the time. Still other applied models mainly focus on visual attention. Finally, the applicability of a HABTA-based agent has not yet been investigated either.

This paper consists of the following sections. In Section 2 the psychological background of human error in the allocation of attention in the domain of naval warfare is shortly described. The understanding of these errors is important for the management of attention allocation. In Section 3 the design requirements of an agent-component *Human Attention-Based Task Allocator (HABTA)* are given. These requirements enable the agent to support the human-agent team by managing attention allocation of the human and the agent.

The main contribution of the present paper is the description of the combined approach of design and validation for the development of applied cooperative agent-components. In Section 4, two complementary methods of experimental validation against the in Section 3 stated design requirements are described. The first experiment validates the model of human attention that is incorporated in a HABTA-component. The validity of the model is determined by comparison of the model's and human's estimation of human attention allocation. The second experiment validates the HABTA-component itself, measuring its effect in terms of human-agent team performance, trust, and reliance. In Section 5 intermediary results of a pilot study are shown as a means to discuss the first experiment described in Section 4, using human data in the domain of naval warfare. In Section 6 the paper ends with concluding remarks and ideas for future research.

2 Human Error in the Allocation of Attention

As is mentioned in the introduction, the domain chosen in this research is naval warfare. One of the important tasks in naval warfare is the continuous compilation of a tactical picture of the situation (see for a description in more detail [3]). In a picture compilation task operators have to classify contacts that are represented on a radar display. The contacts can be classified as hostile, neutral or friendly, based on certain identification criteria (idcrits). Tactical picture compilation is known for its problems in the allocation of attention. To be able to identify contacts, contacts have to be monitored over time. This requires attention, but resources of attention are limited. When a task demands a lot of attention, less attentional resources are available for other tasks (e.g., [4, 5]. In general, two kinds of problems with human attention allocation can be distinguished: underallocation of attention and overallocation of attention.

Underallocation of attention means that tasks or objects that need attention do not receive enough attention from the operator. *Overallocation* of attention is the opposite: tasks or objects that do not need attention do receive attention. Overallocation of attention to one set of tasks may result in underattention to other tasks. Both under- and overallocation of attention can lead to errors. Experience, training, and interface design can improve these limitations, but only to a certain level. Efforts have been done, for example, to fuse tactical information on displays [6]. To be able to investigate whether a support system for attention allocation, like HABTA, can overcome these limitations of attention, it is important to understand these types of errors and more specifically in the domain of naval warfare. In Section 2.1 and 2.2, examples of errors of under- and overallocation when performing a tactical picture compilation task and their possible causes are described.

2.1 Underallocation of Attention

Underallocation of attention means that some objects or tasks receive less attention than they need according to certain normative rules for the task to be performed. Underallocation of attention occurs because of limited resources of attention or because of an incorrect assessment of the task.

When performing a tactical picture compilation task, operators have to monitor a radar screen where the surrounding contacts are represented as icons. The contacts on the screen have to be classified as neutral, hostile or friendly based on observed criteria. This is a complex task and it is essential that attention is allocated to the right objects. Inexperienced operators often allocate too little attention to contacts that they have previously classified as neutral [7]. When the behavior of these contacts changes to that of a hostile contact, this may not be observed because of underallocation of attention to those contacts. One reason for this could be that identity changes are not expected by the operator due to the fact that people are too confident in their identified contacts. Another reason might be that changes in relevant behavior of contacts are not salient enough to be observed without paying direct attention to those objects. Underallocation of attention to objects may also occur because of a lack of anticipatory thinking. This is the cognitive ability to prepare in time for problems and opportunities. In a picture compilation task, classification of contacts that are expected to come close to the own ship have priority over those that are not expected to come close. The reason for this is that there is less need to identify contacts when the own ship is out of sensor and weapon range of those contacts.

Therefore, inexperienced operators often direct their attention only to objects in the direction the ship is currently heading. When unexpected course change is needed because of emerging threats, the ship is sometimes headed toward an area with contacts that are not yet classified [7].

2.2 Overallocation of Attention

Apart from underallocation, overallocation of human attention is also a common problem. Overallocation of attention means that some objects receive more attention than needed according to certain normative rules. Overallocation of attention can occur for example, when operators overestimate the importance of a set of objects or tasks, while underestimating the importance of other objects or tasks. This occurs for example, when some contacts act like distractors and perform salient behavior. Comparable to visual search tasks where objects with salient features generate a pop-out effect (e.g., [8]), those contacts directly attract the attention of the operator (bottom-up). Especially inexperienced operators overrate those salient cues and allocate too much attention to those contacts [7]. Another possibility is that irrelevant behavior of objects is highly salient due to the manner information is presented on the interface. For instance, when a contact's behavior is unexpected, but not threatening, attention is unnecessarily drawn to this contact. In this case, the correct and quick application of identification rules will result in neutral identity and resources become available for the identification of other contacts.

3 Design Requirements

The goal of the efforts described is to come to a generic methodology for developing a component for an agent that supports humans with the appropriate allocation of attention in a domain that suffers from human error in the allocation of attention. As mentioned in Section 2, human attention allocation is prone to two types of errors with several possibilities as causes, such as inexperience and information overload.

In this section the design requirements of an agent-component is described that enables agents to determine whether objects or tasks that are required to receive attention indeed do receive attention, either by the human or the agent, and to intervene accordingly. The component is called an *Human Attention-Based Task Allocator (HABTA)*-component, since it *bases* its decisions to intervene on estimations of *human attention* and intervenes by *reallocating tasks* to either human or agent. It is expected that the combined task performance of the human-agent team will be optimized when the agent consists of such a HABTA-component. This work builds forth on earlier work. In [9] some of the possibilities are already discussed of dynamically triggering task allocation for tasks requiring visual attention, and in [10, 11] the real-time estimation of human attentional processes in the domain of naval warfare is already discussed.

Properly stated design requirements are important for the design of effective agent-systems for a certain purpose and for validating whether the design meets the requirements for that purpose. A HABTA-component has four design requirements, which are the following:

1. It should have a descriptive model, meaning an accurate model of what objects or tasks in the task environment receive the human's attention,
2. It should have a prescriptive or normative model, meaning an accurate model of what objects require attention for optimal task performance,
3. It should be able to reliably determine whether actual attention allocation differs too much from the required attention allocation,
4. It should be able to support by redirecting attention or by taking over tasks such that task performance is improved.

In Fig. 1 the design overview of a HABTA-component is shown that corresponds to the above design requirements. The setting in this particular overview

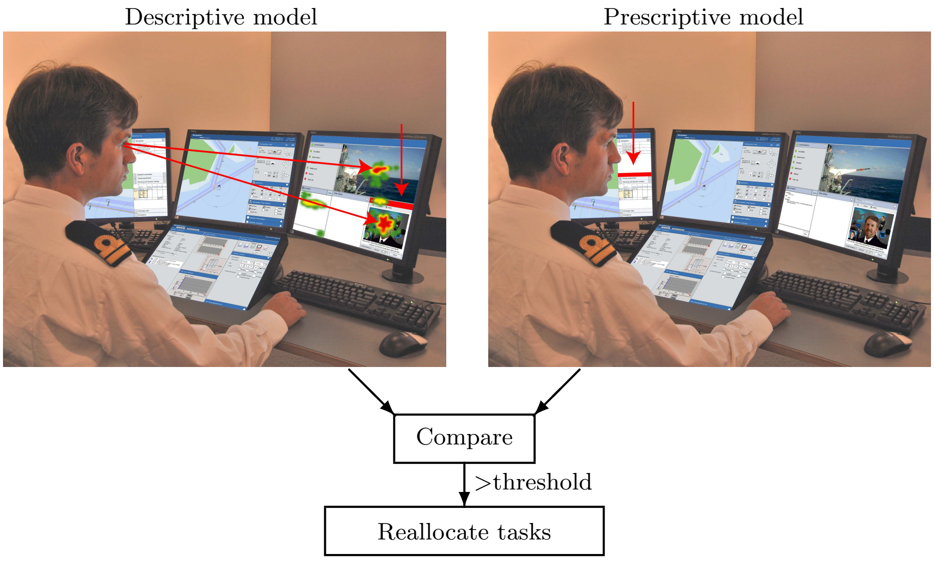


Fig. 1. Design overview of a HABTA-component for a future integrated command and control environment. The discrepancies between the output of the descriptive and prescriptive model result in a reallocation of tasks. The workstation shown in the pictures is the Basic-T [12].

is a naval officer behind an advanced future integrated command and control workstation and compiling a tactical picture of the situation. If the agent cooperatively assists the officer, than the agent should have a descriptive (Requirement 1) and normative model (Requirement 2). When the operator allocates his attention to certain objects or tasks that also require to receive attention, the outcome of both models should be comparable. This means that output of the models should not differ more than a certain threshold. The output of the two models in the example shown in Fig. 1 are clearly different: in the left image, the operator is attending to different objects and corresponding tasks than the right image indicates as being required (see arrows). Because of this discrepancy, which the HABTA-component should be able to determine (Requirement 3), an adaptive reaction by the agent is triggered (Requirement 4). This means that, for instance, the agent either will draw attention to the proper region or task through the workstation, or it will allocate its own attention to this region and starts executing the tasks related to that region, for the given situation.

To prevent that HABTA-based support results in automation surprises, the human-agent team should be able to make and adjust agreements about how they work as a team. It may be, for example, that the human does not want to be disturbed, and the agent is supposed to allocate tasks solely to itself. This option requires a higher form of autonomous task execution by the agent. The other possibility is that the human wants to stay in control as much as possible and therefore only wants to be alerted by the agent to attend to a certain region or execute a certain task. The choice of the agent's autonomy or assertiveness can also depend on a certain estimate of the urgency for reallocating tasks. In the case of tactical picture compilation, human and agent should agree on whether the agent is allowed to take over identification tasks for contacts that are overlooked or not.

On the one hand, the human may be preferred to be dealing with an arbitrary region or task, because the human may have certain relevant background knowledge the agent does not have. But on the other hand, the human is not preferred to be allocated to all objects or tasks at once, because, in a complex scenario, he has limited attentional resources. Hence humans cannot be in complete control, given the fact that both human and agent need each other. Optimal performance is only reached when human and agent work together as a team. Human-agent team work is expected to be effective when the right support is provided at the right time and in the right way. An obvious goal, but there are some potential obstacles in achieving it. Descriptive and prescriptive (normative) models of attention allocation may be inaccurate. Objects that require or receive attention may not be in the output of the descriptive or normative models, respectively. Similarly, objects that do not require or receive attention may be in the output of the models. The agent may conclude that descriptive and normative models differ when they do not, and vice versa. The system may be assertive and wrong, or withholding but right. Attention may be redirected to the wrong region or the wrong set of objects, or tasks are taken over by the agent that should be taken over by the human. Because of the complexity of these consequences of the above design requirements, both the validity of the model and the effectiveness of the agent's HABTA-component should be investigated and iteratively improved. This procedure of investigation and improvement is described in Section 4.

4 Validation

As described in Section 3, HABTA-components require a descriptive and prescriptive model of attention to support attention allocation of humans in complex tasks. Before HABTA-components can be used to support humans, they have to be validated. Validation is the process of determining the degree to which a (cognitive) model is an accurate description of human (cognitive) phenomena from the perspective of the intended use of the model. Again referring to Section 3, for the intended use mentioned in this paper, this means that HABTA-components have to meet the design requirements (1–4) in Section 3.

In the near future two experiments will be carried out to validate a HABTA-component. In Experiment 1 the descriptive model will be validated and optimized (Requirement 1). This experiment aims at determining the sensitivity (d') of the model by comparing it with data retrieved from human subjects executing a complex task that causes problems with attention allocation. Based on the results of the experiment, the d' of the model can be improved by optimizing it off-line against a random part of the same data. It is expected that the higher d', the better the HABTA-component will be able to support the human. If the d' of the descriptive model is not high enough, the HABTA-component will support at the wrong moments and for the wrong reasons, which obviously leads to low performance, trust, and acceptance. In Experiment 2 the applicability of the (improved) descriptive model for attention allocation support is tested (Requirements 2–4). It will be investigated if the support of an agent with the HABTA-component leads to better performance than without HABTA-component.

The remainder of this section is composed of three parts. In Section 4.1 the task that will be used in the above mentioned experiments is described in more detail. After that, the specific experimental design and measurements of the experiments are described in Sections 4.2 and 4.3, respectively. Both experiments still have to be carried out. Preliminary results from a pilot of Experiment 1 will be described in Section 5.

4.1 Task Description

The task used in both experiments is a simple version of the identification task described in [13] that has to be executed in order to buildup a tactical picture of the situation. In Fig. 2 a snapshot of the interface of the task environment is shown.[3] The goal is to identify the five most threatening contacts (ships). In order to do this, participants have to monitor a radar display where contacts in the surrounding areas are displayed. To determine if a contact is a possible threat, different criteria have to be used. These criteria are the identification criteria (idcrits) that are also used in naval warfare, but are simplified in order to let naive participants learn them more easily. These simplified criteria are the speed, heading, bearing, and distance of a contact to the own ship, and whether the contact is in a sea lane or not. When the participant clicks on a contact with

[3] A full color variant of Fig. 2 can be found at [14].

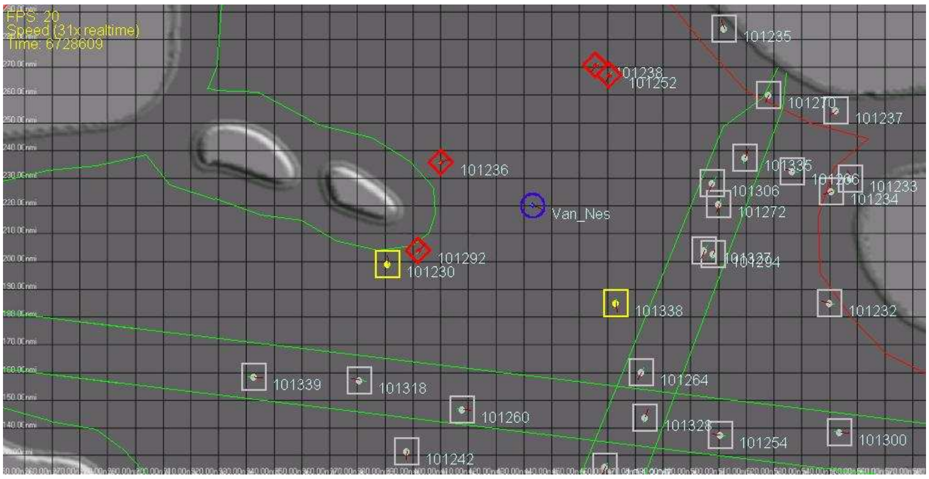


Fig. 2. The interface of the used simplified task environment based on [13]. The green lanes are sea lanes. The blue circle labeled with "Van_Nes" represents the own ship.

the right mouse button this information is displayed. If a participant concludes that a ship is a possible threat or not, he can change the color of the contacts by clicking with the left mouse button on the contact. Contacts can be identified as either a threat (red), possible threat (yellow), or no threat (green). It is not necessary that all contacts are identified. Only the five most threatening have to be identified as a threat (marked as red). The other types of identification (possible threat and no threat) are used to assist the participant in his task. When a contact is marked as green, this means no direct attention is needed. When a contact is marked as yellow, this contact has to be checked regularly to decide if the contact is still no threat. The task has to be performed as accurately as possible. Contacts that are wrongfully identified as a threat will result in a lower score. Performance is determined by the accurateness, averaged over time, of the contacts that are identified as the five most threatening contacts during the task. Behavior of each contact can change during the task and therefore the soundness of classifications (which is not communicated to the participant) may change over time. For instance, a contact can suddenly come closer to the own ship, get out of a sea lane, speedup, or change its bearing or heading.

For Experiment 2 (see Section 4.3) the task is extended to one that includes the support of the HABTA-based agent. The support agent is capable of doing the same as the operator, except with limited background knowledge and therefore limited performance per object. In order to simulate this aspect, for each participant, the measured average performance per contact in Experiment 1 is used in order to set the performance of the agent. The agent can be given a list of objects provided by the its HABTA-component and compile a tactical picture related to those objects.

4.2 Experiment 1: Validation of the Descriptive Model

In Experiment 1 participants perform the task as described in the previous section without support of the agent. The same scenario will be used for all participants. Before the actual task starts, the task will be explained thoroughly to the participants. The task will be illustrated by using different examples to be sure that the participants understand the task and how to decide if a contact is a threat based on the different criteria. All participants have to perform a test to check if they sufficiently understood the rules of classifying the contacts. If they do not perform well, i.e. their score is blow 80%, they receive extra instructions and another test. Also the possible second test has to be performed with a success rate of above or equal to 80%. Then they have to perform a practice trial in which they have to apply the learned rules. After this they get instructions of how to behave when there during the experimental interventions while they are executing their task. This is practiced as well several times, after which the actual experiment begins.

During the task, different variables are measured to determine the d' of the model and to be able to iteratively improve the model afterwards. The following variables will be measured: eye movements, performance, mental workload and at different points in time participants have to mark contacts that received attention according to the participant. The performance and mental workload measures are used as a baseline for comparing the performance and mental workload of the task with and without using the HABTA-component (see Experiment 2). In order to measure the variables, at random moments (varying from 4–6 minutes) the scenario is frozen. During a freeze, the participants have to click on the contacts to which, in their opinion, they had allocated their attention the moment right before the scenario was frozen. The participants also have to motivate why those contacts are selected. Directly after the participant has selected the contacts, mental workload is measured during the same freezes. For this, the mental workload scale from [15] is used (BSMI). On a scale from 0 (not at all strenuous) to 150 (very strenuous) the mental workload of the task has to be indicated. Performance and eye movements of the participants are measured during the task, by calculation according to the rules described in Section 4.1 and by eye-tracker recording, respectively. The patterns of the eye movements (what objects are looked at through time) are compared with the contacts that received attention before the freezes, according to the participant. This is done to be sure that the participants were able to select the objects that received their attention. Those contacts that got a considerate amount of gaze fixations, are expected to have received attention.[4] If the participants do not mention those contacts, it is expected that they are not good at selecting the proper contacts.

After the experiment is performed, the contacts selected by the participants during the freezes are matched with the output of the model in a simulation. The calculation of d' provides information about the sensitivity of the model, i.e. whether the model is able to accurately describe the participant's dynamics

[4] Note that this does not hold vice versa, which would otherwise mean that attention in complex scenarios is easily described using solely fixation data.

of attention allocation. Information about performance, workload, and the description of the participants why contacts are selected, is expected to be valuable for determining in what cases the percentage true positives (hits) is high and percentage false positives (false alarms) is low, which in turn can be used to improve the sensitivity of the model. See Section 5 for the illustration of this process.

4.3 Experiment 2: Validation of the HABTA-Based Support

In Experiment 2 the applicability of the model for supporting attention allocation is tested. The same task as in Experiment 1 has to be performed, except this time the participant is supported by the agent of which the HABTA-component is part of. When there is a discrepancy between the descriptive and prescriptive model, higher than a certain threshold (see Fig. 1), the agent will support the human by either performing the task for the participant or by drawing attention to the contact that should receive attention. Different variables are measured to determine the excess value of the HABTA-based support. Performance and mental workload are measured in the same way as in Experiment 1. Furthermore, trust and acceptance are measured at the end of the scenario.

In order to determine the effectiveness of an agent, it is also important to measure trust and acceptance of that agent and to investigate what factors influence trust and acceptance. Trust and acceptance indicate whether people will actually use the agent. For instance, it says something about whether people will follow the advice of the agent, in the case the agent provides advice. Validated questionnaires are adjusted to be able to measure trust and acceptance in adaptive systems. The trust questionnaire is based on the questionnaire of [16]. An example of a question on this questionnaire is: "Is the agent reliable enough?". The acceptance questionnaire is based on the questionnaire of [17] and [18]. An example of a question on this questionnaire is: "Is the support of the agent useful for me?". The trust and acceptance scores are expected to provide more insight in the results of the experiment. If trust in and acceptance of the agent is low, people will not follow any suggestions made by the agent.

The performance and mental workload without a HABTA-based agent will be compared with those with a HABTA-based agent, using the results of Experiment 1 as a baseline. This is one of the reasons that the same participants are used as in Experiment 1. The other reason is that the measured performance in Experiment 1 is used for setting the performance of the agent. For Experiment 2, it is expected that performance is higher and mental workload is lower when supported with HATBA.

5 Intermediary Results

In this section preliminary results of the experiments described in Section 4 are shown based on a pilot study for Experiment 1, using one arbitrary participant. The actual experiment will be performed with more participants. The pilot is primarily meant to explore the applicability of the experimental method of Experiment 1 to the given task. It is also meant as an illustration of the form and dynamics of the participant's and model's estimation of human allocation of attention. Finally, it is used as a basis for a better understanding of the possibilities of HABTA-based support, which is important for a proper preparation and performing of Experiment 2. This is because this type of support is required in the experimental setup of Experiment 2.

In the pilot study, the participant was required to execute the identification task and to select contacts during the freezes. In contrast with the procedure during the actual experiment, no questions concerning the participant's cognitive workload or motivation for the selected contacts were asked. In Fig. 2 the interface right before a freeze is shown. During a freeze both the participant and the model had to indicate their estimation of what contacts the attention of the participant was allocated to. In the situation presented in Fig.2, the participant selected contacts 101238, 101252, 101236, 101338, 101230, 101292, 101294, and 101327. Between every two freezes certain events can cause the participant to change the allocation of his attention to other attention demanding regions. The preceding course of events of the situation in Fig.2 clearly caused the participant to attend to the contacts close to his own ship "Van_Nes". If the model made a proper estimation of the participant's allocation of attention, the selected contacts by the participant would resemble those selected by the model. Consequently, the performance of the model is best determined by means of the calculation of the overall overlap of the participant's and model's selection of contacts. This calculation is explained below.

There are four possible outcomes when comparing the participant's and model's selection of contacts, namely, a Hit, False Alarm, Correct Rejection, and Miss. The counts of these outcomes can be set out in a 2×2 confusion matrix. Tab. 1 is such a confusion matrix, where T and F are the total amount of the participant's selected and not selected contacts, respectively, and T' and F' are the total amount of the model's selected and not selected contacts, respectively. The ratios of all the possible outcomes are represented by H, FA, CR,

		Participant		
		t	f	$total$
Model	t'	Hits	False Alarms	T'
	f'	Misses	Correct Rejections	F'
	$total$	T	F	

Table 1. Confusion matrix of the participant's and model's estimation of the allocation of attention.

and M, respectively. A higher H and CR, and a lower FA and M, leads to a more appropriate estimation by the model. This is the case because the selected contacts by the model then have a higher resemblance with those selected by the participant. Furthermore, a higher T' leads to a higher H, but, unfortunately, also to a higher FA. Something similar holds for F'. The value of T' therefore should depend on the trade-off between the costs and benefits of these different outcomes.

In Fig. 3 the $15 \times 10 \times 1$ output of the model for the situation presented in Fig. 2 is shown. If the estimated attention on the z-axis, called Attention Value (AV), is higher than a certain threshold, which is in this case set to .035, the contact is selected and otherwise it is not. The different values of AV are normally distributed over the (x, y)-plane. The threshold is dependent on the total amount of contacts the participant is expected to allocate attention to [10]. The AV-distribution in Fig. 3 results in the selection of contacts 101235, 101238, 101252, 101236, 101292, 101230, 101338, and 101260. Using this selection and the selected contacts by the participant, for each contact, the particular outcome can be determined. For each freeze, if one counts the number of the different outcomes, a confusion matrix can be constructed and the respective ratios can be calculated. For Fig. 3, for example, these ratios are $H = \frac{6}{8} = 0.750$, $FA = \frac{2}{19} = 0.105$, $CR = \frac{17}{19} = 0.895$, and $M = \frac{2}{8} = 0.250$, respectively.

To study the performance of models Receiver-Operating Characteristics (ROC) graphs are commonly used. A ROC-space is defined by FA as the x- and H as the y-axis, which depicts relative trade-offs between the costs and benefits of the model. Every (FA, H)-pair of each confusion matrix represents one point in the ROC-space. Since the model is intended to estimate the participant's allocation of attention for each freeze and participant, this means that for N participants and M freezes, there are NM points in the ROC-space.

Once all points have been scatter plotted in the ROC-space, a fit of an isosensitivity curve leads to an estimate of the d' of the model. Isosentitivity corresponds to:

$$d' = z(H) - z(FA)$$

where d' is constant along the curve and $z(x)$ is the z-score of x.[5] Larger absolute values of d' mean that the model is more specific and sensitive to the participant's estimation (and thus has a higher performance). If d' is near or below zero, this indicates the model's performance is equal to or below chance, respectively. If there does not exist a proper fit of a isosensitivity curve, the area under the curve (AUC) can also be used as a model validity estimate. In non-parametric statistics the ROC-graph is determined by the data and not by a predefined curve. If the different values of H and FA appear to be normally distributed, the d' can be obtained from a z-table. In this case, the (FA, H)-pair from Fig. 3 results in $d' = 1.927$. Which is a fairly good score.

[5] The z-score reveals how many units of the standard deviation a case is above or below the mean:

$$z(x_i) = \frac{x_i - \mu_x}{\sigma_x}$$

where μ_x is the mean, σ_x the standard deviation of the variable x, and x_i a raw score.

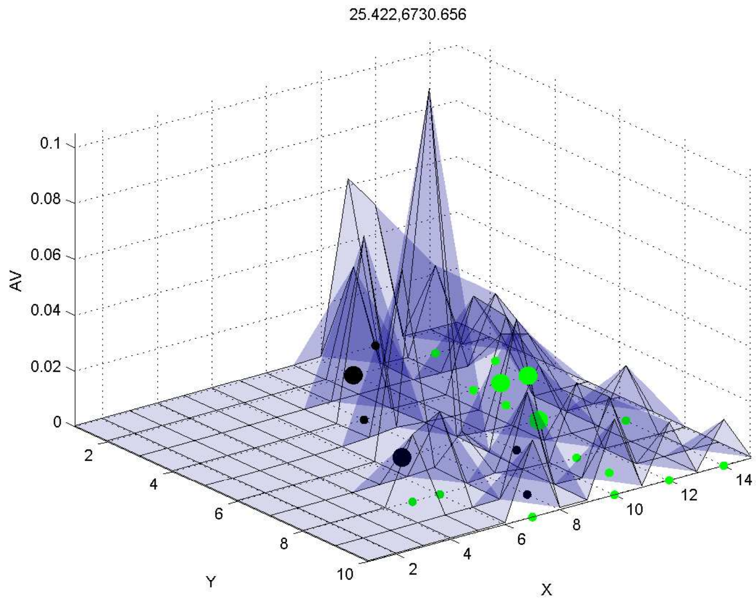


Fig. 3. The output of the model for the situation shown in Fig. 2. The black dots are the selected contacts by the model. Bigger dots mean that there are more contacts on the respective coordinates.

6 Conclusion and discussion

This paper describes the development of an adaptive cooperative agent to support humans while performing tasks where errors in the allocation of attention occur. In general, human attention allocation is prone to two types of errors: over- and underallocation of attention. Several factors may cause over- or underallocation of attention, such as inexperience and information overload. The design is discussed of a component of an agent, called Human Attention-Based Task Allocator (HABTA), that is capable of detecting human error in the allocation of attention and acts accordingly by reallocating tasks between the human and the agent. In this way the HABTA-based agent manages human and agent attention, causing the performance of the human-agent team to increase. The development of such an agent requires extensive and iterative research. The agent's internal structure, i.e. the models describing and prescribing human attention allocation and the support mechanism that is based on those models, has to be validated. In this paper, two experimental designs are described to validate the internal of the agent. The first experiment aims at validating the model of human attention allocation (descriptive model) and the second experiment aims at validating the HABTA-component as a whole, incorporating a prescriptive model and support mechanism.

The results from the pilot of the first experiment presented in this paper have proven to be useful, but the actual experiments still have to be performed. Therefore, future research will focus on the performance and analysis of these experiments. It is expected that the accuracy of the model can be increased hereafter, however 100% accurateness will not be attainable. The results of the first experiment will show if the variables indeed provide enough information to improve the accurateness of the model.

With respect to the second experiment, one might argue to add another variant of support, such as one that is configurated by the participant itself. The participant will then do the same as HABTA does, which might result in him being a fair competitor for HABTA. In this way the effectiveness of HABTA-based support can be studied more convincingly, comparing human-agent performance when either the participant or the agent is managing attention allocation. Deciding on this will be subject in the near future.

If the agent does not support the human at the right time and in the right way, this might influence trust and acceptance of the agent. It is interesting to investigate whether an observable and adjustable internal structure of the agent improves trust and acceptance of the system (see for instance [19] in these proceedings). This also needs further research.

In this paper the development and validation of a normative model (prescriptive model) is not described. Validation of this model is important, as it is also a crucial part of the HABTA-component. Errors in this model will lead to support at the wrong time and this will influence performance, trust, and acceptance. Further research is needed in order to develop and validate normative models.

Finally, in general, agent-components have more value when they can be easily adjusted for other applications. It is therefore interesting to see whether HABTA-based support can be applied in other domains as well.

7 Acknowledgments

This research was funded by the Dutch Ministry of Defense under progr. nr. V524. The authors would like to thank Annerieke Heuvelink, Harmen Lafeber, Jan Treur, Kim Kranenborg, Marc Grootjen, Tibor Bosse, Tjerk de Greef, and Willem van Doesburg for their contribution and helpful comments.

References

1. Grootjen, M., Neerincx, M.: Operator load management during task execution in process control. In: Human Factors Impact on Ship Design. (2005)
2. Wickens, C., McCarley, J., Alexander, A., Thomas, L., Ambinder, M., Zheng, S.: Attention-situation awareness (a-sa) model of pilot error. Technical Report AHFD-04-15/NASA-04-5, University of Illinois Human Factors Division (2005)
3. Chalmers, B.A., Webb, R.D.G., Keeble, R.: Modeling shipboard tactical picture compilation. In: Proceedings of the Fifth International Conference on Information Fusion. Volume 2., Sunnyvale, CA, International Society of Information Fusion (2002) 1292–1299
4. Kahneman, D.: Attention and effort. Prentice Hall, Englewoods Cliffs, NJ (1973)
5. Wickens, C.: Processing resources in attention. Varieties of attention (1984)
6. Steinberg, A.: Standardisation in data fusion. In: Proceedings of Eurofusion'99: International Conference on Data Fusion, UK, Stratford-Upon-Avon (1999) 269–277
7. Verkuijlen, R.P.M., Muller, T.J.: Action speed tactical trainer review. Technical report, TNO Human Factors (2007)
8. Treisman, A.: The perception of features and objects. Attention: Awareness, selection, and control (1993)
9. Bosse, T., Doesburg, W.v., Maanen, P.-P. van, Treur, J.: Augmented metacognition addressing dynamic allocation of tasks requiring visual attention. In Schmorrow, D.D., Reeves, L.M., eds.: Proceedings of the Third International Conference on Augmented Cognition (ACI) and 12th International Conference on Human-Computer Interaction (HCI'07). Volume 4565 of Lecture Notes in Computer Science., Springer Verlag (2007)
10. Bosse, T., Maanen, P.-P. van, Treur, J.: A cognitive model for visual attention and its application. In Nishida, T., ed.: Proceedings of the 2006 IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT-06), IEEE Computer Society Press (2006) 255–262
11. Bosse, T., Maanen, P.-P. van, Treur, J.: Temporal differentiation of attentional processes. In Vosniadou, S., Kayser, D., eds.: Proceedings of the Second European Cognitive Science Conference (EuroCogSci'07), IEEE Computer Society Press (2007) 842–847
12. Arciszewski, H., Delft, J.v.: Automated crew support in the command centre of a naval vessel. In: Proceedings of the 10th International Command and Control Research and Technology Symposium. (2005)

13. Heuvelink, A., Both, F.: Boa: A cognitive tactical picture compilation agent. In: Proceedings of the 2007 IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT 2007), IEEE Computer Society Press (2007) forthcoming
14. http://www.few.vu.nl/~pp/public/vanmaanendekoningvandongen-AmI07.pdf
15. Zijlstra, F.R.H., Doorn, L.v.: The Construction of a Scale to Measure Perceived Effort. Department of Philosophy and Social Sciences, Delft University of Technology, Delft, The Netherlands (1985)
16. Madson, M., Gregor, S.: Measuring human-computer trust. In: Proceedings of the Australasian Conference on Information Systems. (2000)
17. Venkatesh, V., Morris, M.G., Davis, G.B., Davis, F.D.: User acceptance of information technology: Toward a unified view. MIS Quarterly **27**(3) (2003) 425–478
18. Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly **13**(3) (1989) 319–340
19. Mioch, T., Harbers, M., van Doesburg, W.A., van den Bosch, K.: Enhancing human understanding through intelligent explanations. In: Proceedings of the First International Workshop on Human Aspects in Ambient Intelligence, Springer Verlag (2007) forthcoming

Affective Human Factors Design with Ambient Intelligence

Jianxin (Roger) Jiao, Qianli Xu and Jun Du

School of Mechanical & Aerospace Engineering,
Nanyang Technological University, Singapore
jiao@pmail.ntu.edu.sg

Abstract. The need to include customer's affective needs in product design presents a new direction beyond tradtional human factors and ergonomics. While the human-product interactions have been extensively studied, the interactions of these elements with the ambience have been largely ignored. This gives rise to a nascent research perspective called affective human factors design, which aims at addressing human's emotional responses and aspirations and to achieve aesthetic appreciation and pleasurable experience through human-product-ambience interactions. This paper presents a framework of affective human factors design with ambient intelligence to achieve the extensive interactions among these elements. Ambient intelligence establishes a multidisciplinary technology core that incorporates affective design, human factors and ergonomics, product development, and specific application sectors. A few application scenarios reveal the most important characteristics and emerging trends in this research area.

Keywords: Affective Design, Human Factors, Ambient Intelligence.

1 Introduction

The interaction between human factors and design is a common theme in the literature and a constant challenge in the practical working-out of human use issues in the design of products and systems. It is suggested that the discourse has moved on from the goal of merely attaining system usability [1]. The 'new' human factors must further support designs that address peoples' emotional responses and aspirations, whereas usability alone still demands a great deal of attention in both research and practice. Consideration of these needs has generally fallen within the designer's sphere of activities, through the designer's holistic contribution to the aesthetic and functional dimensions of human-product interactions. They have thus tended to be interpreted and explored through creative, subjective design processes rather than through the application of analytical, objectively determined methods. Such a rationale of product design represents a nascent research perspective, namely affective human factors design.

Affective human factors design originates from the field of human-computer interaction and more specifically from the developing area of affective computing. It used to address the delivering of affective interfaces capable of eliciting certain emotional experiences from users. Similarly, affective product design attempts to define the subjective emotional relationships between consumers and products, and to explore the affective properties that products intend to communicate through their physical attributes. It aims to deliver artifacts capable of eliciting maximum physiopsychological pleasure that consumers may obtain through all of their senses.

The fulfillment of affective design necessities a new dimension beyond the traditional human-product interactions, namely the ambience. Accordingly, a coherent consideration of the interactions between product, human, and ambience suggests to be more profound for the discipline of affective human factors. Such a consideration is in line with the wisdom of product ecosystems [2], which essentially entails a scenario of affective design of the entire system with customer experience in the loop (Fig. 1). Products can interact with its ambience and such an interaction influences the customer's perceptions due to the particular context created. Hence, affect is a combination of two elements, namely customer perception and customer experience. Accordingly, the aim of affective design is to address human's emotional responses and aspirations, and to achieve aesthetic appreciation and pleasurable experience through human-product-ambience interactions.

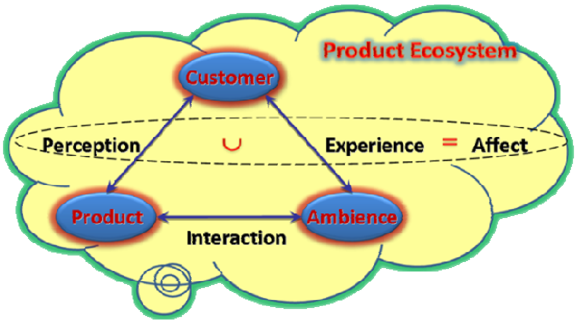


Fig. 1. Interactions between human, product, and ambience in a product ecosystem

However, it remains a challenging task to model and evaluate product ecosystems with considerations of affective human factors. Above all, the construction of a product ecosystem requires extensive human-product-ambience interactions, which could not be realized without advanced product technologies and enhanced analytical methods. In this respect, ambient intelligence suggests itself to be a promising solution to product ecosystem design. It takes the integration provided by ubiquitous and pervasive computing one step further to realize context-aware environments that are sensitive and responsive to the presence of people [3].

Considering the challenges of affective human factors design and the opportunities provided by ambient intelligence, this paper aims at developing a comprehensive solution framework of affective design with ambient intelligence by incorporating human factors and ergonomics, information and communication technologies, engineering design, product innovation, and social and psychological sciences. Section 2 presents the overall structure of the framework. The major technical issues and the proposed solution strategies are discussed in Section 3, together with application examples for illustrating the rationale. Summary and possible research directions are discussed in the last section.

2 A Framework of Affective Human Factors Design

The key issue of affective human factors design with ambient intelligence manifests itself through the development and utilization of electronic intelligent environments that are sensitive and responsive to the presence of people. This research explores the potential of applying ambient intelligence technologies for the enhancement of acquisition, analysis, and fulfillment of affective customer needs. In particular, the proposed framework is geared towards the following goals: (1) examine the feasibility and potential of ambient intelligence concept and establish the frame of references in the respective affective human factors design domains; (2) identify fundamental issues and formulate a technological roadmap of affective human factors design research.

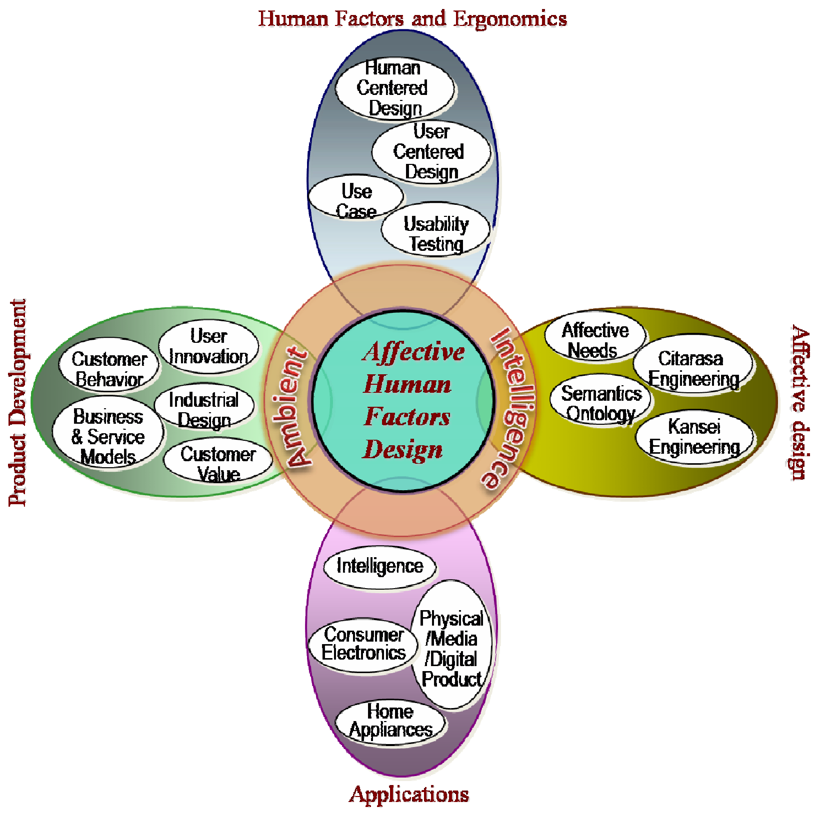


Fig. 2. Framework of the affective human factors design with ambient intelligence

Towards this end, a comprehensive technical framework is proposed, as shown in Fig. 2. This framework features a technology core of ambient intelligence in support of affective human factors design. It expands in four domain areas, including affective design, human factors and ergonomics, product development, and the specific application sectors. Among them, affective design and human factors and ergonomics define the problem context and the fundamental requirements of product ecosystem design. Product development accommodates the traditional engineering considerations in design and production. The application sectors provide various scenarios with respect to the particular requirements. Innovative solutions to affective human factors design may be developed by taking useful ingredients from these multiple disciplines.

3 Fundamental Issues

3.1 Ambient Intelligence

The strength of an ambient intelligence environment is to support affective design with context-aware adaptive applications. With ambient intelligence embedded in the product ecosystem, the behaviors and reactions of the customers can be captured in real time without interrupting the customers' normal activities. In this regard, a multi-layer ambient intelligence architecture is proposed as shown in Fig. 3. Different layers serve to link low-level details (hardware layers) with high-level view (software layers). The major responsibilities of each layer are discusses next.

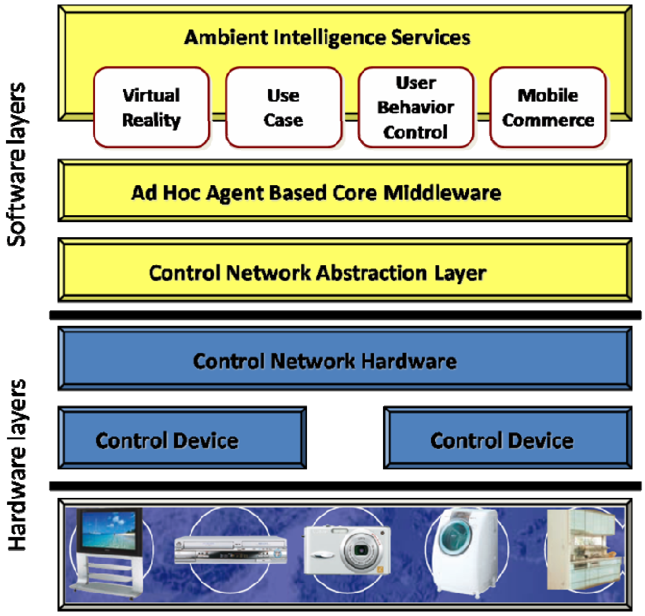


Fig. 3. A multi-layer ambient intelligence environment for affective design

- The bottom level layer is composed of domestic devices that comprise diverse virtual reality-enabled affective use cases such as home appliances in a living room or kitchen.
- A control device layer is defined to add a computing unit at the domestic device layer in order to obtain a particular application of pervasive computing.
- The control network hardware layer allows the interconnection among control devices, and thus satisfying ubiquitous communication.
- The control network abstraction layer is designed based on fuzzy markup language and is used to integrate different standards used to realize the control network layer.
- The ad hoc agent-based core middleware layer resorts to mobile agent technologies and is responsible of allocating the limited computation resources to numerous devices.
- At the top service level, various techniques, such as virtual reality, use case, data mining, mobile commerce, etc., are used to improve the elicitation of customers' affective needs and mapping of these needs to affective design elements.

3.2 Affective Design

A general roadmap can be established in terms of a series of affective mapping processes, including affective needs elicitation, analysis, and fulfillment. A few major issues and solution strategies are discussed next.

3.2.1 Affective Needs

As mentioned earlier, affect refers to customer's psychological response to the perceptual design details (e.g. styling) of the product [12]. Affect is a basis for the formation of human values and human judgment. For this reason, it might be argued that models of product design that do not consider affect are essentially weakened [13]. However, it has been difficult for customers to articulate the affective needs, and hence for designers to understand these needs. Such difficulties could be alleviated by well-defined semantic ontology and various analytical techniques, such as data mining [17] and conjoint analysis [10].

3.2.2 Semantic Ontology

The purpose of semantic ontology is to describe the affective needs that are communicable among customers and designers from different sectors using a limited number of terminologies that are as small as possible but comprehensive enough to cover the majority aspects of affective design. Each type of product ecosystem is supported by a set of affective terminology/taxonomy based on different customer requirements with respect to the particular product systems (e.g., automobile, living room, shopping mall, etc.). The development of semantics starts with customer survey, preferably carried out in an ambient intelligence-enabled environment. Next, semantic scales can be constructed for product ecosystem evaluation, which involves the collection of a large number of descriptive words for the product ecosystem, and the clustering of the words that are similar in meaning into categories [14]. From each category, one or several words are chosen to represent the category and be used on a semantic scale in order to evaluate the product ecosystem. Finally, the semantic scale assessment can be interpreted by domain experts to delineate the usage of the terminologies.

3.2.3 Citarasa Engineering

Citarasa engineering is a new methodology that integrates cognition and affect in uncovering customer needs [15]. *Cita* is a Malay word that denotes intent, hope, aspiration, and expectation that are cognitively-linked, while *rasa* indicates taste and feels that is related to affect/emotion. While affect refers to feeling responds, cognition is used to interpret, make sense of, and understand user experience [16]. Citarasa engineering hosts a combination of techniques that will be applied in the elicitation of customer needs, measurement of customer satisfaction, the identification of mapping relationship between customer needs and design elements, and the design of product portfolio to fulfill the customer needs.

3.2.4 Kansei Engineering

Developed in Japan from the 1970s, Kansei engineering is a technology for translating a consumer's feeling and image of a product into design elements [7]. Kansei in the Japanese language contrasts with Chisei. It is the subjective feeling and aesthetic aspect of customers' mindset, as opposed to Chisei's rational knowledge set. Both together determine how people interact with the environment around them.

While Kansei words excel in describing affective needs, the mapping relationships between Kansei words and design elements are often not clearly available throughout the innovation process of industrial design. Designers are often not aware of the underlying coupling and interrelationships among various design elements in regard to the achievement of customers' affective satisfaction. It is necessary to discern customer needs and product specifications, and as a result the mapping problem in between is the key issue in 'design for customers'. This leads to that affective design involves not only general fields of engineering design and industrial design, but also a number of issues related to the system-wide operations.

3.3 Human Factors and Ergonomics

Human factors and ergonomics have traditionally concerned with cause and effect relations between products and human performance, as measured by responses to tasks or by physical pain. With considerations of affective needs, it is expanding its remit to include emotional relations with the phrase 'cause and affect', in contrast to effect [19]. Research in this area requires extensive investigation of human-product interactions.

3.3.1 Human- or User-Centered Design

Human- or user-centered design process for interactive systems stands for the title of ISO 13407/1999. It supports managers of hardware and software design projects, and reflects realization of the importance of users in human-computer interface design [4]. As shown in Fig. 4, it advocates specifying the context of use for an identified need, translating that to requirements, producing and evaluating design solutions, and iterating until a satisfactory solution is attained. It implies a process but does not offer the methods for use within it. Its context is purely focused usability ergonomics; but it could likewise be applied to tackle affective issues.

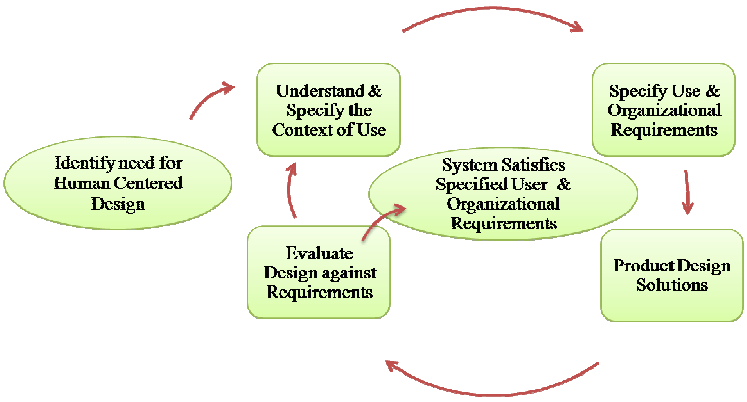


Fig. 4. Human/User-center design processes for affective design

3.3.2 Use Case

A use case is a popular technique for capturing functional requirements of systems. It allows description of sequences of events that, taken together, lead to a system doing something useful [5]. Each use case provides one or more scenarios that convey how the system should interact with the users (called actors) to achieve a specific business goal or function. Use case actors may be end users or other systems. Within the context of affective design, use cases are employed to build up testing scenarios for probing a user's tangible and intangible requirements via interacting with the product and its operating environment.

3.3.3 Usability Testing

Usability testing is a useful tool of measuring how well people can use a product and interact with the environment. Usability testing generally involves a controlled experiment to determine to what extent the users respond in terms of time, accuracy, recall, and emotional response. The results of the first test can be treated as a baseline or control measurement, whilst all subsequent tests can be compared with the baseline to justify improvement [6].

3.4 Product Development

Affective human factors design generally involves two stages: to understand customers' affective needs and subsequently to fulfill these needs in terms of product and system design, as shown in Fig. 5. While existing methods are mostly devoted to the first stage, the translation of affective needs into the product and system is even more challenging a task. In most cases, it is very hard to capture the customers' affective needs due to their linguistic origins. Since subjective impressions are difficult to translate into verbal descriptions, affective needs are relatively short-lasting emotional states and tend to be imprecise and ambiguous. Sometimes, without any technical experience, the customers do not know what they really want until their preferences are violated. In practice, customers, marketing folks and designers always employ different sets of context to express their understanding of affect information. Differences in semantics and terminology impair the coherence of transferring affective needs effectively from customers to designers. Furthermore, the sender-receiver problem which may arise during the communication process between customers and designers is a further reason leading to the misconception of customer affective needs.

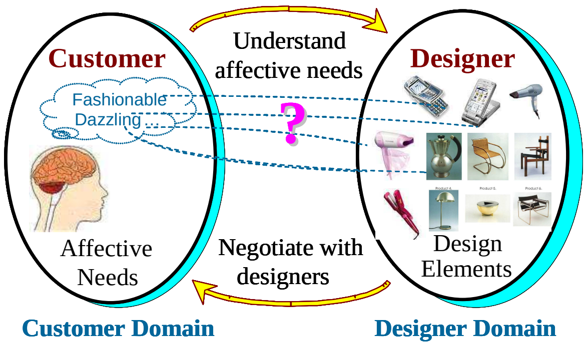


Fig. 5. The domain mapping process underlying affective product development

Inherently, affective human factors are subjective and mostly perception-based. To support affective design, these affective needs must be able to be manipulated during product design, that is, quantifiably in analytical forms. To rectify such deficiency in traditional human factors and ergonomics methods, it is necessary to adopt well established techniques of customer modeling and analysis from the business field, such as marketing research and product positioning. More specifically, discrete choice models will be applied to analyze customer behaviors and interactions with the product [8]. To alleviate difficulties in measure affect, utility theories [9] and conjoint analysis [10] are introduced to quantify customer-perceived value of affective design.

Another important area of understanding customers is user innovation, which has received limited attention in the current practice of affective design. User innovation refers to innovations developed by consumers and end users, rather than manufacturers. Von Hippel has discovered that most products and services are actually developed by users, who then give ideas to manufacturers [11]. This is because products are developed to meet the widest possible need. When individual users face problems that the majority of consumers do not, they have no choice but to develop their own modifications to existing products, or entirely new products, to solve their issues. Often, user innovations will share their ideas with manufacturers in hopes of having them produce the product – a process called free revealing. Built upon the popular lead user method, this project will further develop business and service models, incorporating with human-product-ambience interactions.

3.5 Application Scenarios

While affective design generally refers to the broad affective aspects of human factors, this research targets solid research programs by scrutinizing a concrete problem context. Considering the emerging trend of industries moving to mass customization and personalization, this research addresses the important situation of product customization and user personalization [18]. Consumer electronics products are deemed to be most indicative for this trend, and thus become one focus of this study. Vehicle design has many ingredients of customer affective needs. As a useful startup, this research investigates the interior design of the living compartment of a truck cab. Other commercial applications may be manifested through, for example, cabin comfortableness of passengers/patrons and attendants, such as an airplane, train, yacht, subway, exhibition hall, shopping mall, and alike.

3.5.1. iHome

An example of application scenario is developed by Philips Homelab, namely iHome, in which 'i' refers to the 'my' factor and an 'intelligent' sense of customization involving both the product and the ambience (http://www.research.philips.com/technologies/misc/homelab/). As shown in Fig. 6, examples of iHome may be the affective design of home appliances within a kitchen, living room, or restroom. Each of these scenarios can be constructed as virtual reality-enabled use cases in an ambient intelligence environment, through which various aspects of affective design can be explored. Within an iHome ambient intelligence environment, an affective design application is enacted as a coherent service system, including hardware and software, along with physical, media, or digital products. Various system-wide solutions can be further examined with respect to many business and service concerns.

3.5.2. Truck Cab

The interior design of truck cockpits involves both the product and the ambience through human interactions. In an effort to elicited customer affective needs, a mixed-reality environment has been installed with surveillance systems embedded (Fig. 7). The mixed-reality environment is used to simulate and rapidly reconstruct the actual truck cab environment at low cost. The surveillance system is used to capture customer response without interrupting the customer experience.

Fig. 6. Affective design with ambient intelligence application scenario: iHome

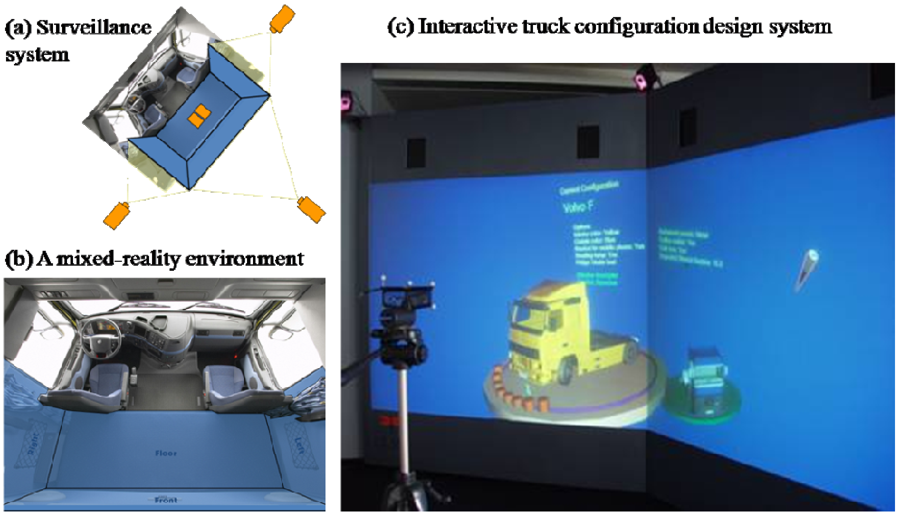


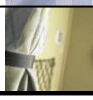
Fig. 7. Ambient intelligence environment of truck cab

Based on the system, a number of affective descriptors are elicited and categorized, forming the vehicle affective needs ontology, an excerpt of which is shown in Table 1. The elements for constructing the truck cab ecosystem are defined as customizable design elements to be included in different versions of the actual products. Typical design elements are listed in Table 2. These results will be used in the ongoing project for defining the mapping relationship between customers' affective needs and design elements, as well as product configuration design considering producer production capabilities.

Table 1. Definition of affective needs of truck cab

Affective needs	Definition	Affective needs	Definition
Durable	durable long-lasting hard-wearing …	Soft	soft silky tender …
Solid	solid firm hard …	Spacious	ample wide not-cramped …

Table 2. Design elements for truck cab ecosystem

Code	Description	Figure	Code	Description	Figure
Y1	Interior color_blue		Y6	Panel for switches	
Y2	Interior color_yellow		Y7	Bracket for mobile phone	
Y3	Interior color_grey		Y8	Reading lamp	
Y4	Curtain color_blue		Y9	Fridge under bed	
Y5	Curtain color_yellow		Y10	Fridge in rear shelf	

4 Concluding Remark

The proposed framework approaches affective design from an interdisciplinary perspective. To overcome the limitations of existing methods, the affective design with ambient intelligence framework orchestras relevant solutions from such fields as human factors and ergonomics, information and communication technologies, product innovation, industrial design, as well as business and management. Affective human factors have been traditionally dealt with in human factors and ergonomics with focus on the human-product interactions only. This research identifies an important dimension, namely interactions with the ambience, which has received very limited attention. In this regard, human-product-ambience interactions become the main theme of affective design, and in turn the main challenge is how to bring the environment dimension into the affective design horizon. Realizing the unique strength and potential of ubiquitous and pervasive computing, this research proposes to incorporate ambience intelligence into the affective design process, and thus enables the modeling of product ecosystems. Therefore, the proposed framework enriches the affective design problem context itself.

Main technical challenges are associated with the need of an ambient intelligence environment to support affective design. A few typical features to be realized in such an environment include:

Embedded: Since many devices are plugged into the network, the resulting system consists of multiple devices, computing equipment, and software systems that must interact with one another. Some of the devices are simple sensors, while others may be actuators owning a crunch of control activities within an ambient intelligence environment (e.g., central heating, security systems, lightning systems, washing machines, refrigerators, etc.). The strong heterogeneity makes difficult a uniformed policy-based management among diverse user interactions and services.

Context aware: A fundamental role of ambient intelligence is the capability of context sensing. This central concept of context awareness represents the possibility for the ambient intelligence system of biasing itself and its reactions to the environment. This means knowledge of many statically- and dynamically-changing parameters in relation to consciousness. In particular, affective design involves intensive user-centered contextual data, which necessitates the exploitation of relationships between the human concept of consciousness and the ambient intelligence idea of context.

Personalized: An ambient intelligence environment is supposed to be designed for people instead of generic users. This means that the system should be flexible enough to tailor itself to meet individual human needs. This is because affective design always involves highly customized products and personalized environments.

Adaptive: The affective design with ambient intelligence system, being sensible to the user's feedback, is capable to modify the corresponding actions have been or will be performed. This is consistent with the mass customization situation, where customers always want to make informed decisions of their own.

References

1. Jordan, P.W.: Designing Pleasurable Products: An Introduction to the New Human Factors. Taylor and Francis, London (2000)
2. MacQueen, M.: Ecosystem Thinking Needed for Product Design, http://macqueen.com/
3. Ducatel, K., Bogdanowicz, M., Scapolo, F., Leijten, J., Burgelman, J-C.: Scenarios for Ambient Intelligence in 2010. ISTAG Report, European Commission (2001)
4. Beyer, H., Holtzblatt, K.: Contextual Design: Defining Customer-Centered Systems. Morgan Kaufmann, San Francisco (1998)
5. Bittner, K., Spence, I.: Use Case Modeling, Addison Wesley Professional (2002)
6. Virzi, R.A.: Refining the Test Phase of Usability Evaluation: How Many Subjects is Enough?. Human Factors, 34(4), 457--468 (1992)
7. Nagamachi, M.: Introduction of Kansei Engineering. Japan Standard Association, Tokyo (1996)
8. Train, K.: Discrete Choice Methods with Simulation. Cambridge University Press (2003)
9. Edwards, W.: Utility Theories: Measurements and Applications. Kluwer Academic Publishers (1992)

10. Green, P., Carroll, J., Goldberg, S.: A General Approach to Product Design Optimization via Conjoint Analysis. Journal of Marketing. 43, 17--35 (1981)
11. Von Hippel, E.: Democratizing Innovation. MIT Press, Cambridge (2005)
12. Demirbilek, O., Sener, B.: Product Design, Semantics and Emotional Response. Ergonomics. 46(13/14),1346--1360 (2003).
13. Helander, M.G., Tham, M.P.: Hedonomics-Affective Human Factors Design. Ergonomics. 46(13/14), 1269--1272 (2003)
14. Karlsson, B., Aronsson, N., Svensson, K., Using Semantic Environment Description as a Tool to Evaluate Car Interiors. Ergonomics. 46(13/14), 1408--1422 (2003)
15. Helander, M.G., Khalid, H.M., Peng, H.: Citarasa Engineering for CATER System, In: Proc. Int. Conf. on Work with Computer Systems, Stockholm, Sweden, 21–24 May (2007)
16. Khalid, H.M, Helander, M.G.: Customer Emotional Needs in Product Design. Concurrent Engineering. 14(3), 197--206 (2006)
17. Jiao, J., Zhang, Y., Helander, M.: A Kansei Mining System for Affective Design. Expert Systems with Applications. 30, 658--673 (2006)
18. Jiao, J., Ma, Q., Tseng, M.M.: Towards High Value-added Products and Services: Mass Customization and Beyond. Technovation. 23, 809--821 (2003)
19. Talbot, J.: New Interactions: the Role of ‘Affective Human Factors’ in Design. Ergonomics Australia. 19, 21--25 (2005)

Smart Home Technology for the Elderly: Perceptions of Multidisciplinary Stakeholders

Anne-mie Sponselee[1], Ben Schouten[1], Don Bouwhuis[2], and Charles Willems[3]

[1] Fontys University of Applied Sciences, Hulsterweg 2-6, P.O.Box 141, 5900 AC Venlo, the Netherlands. email: {a.sponselee, ben.schouten}@fontys.nl

[2] Eindhoven University of Technology, Den Dolech 2, 5612 AZ Eindhoven, the Netherlands. email: d.g.bouwhuis@tue.nl

[3] iRv/Vilans, Zandbergsweg 111, 6432 CC Hoensbroek, the Netherlands. email: ch.willems@irv.nl

Abstract. The "implementation" and use of smart home technology to lengthen independent living of non-instutionalized elderly have not always been flawless. The purpose of this study is to show that problems with smart home technology can be partially ascribed to differences in perception of the stakeholders involved. The perceptual worlds of caregivers, care receivers, and designers vary due to differences in background and experiences. To decrease the perceptual differences between the stakeholders, we propose an analysis of the expected and experienced effects of smart home technology for each group. For designers the effects will involve effective goals, caregivers are mainly interested in effects on workload and quality of care, while care receivers are influenced by usability effects. Making each stakeholder aware of the experienced and expected effects of the other stakeholders may broaden their perspectives and may lead to more successful implementations of smart home technology, and technology in general.

Keywords: smart home technology, perception, technology acceptance

1 Introduction

The most important developments in society for smart home technology are the socialization of care, extramuralization, and ageing [1]. Socialization of care means that people in need of care are no longer concentrated in large-scale institutions, but are returned a full-fledged place within society. Instead of concentrating on people's disabilities, one looks at a person's possibilities. Supporting aging adults to stay in their homes independently for a longer period of time concedes to the wishes and needs of many people in need of care, aiming for an improvement of quality of living and daily life. Extramuralization leads to less intramural residential facilities, remarkably more small-scale extramural facilities, but also to (re)new(ed) organization of services and an increased use of technological resources. Additionally, ageing plays an important role. The fact that the amount of elderly people is growing, and people become older as well, leads to an enormous increase of care demand. The demand for houses for the elderly, for care, and for services will therefore grow in the coming years, while a shortage in care personnel is expected. The use of smart home technology to support independent living is hereby inevitable [2].

2 Multidisciplinary Stakeholders

Introducing smart home technology in care settings involves more than a technological innovation. It comprises of new processes and organisational changes. These changes all have to occur within regulations and financial rules that are most likely not adjusted to the use of technology. The stakeholders involved in the process of implementing smart home technology in extramural care setting therefore consist of: designers, care receivers, caregivers, care institutions, service providers, housing corporations, insurance companies, and the government. The perceptual worlds of these stakeholders vary due to differences in background and experiences, which lead to different interpretations on how smart home technology can be helpful in supporting independent living of elderly people.

In this paper we focus on the perception of the caregiver, the care receiver, and the designer. Problems in the interaction between the user or end-user and the technology can partially be ascribed to the design, which indicates the importance of considering both the care receiver's perception and the designer's perception. The caregiver, however, is also expected to be a user of smart home technology, and as the care processes also have to be taken into account, we consider the caregiver as third stakeholder.

2.1 The Care Receiver

Care receivers, in our case elderly people, are often considered as technofobic. Although this perspective does not apply to all elderly people, and is remonstrated by several studies (e.g. [3,4]), we believe elderly care receivers are less keen on new technologies than young people are and show lower technology usage rates [5]. One of the reasons for this technofobic perspective might be "self preservation". Elderly people have to deal with more and more limitations due to the ageing process, which make them more vulnerable and dependent than before. As older adults have less experience with computers and other (new) technologies [5], they are confronted with their (cognitive) limitations when they have to work with it, which makes them more afraid of making mistakes. Their (computer) anxiety results in unutilised chances to live and function independently [5], or to be enabled and empowered by technological possibilities [6]. Elderly people may therefore seem peevish and conservative, as they do not want their current life to be influenced too much by external factors.

An older person appears to become less technofobic when he or she knows and understands the usefulness of the technology [7]. Unfortunately, older adults often do not realise what advantages technology can bring them [5, 6]. Proper guidance and information when new (smart home) technology is introduced and used is therefore important to let the older care receiver get a positive view on the technology, and realise the possible benefits of it.

2.2 The Caregiver

Although most caregivers are younger than the care receivers, caregivers are also known for their technophobia. This reluctance towards technology can partially be explained by the caring character of caregivers. People who choose for care delivery or nursing as a professional occupation often prefer working with people who need them. Giving away personal contact, through a technology intervention, raises an aversion. Even in situations where technology replaces physical presence by virtual presence, like telecare, the caregiver gets the feeling he or she has to renounce that which is experienced as "caring". Caregivers also experience a reduction in time spent with clients as a direct decrease in quality of care [8]. According to caregivers' perception, technological developments that are cost-cutting - and are developed for that purpose - result in a loss of quality of care. Raappana et al. [8] state that technology and care service are commonly not felt as being connected, which results in unwillingness by caregivers to use technology, and difficulties when new technologies are introduced. This perception may be due to a lack of abilities and skills among caregivers, which leads to feelings of incapability, with decreased work motivation and distress as a result. Fortunately, caregivers are willing to see utilisation of a safety system as part of their professional care skills - unlike the use of a personal computer as such - and describe technology as a positive change in the (quality of) work [8].

Even though the study by Raappana et al. shows that safety technology is viewed as useful, the implementation of smart home technology results in extra work for caregivers. The caregivers' unfamiliarity with the technology, the lack of skills among care substitutes, along with an increased number of false alarms, result in time-consuming efforts for caregivers. This indicates the importance of professional training to reduce both the expected and the experienced extra workload of the caregivers. Another concern of caregivers is that they expect elderly people to become even more lonely when technology is introduced into the care process. However, screen-to-screen contact may increase social contacts among elderly and between elderly and the community or their relatives. This emphasises the importance of proper guidance and training for caregivers when smart home technology is introduced. The use of (care)technology should actually become part of care education, in which both usage and implications of smart home technology are taught. Besides, by means of good orientation on useful technologies most negative effects that caregivers experience can be eliminated.

The idea that all care workers are reluctant towards using technology needs some nuances. In a study among care workers by Lyons et al. [9] administrators appear to judge computers much more positively than physicians and nurses do, not surprisingly as computers were first introduced in the administration work field. Physicians, however, declare to be unmotivated to learn how to use computers, while nurses feel insecure and perceive the computer as a barrier between themselves and their clients. Especially the differences between managers and the nursing staff is of importance, as decisions - on the use of technology - are mostly taken by managers, without much consultation with those people who have to work with it later.

2.3 The Designer

The designer, on the other hand, is nothing but technofobic. This, at the same time, is his weakest point. For a technician it is hard to imagine the perceptual world of a technofobic user. The focus of the designer is mainly on the functionality of the technology, achieving the effective goals. The advantage of this focus is that the designer is well-aware of the benefits that the technology can bring. However, due to an often experienced vocabulary difference between the designer and the users and end-users (e.g. [10]), the designer may not be able to convince the technofobic (end)users of the usefulness and the benefits that smart home technology can have.

3 Perceptual Differences on Prevention and Privacy between Caregivers and Care Receivers

Prior to the implementation of smart home technology in care processes, a decision is made on the proper technologies which the care receiver needs. These decisions, however, are often made by the managers, based on recommendations by technicians, and have resulted in choices that were not in accordance with the actual needs of the care receiver [11, 12]. The reasons for the perceptual differences are explained here.

As mentioned before, caregivers are known for their caring character. In their perception, the care receiver is most important, but this also means that no risks will be taken. This protective view, possibly also due to a need for controllability, may in some cases result in situation that are too safe, in which elderly people are insufficiently stimulated to undertake actions by themselves and thus stay independent. Elderly people who move to a care facility often show a great regression in their functioning and their abilities, due to the increased support in comparison with the home situation. One of the goals of the caregivers should therefore be: continuation and stimulation of the independence and autonomy of the care receiver. Becker [13] refers to this inevitable change as “making care humane”. He considers caregivers as the “suppliers of human luck”. The protective mentality of caregivers also results in the protection of privacy of the care receivers. Caregivers are well-aware of the fact that people who are in need of care always have to deal with a loss of privacy. Protecting the remaining part of their privacy is one of the main issues for caregivers. Several studies, however, have shown that people in need of care are willing to lose some privacy if they get more independence or quality of life in return [14, 15].

In the study by Kearns et al. [14] the perspectives of 6 focus groups, including elderly nursing home residents, volunteer caregivers, care staff, medical surgical

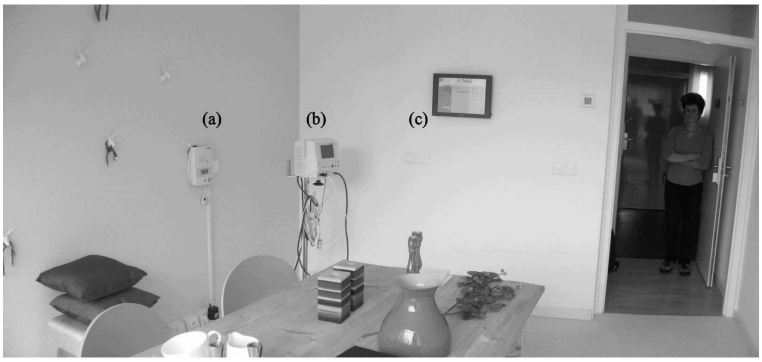


Fig. 1. Demonstration room with smart home technology: (a) Alarm unit; (b) Telemedicine monitoring system; (c) Touch screen with electronic patient file.

staff, and engineers, were combined to find requirements of elopement management systems. Although all focus groups agreed on the use of a non-stigmatizing device to attach to a wanderer (an inconspicuous device should, for instance, resemble a necklace or a watch), there was a different perception on the use of an implanted "tracking chip". The elderly focus group was less reluctant towards using an implanted chip than expected. When privacy and ethics were brought up by the researcher, it was quickly diminished as secondary. Apparently safety and independence are more important to elderly people than privacy. Kearns et al. refer to this as the liberating role of technology.

In a study by Willems [15] both elderly people and caregivers were asked to interpret the smart home technology available in a demonstration facility. One of the rooms of the demonstration facility is depicted in Fig. 1. The study showed that caregivers are more focussed on safety and security technology that may prevent harm and injuries, while elderly people are more eager to agree on care technology. Although the elderly subjects found an alarm unit useful for safety issues, they did not believe it would be necessary for them. This finding can be explained by the negative stigmatizing association that seems to overrule the positive safety effect of an alarm unit. On the other hand, elderly people are willing to hand in some of their privacy in order to facilitate the care giving process. Caregivers only agreed on the technology when they were sure the care receivers fully accepted the technology.

There is clearly a different interpretation between caregivers and care receivers in how smart home technology can be helpful supporting independent living of elderly people. While "implantables" seem accepted by the elderly, caregivers as well as researchers [16] believe that even more common technology applications likethe use of cameras for remote monitoring of people with mental disabilities is ethically not acceptable. Caregivers should realise that in situations where cameras are used to increase the safety and independence of the mentally disabled, care receivers will probably agree on the use of cameras despite the loss of privacy. The lack of concern about privacy might be ascribed to technological and social developments, such as cell phone networks, cameras in public spaces, blogs and home videos on the world wide web, and reality television shows like "Big Brother". Privacy is becoming a global good, and in some situations less relevant than safety.

Another explanation for the different perceptions between caregivers and care receivers, may be the fact that caregivers have certain habits that do not always correspond to the clients' needs [17]. As the care giving process often does not involve technology use, caregivers may react quite reluctant towards the implementation of smart home technology, and its accompanying procedures. This preference for care giving in the way people are accustomed to irrespective of care receivers' needs, does also apply to situations in which technology is used. Patient lifts, for instance, are often used in situations where clients actually do not need a lift yet [18]. Apparently caregivers accept those technologies they are accustomed to in care giving situations. These technologies, however, are often prescribed by protocols in which the labour conditions are set, which are rigidly applied to all situations (independent of client needs). Caregivers should better deviate from routines and make use of only those technologies, including new ones, that serve the needs of the care receiver. To implement smart home technology properly, caregivers have to become aware of their habits by analysing their perceptions and their way of acting.

Studies on technology perceptions of caregivers and care receivers are important to understand the acceptance of smart home technology, although these perceptions may change over time, due to experience. In the study by Willems [15] the answers were given in foresight: respondents were asked to reply on a situation which they were not yet familiar with. Answers may therefore be different than if respondents are living in a smart home, or have to work with the technology in a care setting. A passive alarm, for example, a basic functionality in many smart homes in the Dutch "Vitaal Grijs" program, appeared not to be as effective as expected before implementation [19]. On the basis of (negative) user experiences the functionality was disabled or removed in most houses. This indicates a difference between expected benefit and experienced benefit. But also questions like "willingness to pay for", as in the study by Willems [15], may result in responses different than can be expected on the basis of actual purchases and use. In case of an active alarm unit, elderly people are reluctant towards buying the technology, as initial costs are relatively high while benefits are unknown. After (effective) use people appear to judge this technology and its costs positively, as the usefulness becomes obvious (see [20] on the role of usefulness in technology acceptance).

4 Perceptual Differences on Requirements between Designers and Care Receivers

A design engineer of smart home technology for older care receivers should be able to understand the needs and wishes of the users. The designer, however, has to deal with a potentially technofobic user but also with an older user. The

Fig. 2. The room controller allows you to control lighting, room temperature, and television, for example. This room controller is negatively evaluated by both care receivers and caregivers, mainly due to poor legibility [15].

process of aging brings along many limitations or disabilities that are difficult to imagine for a non-limited and non-disabled designer. During a symposium [21] this gap was described as: "young males have to design technology for old females". Although the emphasis should not be so much on the gender difference, the age difference is truly a relevant factor [22]. As described earlier, aging often brings along changes in vision, hearing, attention, and memory. Additionally, physical disabilities due to rheumatoid arthritis or paralyses due to a Cardiovascular Accident (CVA) happen more and more often. Designing technology considering an average adolescent would not be very useful in this case, as an adolescent differs strongly from an elderly person on these physical factors. In case of sound-signals, for example, the designer must be aware of the fact that elderly people can not or hardly hear sounds of 2000Hz and above. Also, no robust actions should be needed for handling the devices, and no difficult procedures should be required. Buttons have to be larger, symbols or texts should be well-legible, and thus larger, due to decreased vision. An example of such designer/user gap is found in Fig. 2 that shows a room controller with poor usability. The design of the interface does not correspond with the abilities and expectations of the (elderly) user. The LCD screen, for example, is difficult to read, due to bad illumination and low contrasts. The use of both sides of the device as buttons is not in correspondence with intuitive use, which is important particularly for elderly users, as they have difficulties learning new skills. Depending on the limitations of the end-user, the design requirements should be altered, in favor of the user (see for example [23] on design principles for elderly). Although all design principles may be relevant when designing for the older care receiver, the consequences of their limitations for the design obviously depend on the intended functionality, and should therefore be considered separately for each technological design.

A solution to the difficulties elder care receivers experience when using (smart home) technology may be found in "inclusive design", "design for all", or "universal design". Designers of technology for the elderly have been requested for inclusive design by gerontologists for quite some time [24, 25]. Inclusive design implies that older and disabled people are part of the potential user groups in all product development processes. The design for older (and disabled) people, however, requires special attention for their needs and abilities. We may question whether designing for "all", including the elderly and the disabled, is useful and appealing to a young non-limited person. We believe it is more important that the designer of new technology takes into account those needs and wishes of the user he is designing for. This design process, however, should not only focus on the technological usability specifications, as Nielsen proposes in his user-centered design [26]. As in scenario-based models [27, 28], the technology should be viewed from different approaches. However, it should concern an iterative process in which not only the expectations people have of the technology and its interaction with their environment are taken into account, but also their eventual experiences with the technology. We expect best results when the design process involves all relevant stakeholders, at several stages of the process.

5 Perceptual Differences on Functionality between Designers and Caregivers and Relatives

The design engineer, or technician, clearly believes in the functionality of the technology. The other stakeholders, in most cases, rely on the designer's knowledge and promises. This may result, however, in expectations that are too high. The study by Raappana et al. [8], for example, shows that relatives and caring family members were satisfied with the technology, as they had the feeling that the safety of their relative was secured. One of the problems caregivers saw in the interaction between caring family members and the technology, is that they relied more on the technology that was actually possible. Relatives should be informed that the technologies cannot replace all health monitoring, while technicians should be honest about the (im)possibilities of technology [10].

Another barrier designers experience with caregivers is the so called *not invented here syndrome* [18,29]. The fact that the technology is not solely designed for care purposes, or that is designed for another care institution is often used as a reason not to accept the technology in the care professional's own organization. Care institutions however should better be open to knowledge of, and experience with technologies used in other places in order to learn from it and make better (smart home system) decisions.

6 Analyses of Multidisciplinary Stakeholders' Perceptions

To decrease the perceptual differences between the stakeholders, we propose an analysis of the expected and experienced effects (E-E Analysis) of smart home technology in care situations for each group. This means we are not only aiming at effects in relation to “effectiveness” - is the technology doing what it is supposed to do? - but also effects on the relationship between caregiver and care receiver, effects on the well-being of the client, on the nature of care giving, and matters like privacy, safety, security, and many more [18].

We are not only dealing with a gap between perceptions of various stakeholders, but also a difference between technological possibilities, related expectations, and the eventual use of the technology. The expectations and the actual use, including the subjective evaluations of the use, differ along the stakeholders and should be taken into account for successful implementation of smart home technology. This is why the analysis should include the expected and experienced effects of smart home technology of each stakeholder. The survey of these effects on all levels requires a multidisciplinary vision on this issue.

For the E-E analysis of effects the attribute-consequence-value *(A-C-V) model* [30] can be used, to get to higher and lower level effects. Attributes relate to aspects of the product or service, like functionality and design, on a very basic level. Consequences concern the functional and psychological effects of the technology, e.g. technology acceptance, while values resemble higher order merits, such as goals. The next step is to survey these attributes, consequences and values for each stakeholder involved. It is important to analyze the different layers of the technology, ranging from the functionality of the system to the behavior of people. While the designer may only be looking at the functionality, and whether or not the technology functions right, the user is interested in lower level effects, like the usability of the interface or the effect of the environment on the technology and vice versa.

To increase the acceptance and use of smart home technology, the technology should fit into the daily routines of users and end-users. The designer must be aware that his design determines how the technology intervenes with the orderliness of life-supporting everyday activities. The design of the technology may have an impact on timeliness, reliability, dependability, safety, and security [6]. Cheverst et al. propose a full user needs assessment, to analyze how the (end)user interacts with the technology from a psychological, emotional, physical, and social perspective. Also broader social and ethical effects of the technology should be identified and taken into account by the designer. As long as there are difficulties with the acceptance of smart home technology, the designer must consider an iterative design process [31], in which problem specification, matching the system to the real world, and the evaluation are an ongoing process [6]. The analysis of the effects should thus include a user-technology interaction assessment on a daily routine scale.

The E-E analysis thus displays possible mismatches between stakeholders as well as between expected and experienced effects. The implementation of a passive alarm, as mentioned earlier [19], is a good example of these differences between stakeholders and expectations and experiences. Care receivers expected great usefulness of a passive alarm, as it would give them feelings of safety and security. Their experiences after implementation, however, were feelings of insecurity and unreliability due to a high amount of false alarms. A false alarm is generated when elderly people forget to turn the switch in their house to indicate whether people are home or not. The - misplaced - expectation of the design engineer and the caregivers was that care receivers would be able to learn this new routine. By taking all of the effects into account in an iterative smart home design process, the design would better not contain a switch that needs action by an elderly user. A more valuable and less preferred solution that was chosen in the “Vitaal Grijs” project, however, was to disable or remove the technology [19].

The E-E analysis can not only be made by taken into account the expected effects on each stakeholder, but also by actually including the (end)users in the design process. Several studies have focussed on involving caregivers or elderly care receivers in the designing process [22]. In specially built user centers, user experiences can be tested beforehand, in the prototyping phase [32,33]. Another method occasionally applied is the use of drama [34]. These time-consuming processes, however, become less urgent when designers are aware of the perceptual world of caregivers and care receivers.

Analyzing the expected and experienced effects of smart home technology for each stakeholder involved, leads to better insight into human-technology interactions, which will result in better choices in the design process and system development. The possibility that the technology will not be accepted by the (end)users decreases, which will cut down expenses. At the end, the analysis may lead to the development of standardization in smart home technology. The downside of the analysis are the extra work and initial costs involved, although this will be compensated by the increase in technology acceptance. As the benefits of the investment are unclear until later, the return of investments appears negative at first. This is also the reason why care organisations are quit reluctant towards large-scale implementation of smart home technology. The initial costs of the technology and the organisational changes are relatively high, while the benefits (reduction in workload and costs) only become obvious after even more investments (increased workload). Additionally, we may also question whether elderly care receivers as well as caregivers actually know what is best for them. The latter implies that a multidisciplinary view, by combining all stakeholders' perceptions is crucial for effective smart home technology implementation.

7 Discussion

To decrease the perceptual differences between the stakeholders, we proposed an analysis of the expected and experienced effects of smart home technology (E-E Analysis) in care situations for each group. For designers the effects will involve effective goals, caregivers are mainly interested in effects on workload and quality of care, while care receivers are influenced by usability effects. It is not the case that technological possibilities are insufficient to solve the problems with smart home technology in care situations. Actually, on a technological level even more is possible than is yet applied in so called "smart" technology. Maybe the problem lays more or less in the functionalities of smart home technology that do not correspond to the actual needs of the care receivers or caregivers. Even though many researchers have stated that user requirements should be taken more into account in smart home projects, much technology development is driven by technological possibilities (technology push). The actual users obviously need to get involved in the development and implementation process of smart home technology. By involving the care receiver and the caregiver in the process, the designer may gain more insight into the true perceptions of the stakeholders he or she is designing for. As a result, the list of functional requirements for a smart home system or a smart home project must consist of more than just technological functionalities, and should comprise all stakeholders' attributes, consequences, and values. Finally, stakeholders should not only be aware of the expected effects, but also of the actual experienced effects, which may influence the list of requirements.

References

1. Provincie Utrecht: Wonen-Welzijn-Zorg: Behoefteontwikkeling en Aanbod in Gemeenten, 2003-2020 [Housing-Welfare-Care: Supply and Demand Development in Communities, 2003-2020]. Companen, Arnhem (2003)
2. Raad voor de Volksgezondheid en Zorg: Arbeidsmarkt en Zorgvraag [Labour Market and Care Demand]: Advies Uitgebracht door de Raad voor de Volksgezondheid en Zorg aan de Minister en Staatssecretaris van Volksgezondheid, Welzijn en Sport. RVZ, Raad voor de Volksgezondheid & Zorg, Den Haag (2006)
3. Rogers, W., Cabrera, E., Walker, N., Gilbert, D., Fisk, A.: A Survey of Automatic Teller Machine Usage across the Adult Life Span. Human Factors: The Journal of the Human Factors and Ergonomics Society, 38 (1996) 156-166
4. Dyck, J., Smithers, J.: Age Differences in Computer Anxiety: the Role of Computer Experience, Gender and Education. Journal of Educational Computing research, 10 (1994) 239-248
5. Czaja, S., Charness, N., Fisk, A., Hertzog, Ch., Nair, S., Rogers, W., Sharit, J.: Factors Predicting the Use of Technology: Finding from the Center for Research and Education on Aging and Technology Enhancement (CREATE). Psychology and Aging, 21 (2006) 333-352
6. Cheverst, K., Clarke, K., Dewsbury, G., Hemmings, T., Hughes, J., Rouncefield, M.: Design with Care: Technology, Disability and the Home. In: Harper, R. (ed.): Inside the Smart Home. Springer-Verlag, London (2003) 163-179
7. Eyck, A.: Domotica in Senioren Woningen. Zijn Ouderen Bereid deze Nieuwe Voorzieningen te Accepteren en Zijn ze ook in Staat om ze te Gebruiken? [Smart Home Technology in Adapted Homes. Are Elderly Prepared to Accept these New Facilities and are They Able to Use Them?] Maastricht University, Maastricht (2004)
8. Raappana, A., Rauma, M., Melkas, H.: Impact of safety alarm systems on care personnel. Gerontechnology 6 (2007) 112-117
9. Lyons, S., Tripp-Reimer, T., Sorofman, B., DeWitt, J., BootsMiller, B., Vaughn, T., Doebbeling, B.: VA QUERI Informatics Paper: Information Technology for Clinical Guideline Implementation: Perceptions of Multidisciplinary Stakeholders. Journal of the American Medical Informatics Association 12 (2005) 64-71
10. Nouws, H., Sanders, L.: Monitoring Toekomst Thuis. Eerste Tussenrapportage: Leerervaringen Voorbereidingsfase [Monitoring Future Home. The First Interim Report: Experiences of the Preparation Phase]. De Vijfde Dimensie, Amersfoort (2007)
11. Smart Homes: Evaluatieonderzoek Woonzorgcomplex "Rozenhof II" te Velden en Woonzorgcomplex "De Waag" te Arcen [Smart Homes: Evaluation Research Sheltered Housing "Rozenhof II" in Velden and Sheltered Housing "De Waag" in Arcen]. Smart Homes, Eersel (2004)
12. van de Leeuw, J.: Veilig en Comfortabel Wonen met Domotica. Beschrijving en Analyse van Vraaggestuurde Domoticaprojecten in de Ouderenhuisvesting [Safe and Comfortable Living with Smart Home Technology. Description and Analysis of Demand Driven Smart Home Projects in Housing for the Elderly]. IWZ, Utrecht (2005)
13. Becker, H.: Hans Becker, Bestuursvoorzitter Humanitas: Leverancier van Geluk [Hans Becker, Chairman of Humanitas: Supplier of Luck]. Zorg + Welzijn Magazine 3 (2006)

14. Kearns, W., Rosenberg, D., West, L., Applegarth, S.: Attitudes and Expectations of Technologies to Manage Wandering Behavior in Persons with Dementia. Gerontechnology 6 (2007) 89-101
15. Willems, Ch.: Evaluatie van de Demonstratiefaciliteit Domotica [Evaluation of the Demonstration Facility Domotics]. Gerealiseerd door De Erkenkamp. iRv, Hoensbroek (2007)
16. van Berlo, A.: Ethics in Domotics. Gerontechnology 3 (2005) 170
17. Bussemaker, J.: Zorg voor Ouderen: Om de Kwaliteit van het Bestaan [Care for the Elderly: For the Quality of Life]. Ministerie van VWS (2007) 1-9
18. Soede, M., de Witte, L.: Inaugurale Redes. Uitgesproken bij de Installatie als Lectoren Technologie in de Zorg aan de Hogeschool Zuyd [Inaugural Speeches. Delivered at the Appointment as Lecturers Technology in Care at the Hogeschool Zuyd]. Technologie in de Zorg / Hogeschool Zuyd, Heerlen (2007)
19. Bosch, A., van Daal, P., Dorrestein, A.: Thuis met Domotica. De Ervaringen van Ouderen in Zes Brabantse Domoticaprojecten. [Home with Domotics. The Experiences of Elderly People in Six Regional Projects] PON, Tilburg (2002)
20. Melenhorst, A.-S.: Adopting Communication Technology in Later Life. The Decisive Role of Benefits. J.F. Schouten School for User-System Interaction Research, Technische Universiteit Eindhoven, Eindhoven (2002)
21. Verweij, R.: Symposium ter Gelegenheid van de Inaugurele Rede van [Symposium on the Occasion of the Inaugural Speech of] dr. L. de Witte en dr. M. Soede, dd. June 21, 2007. Hogeschool Zuyd, Maastricht (2007)
22. Hawthorn, D.: Enhancing the Contributions of Older People in Interface Design. Gerontechnology 5 (2006) 4-15
23. Fisk, A., Rogers, W., Charness, N., Czaja, S., Sharit, J.: Designing for Older Adults: Principles and Creative Human Factors Approaches. CRC Press, Boca Raton, USA (2004)
24. Newell, A.: Older People as a Focus for Inclusive Design." Gerontechnology 4 (2006) 190-199
25. Whitney, G., Keith, S.: Active Ageing through Universal Design. Gerontechnology 5 (2006) 125-128
26. Nielsen, J.: Usability Engineering. Morgan Kaufmann, San Diego (1993)
27. Rosson, M., Carroll, J.: Scenario-based Design. In: Jacko, J., Sears, A. (eds.): The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications. Lawrence Erlbaum Associates, Inc., Mahwah, NJ (2002) 1032-1050
28. Marzano, S.: The Culture of Ambient Intelligence. In: Aarts, E., Encarnao, J. (eds.): True Visions. The Emergence of Ambient Intelligence. Springer, Berlin Heidelberg (2006) 35-52
29. Kittz, It Cares, I&C Services Nederland: ICT in de Care 2004-2006. Ministerie van VWS, Woerden/Groningen (2004) 1-26
30. Chiu, C.-M.: Applying Means-End Chain Theory to Eliciting System Requirements and Understanding Users Perceptual Orientations. Information & Management 42 (2005) 455-468
31. Clarkson, P., Keates, S.: A Practical Inclusive Design Approach. Proceedings of INCLUDE 2001, London (2001) 72-73
32. Dewsbury, G., Dickinson, A.: Enabling and Applying Person-Centred Design for Older Adults. Gerontechnology 5 (2006) 56-57
33. de Ruyter, B., Aarts, E., Markopoulos, P., IJsselsteijn, W.: Ambient Intelligence Research in HomeLab: Engineering the User Experience. In: Weber, W., Rabaey, J., Aarts, E. (eds.): Ambient Intelligence. Springer, Berlin Heidelberg (2005) 49-62

34. McKenna, S., Marquis-Faulkes, F., Newell, A., Gregor, P.: Requirements Gathering Using Drama for Computer Vision-Based Monitoring in Supportive Home Environments. Gerontechnology 5 (2006) 29-45

Enhancing Human Understanding through Intelligent Explanations

Tina Mioch[1,2], Maaike Harbers[1,2], Willem A. van Doesburg[2], and Karel van den Bosch[2]

[1] Institute of Information and Computing Sciences, Utrecht University, P.O.Box 80.089, 3508 TB Utrecht, The Netherlands
[2] TNO Defence, Security & Safety, Kampweg 5, 3796 DE Soesterberg, The Netherlands
{tina.mioch,maaike.harbers, willem.vandoesburg,karel.vandenbosch}@tno.nl
http://www.tno.nl

Abstract. Ambient systems that explain their actions promote the user's understanding as they give the user more insight in the effects of their behavior on the environment. In order to provide individualized intelligent explanations, we need not only to evaluate a user's observable behavior, but we also need to make sense of the underlying beliefs, intentions and strategies. In this paper we argue for the need of intelligent explanations, identify the requirements of such explanations, propose a method to achieve generation of intelligent explanations, and report on a prototype in the training of naval situation assessment and decision making. We discuss the implications of intelligent explanations in training and set the agenda for future research.

Key words: Explanations, simulation-based training, intelligent tutoring systems, cognitive modeling, feedback, learning

1 Introduction

Human well-being and performance are highly affected by the environment in which a person operates. People are always trying to improve their conditions, from increasing the temperature when it is cold to developing more and more advanced computer systems to aid them in their daily work. A recent development in the enhancement of environments is the incorporation of mechanisms that show some understanding of humans. Such mechanisms use sensors to acquire information about human functioning and analyze this information to adapt to human needs. An environment containing systems with these mechanisms is called intelligent.

Some of the applications exposing such ambient intelligence require interaction between the human user and the system. For example, decision-support systems have to communicate their advice to the person who is in charge of making a decision, and tutoring agents need to convey instructions and feedback to a student. Human-system interaction has two sides: the system or agent has to transmit information to the human, but it also has to understand the human. An agent reminding elderly or disabled people to take their medicine not only has to convey this message, it must also be able to understand when someone says he or she has already taken the medicine. One of the requirements of good human-system interaction is that the human understands and accepts a system's message. The quality of interaction between the human and the system is an important factor in the endeavor to improve human comfort and performance.

We claim that one of the factors contributing to the quality of human-system interaction is intelligent explanation. Providing explanations along with presented information is not something new. Various explanation components have been developed in recent decades for software systems, such as intelligent tutoring systems, decision-support systems and expert systems [8, 4, 9, 12]. It is supposed that the more these explanations are tailored to the specific needs of the user, the better the user is served. A system could make distinctions between users on the basis of their knowledge, speed of learning, most efficient learning method, preferences, etc. Most of the existing explaining components do not take features of the specific user into account, but treat all users in the same way.

In this paper we will clarify that in order to improve the effectiveness of explanations, systems should be equipped with capacities that refer to the users' mistakes, performance, beliefs, knowledge, intents and the like in their explanations. First we will take a closer look at agent explanation in ambient systems: what are the requirements to make an explanation useful, and what type of explanations can be distinguished? Then we will discuss how such an explanation mechanism can be implemented in a feedback system of a simulation-based training environment.

2 Intelligent Explanations

Most people take the information on a digital clock for granted; there is no need for further explanation about the current time. However, a user is not always sufficiently informed by such basic information. Even though a system may be correct in stating 'It is time to buy a new computer', this announcement might raise some questions. He or she wants to know why the computer believes this; is the computer too outdated for its purpose, or is the computer broken? If so, it would be interesting to know what part is not functioning and whether there are possibilities to update or repair the computer. This example shows that in some cases it is not sufficient if a system just presents its conclusion. An accompanying useful explanation will make more sense to the human user and it will increase both the human's understanding and acceptance of the system [20, 21].

Explanations exist in many forms. Furthermore, one single event can be explained in different ways. One explanation is not by definition better then another; the desired explanation depends on the context in which it is given. For example, a possible answer to the question 'Why did the apple fall?' is 'Because I dropped it'. In some situations however, the explanations 'Because I stumbled' or ‘Because he pushed me’ would be more useful. A whole other type of explanation of why the apple fell is ‘Because of the gravitation force’. Dependent on the context, people need a particular type of explanation. An explanation system should be able to estimate the information need of the user and provide an explanation accordingly.

Another difficulty to overcome in providing explanations is *timing* [20]. In some applications it is obvious that each time new information is presented it should be accompanied by an explanation. For instance in diagnosis systems, every given diagnosis should be accompanied by an explanation of how the system came to this result. In contrast, in systems that constantly provide new information, there are no predefined moments in which explanations should be given. For instance, a navigation system has to decide for itself when the user needs new instructions. So in a complex and open environment, an explanation system should be able to determine when and how often the user needs explanations.

Furthermore it is desirable that explanations are adapted to the receiver, as not all people are the same and thus might need different explanations. Whereas novices tend to need extensive explanations, experts generally prefer explanations in which the to them obvious steps are skipped [17]. Besides level of expertise, other human factors such as knowledge, intents and emotions could be taken into account. An explanation commenting on an assumed strategy of a student could be: ‘Because you performed action a_1, I think your plan must be P. This is not a good strategy because you do not have enough resources to perform action a_3, which is also part of plan P’. An explanation involving emotions is: ‘The other agents acted this way, because your angry words scared them’. Hence, intelligent explanations should be adapted to the user’s perspective to enhance understanding and learning.

3 Related Work

In the past twenty years, much research has been done on intelligent tutoring systems (ITS) [14, 13, 4], which are systems that teach students how to solve a problem or execute a task by giving explanations. Such systems have been successfully designed for the training of well-structured skills and tasks (e.g. LISP programming [11] or algebra [10]), which are relatively closed, involve little indeterminacy, and do not involve real-time planning. For the training of real world tasks, these conditions do not always apply [1]. Real world tasks are often complex, dynamic and open in the sense that outcomes of actions may be unpredictable. These features make it difficult to design training, because it is usually not possible to represent the domain by a small number of rules. Moreover, the space of possible actions is large. For instance, the military uses simulation-based training systems to train tactical command and control [2]. In such training, the student responds in real-time to simulated problems, so the system needs to be able to evaluate whether the actions taken are correct and whether they have been executed at the right time. A complicating characteristic of evaluating tactical performance is that there is often no single ‘right’ way to accomplish a task, but that there exists more than one good solution for a problem, depending on the context [3]. In addition, a training system should not only evaluate a student's behavior, but, in case of errors, it should also take cognitive processes underlying that behavior into account. The result should be suitable for inferring the student's strategy. The demands on context-sensitivity and performance diagnosis make it hard to generate appropriate feedback.

In recent years, the challenge of developing and providing explanations in open, complex and dynamic environments has been accepted by the international research community [7, 15, 5] and first steps forward have been made. Livak, Heffernan and Moyer [7] developed a prototype that uses a cognitive model to provide both tutoring and computer generated forces capabilities. The actions of the student are evaluated by comparing the student's behavior to the ideal behavior of an expert. If the student deviates from the behavior that the expert model demonstrates, feedback is returned to the student. The feedback that is given is a low-level explanation of why the particular action at that moment was not correct. The explanations only refer to a particular action, and no reference to consequences of actions are given. Furthermore, the tutoring agent does not maintain a model of the user to take his beliefs and intentions into account. As a consequence, the tutoring agent is not capable of adjusting its feedback to the specific knowledge and intentions of the student.

Other research focuses on the debriefing phase of the training by letting the simulation entities explain their reasons for executing particular actions to the student. Examples of this approach are the explanation system *Debrief* [8] and the *XAI system* [9]. Debrief is used to generate explanations for the individual agent's actions in the debriefing phase of the simulation, together with information about what factors were critical for taking that action. The XAI System allows the student to further investigate what happened during the exercise. In order to generate explanations, the software agents log important actions annotated with abstract information about underlying reasons of the actions as well as their consequences. Both Debrief and XAI explain the reasoning behind the executed actions on demand, expecting the student to ask the relevant questions. No assessment of the student is made, so no directed feedback can be given to the student. In addition, the explanations are directly related to knowledge about the task, giving a low-level reason for a particular action. For example, for the task of clearing a room, an agent might answer the question 'Why did you throw a grenade into a room?' by stating that 'A grenade suppresses enemies that are in the room'. It would be more informative for the student to give an explanation on a higher conceptual level, including e.g. beliefs and intentions of the agent. Such an explanation would for example be 'I believed that the enemy was in the room. My goal is to clear all rooms. By throwing grenades into the room, I intended to suppress the enemy'.

As can be seen, research on explanations has been recognized as being important in training simulations to enhance the student's learning experience. However, a lot of research is still required. Questions that still need answering are for example how to obtain insight into the cognitive processes of the student, and how to support students in acquiring an understanding of the relationships between their behavior and the consequences in the environment. To achieve understanding, explanations must be given about processes in the environment. However, as stated above, explanations in simulation-based training systems are often not profound enough to achieve this result.

4 Types of Feedback

When building simulation-based training systems, three types of feedback can be distinguished. They differ in the types of information that they take into account and the sophistication of explanations they give:

Result-based feedback: Feedback is based only on observable student behavior. Correct results, formulated by domain experts, are hard-coded into the scenario, and feedback is generated by comparing the student behavior with the correct behavior. The feedback states only whether the student has completed the task successfully, and if not, what the correct behavior should have been.

Model-based feedback: Feedback is not only based on explicit student behavior, but also on contextual knowledge of the simulation environment and explicit task knowledge. Using the different kinds of information, the feedback is generated by reasoning about an internal model.

Cognition-based feedback: As with model-based feedback, feedback is based on explicit student behavior, knowledge about the simulation environment and task knowledge. In addition, a user model is developed that makes it possible to infer cognitive strategies of the student to facilitate even better feedback.

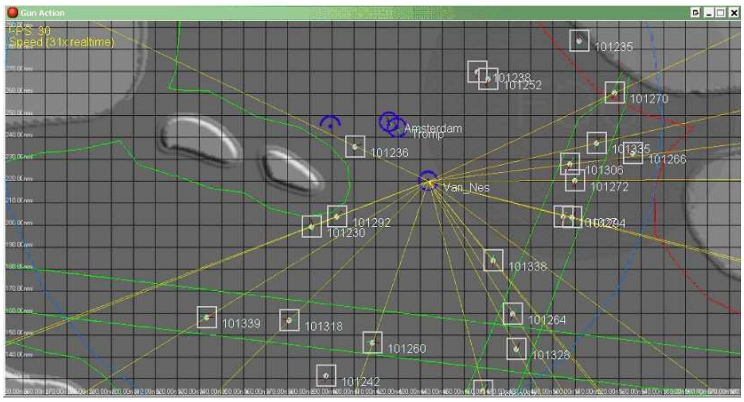


Fig. 1. Screenshot of our training scenario. The squares indicate radar tracks, the circles represent tracks of own forces.

We will illustrate how the types of feedback differ from each other in the context of a navy task, namely the tactical picture compilation task. Developing a tactical picture of the environment is an essential part of any military mission. In the tactical picture compilation task, the tracks in a radar picture have to be classified into categories such as the type of vessel, and the probable intention of the vessels has to be determined. An illustration of a situation in which a student has to develop a tactical picture can be seen in figure 1. Assuming that the student has to decide at a particular point in time which track poses the largest threat to the ship, the following examples of explanations illustrate the types of feedback that the system might give.

Result-based explanation *Your answer is incorrect: You have chosen track 101304 as the most threatening track. It should have been track 101112.*

Model-based explanation *The expert disagrees with your answer: You have chosen track 101304 as the most threatening track. However, its speed is not very high. Additionally, there are a number of ships that are moving at the same speed AND are closer to you. The expert thinks the most threatening track is 101112.*

Cognition-based explanation *The expert disagrees with your answer: You have chosen track 101304 as the most threatening track. However, its speed is not very high. Additionally, there are a number of ships that are moving at the same speed AND are closer to you. The expert considers 101112 to be the most threatening track.*

You have assessed tracks coming from the harbors, probably because you might suspect the enemy to reside in the harbor. However, a good strategy is to start investigating close to your vessel and progress outward. We have not seen you do this in the scenarios you have played thus far. This tip should help you protect your ship by preventing enemy vessels to get too close unseen.

These examples illustrate that generating model-based explanations are the minimum requirement for *intelligent* explanations. Cognition-based explanations are the most sophisticated and would be the type of explanations from which a student learns the most. For that reason, our goal is to build a training system that generates cognition-based explanations.

5 Intelligent Tutoring Agent for the Royal Netherlands Navy

For the Royal Netherlands Navy we investigate the possibility of developing an agent that fulfills the task of an instructor in a training simulation. We focus on the functionality of evaluating student performance and deliver this evaluation along with an explanation. The task that is chosen is a modification of the tactical picture compilation task as described in section 4. In the modified task, the student is presented with a radar picture at a particular point in time, showing a number of radar tracks. The student has to gather and integrate information on these tracks to form a mental tactical picture of the situation. Then the student has to decide which track poses the largest threat to his own ship. Time does not play a role as the picture is static and represents a situation at a particular point in time. Factors in this task that have to be taken into account are for example the speed of vessels, distance from own ship, whether they adhere to shipping lanes and whether they are inbound.

We are developing a training system that uses cognition-based explanations. To meet this objective, the following method is introduced. To generate feedback, an expert agent is executed, as are agents that deviate in some aspect from the expert, representing typical mistakes of students. These deficient agents intentionally fail to take one or more particular factors of the task environment into account, or are deficient in another way. It is assumed that errors of the student are the result of incorrect beliefs or an incorrect strategy. The assessment the student makes of the situation after examining the screenshot is compared to the assessment of the expert and the deficient agents. Four outcomes of the comparison can be differentiated.

First, the assessment of the student might not correspond to either the expert's assessment or to any of the deficient agents' assessments. In that case, the student did not complete the task satisfactorily nor did he make a typical mistake represented by any of the deficient agents. An explanation is then generated that explains why the assessment of the expert is preferable to the assessment of the student based on the comparison between the two performances and on task knowledge. This includes for example knowledge about the environment and knowledge about the importance of different relevant factors.

Second, the assessment of the student corresponds to the assessment of the expert agent, and to none of the assessments of the deficient agents. In this case, it is assumed that the student has solved the task satisfactorily, and accordingly, positive feedback is returned. As the environment is open, it is of course possible that the student's assessment to the problem is correct, but that the student has just been lucky, without the assessment being based on correct beliefs and a correct process of obtaining the assessment. However, over several trials of the training simulation, the incorrect strategy of the student will eventually fail and the student will then receive a feedback that shows that his beliefs are wrong.

Third, the student's assessment might correspond to the assessment of one particular deficient agent, without matching the expert agent or any other deficient agent. As it is assumed that the deficiency of the agent corresponds to the beliefs or strategies of the student, a diagnosis of the student's state of mind can be made and an explanation be generated. The explanation that is returned corresponds to the deficiency of the agent.

Fourth, it is possible that the student's assessment corresponds to several assessments, either of several deficient agents, or one or more deficient agents and the expert agent. This is possible because there are often many possible ways to arrive at the same assessment. Then, the response alone is not sufficient for deciding what feedback is appropriate. We need information about the processes that resulted in the selection of that response and which beliefs and strategies the student used to obtain his response. If we can do this validly, then we can return feedback containing an appropriate explanation. In this case, the user model is of importance, because it gives extra background information about the process that led to the assessment. On the basis of the inferred beliefs and strategies of the student, it is possible to choose the most corresponding of the matching assessments and return the appropriate explanation.

Our prototype does not yet take performance over time into account. In reality, the history of the situation should be used in situation assessment. For example, an apparently non threatening radar track (taking only properties such as, speed, distance, bearing and adherence to shipping lanes into account) may in fact be highly suspect because it has recently varied its speed and has intermittently crossed the shipping lane. A student that is sensitive to this information correctly assesses this track as threatening. A system that cannot use such information may then, erroneously, 'correct' the student and thus fault the student for his judgment even though the student actually outperforms the expert model. Such problems are typical for evaluating performance in complex, dynamic and open tasks. To overcome these problems, it is more useful not to evaluate the assessments of a student but the cognitive strategies that have led to the assessment.

A problem is that cognitive strategies are not observable. We are faced with the problem to construct a user model containing hypotheses about the strategies of the student without the ability to observe these strategies. We choose to overcome this problem by arranging the training simulation in such a way that the user is forced to provide some information about his strategy. We achieve this by allowing the student access to all information that is available in the actual operational environment, but only on explicit request of the student. For example, by initially hiding the shipping lanes and allowing these to be seen for a short period of time we gain evidence that the student is checking adherence to shipping lanes. By observing the pattern of behavior while the student is executing the task, we can build a hypothesis of the strategy that the student employs. Moreover, we can test the hypothesis by selecting a subsequent scenario and predicting the steps that the student will take. If the hypothesis is confirmed, we can then confidently proceed in providing feedback on the strategy level rather than on the performance level. In addition, it enables us to select those scenarios that practice the particular aspect which the student finds difficult. Because the user model is taken into account, the feedback is based on the perceived process of decision-making of the student, which includes an interpretation of the student's actions. By giving an explanation that has relevance to the student's actions and underlying beliefs, acceptance and understanding of the feedback are endorsed.

6 Discussion

In this paper we argued for the importance of intelligent explanations in human-system interaction. We clarified why explanations should be user-specific and what aspects should be taken into account in order to achieve this. There are different ways to generate model-based or cognition-based feedback; we use a method of modeling the user, an expert agent and deficient agents. The behavior of the user is compared to that of the agents. We argue that the results of such comparisons in combination with the user model yield insights about the user which make it possible to provide explanations fit to the particular user.

As mentioned before, we are currently implementing our method of explanation generation in a training system for the Royal Netherlands Navy. Once the system is ready, we will evaluate whether the explanations generated by the proposed method will improve the users' performances. We aim to extend the method to other situations and to apply it to more complex versions of the task, for example involving a time component, and to other tasks than tactical picture compilation. Therefore the expert model, the user model and the deficient agents need to be adapted to the new demands. Despite these changes and modifications, the core mechanism of the approach remains the same. So if the explanations are satisfying in the simple case, we are confident that the system is able to generate intelligent explanations in more complex versions of the task and other tasks as well.

A system providing the desired intelligent explanations referring to knowledge, plans, intentions and the like will yield many advantages. First, good explanations will help the user in his or her learning process, because they will improve conceptual understanding [18]. Besides this advantage while learning the task, good explanations prolong the duration of an acquired skill [16]. A practical advantage is the reduction of training costs. When systems are able to generate human-like explanations, fewer instructors are needed to complete training and this will save costs. The usual weighing between costs and quality no longer has to be made. Finally, because students are no longer dependent on the presence of a trainer, they are more flexible and can train a task whenever they want.

In future research, it could be investigated how expert and deficient agents can be modeled more efficiently. Especially a practical way of modeling deficient agents is useful, because in complex tasks many of them are needed. For this, the behavior of real students can be used. Also the relation between different deficiencies could be examined: what behavior does a user with a combination of different deficiencies show and how is this reflected in the modeled deficient agents? Further, more attention could be paid to the presentation of explanations: which way of presenting leads to the highest performance? It could also be investigated for what type of tasks the intelligent explanations turn out to be the most useful.

Acknowledgements This research has been supported by the GATE project, funded by the Netherlands Organization for Scientific Research (NWO) and the Netherlands ICT Research and Innovation Authority (ICT Regie) and by the research programs "Cognitive Modelling" (V524) and "Integrated Training and Instruction" (V406), funded by the Netherlands defence organisation.

References

1. Zachary, W., Cannon-Bowers, J., Bilazarian, P., Krecker, D., Lardieri, P., Burns, J.: The Advanced Embedded Training System (AETS): An Intelligent Embedded Tutoring System for Tactical Team Training. International Journal of Artificial Intelligence in Education (1999) 257-277
2. Doesburg, W.A. van, Bosch, K., van den : Cognitive Model Supported Tactical Training Simulation. Proceedings of the Fourteenth Conference on Behavior Representation in Modeling and Simulation (BRIMS), Universal City, CA (2005) 313-319
3. Hutchins, S.G., Kemple, W.G., Porter, G.R., Sovereign, M.G.: Evaluating Human Performance in Command and Control Environments. Proceedings of the 1999 Command and Control Research and Technology Symposium, Newport, RI (1999) 50-52
4. Murray, T.: Authoring Intelligent Tutoring Systems: An Analysis of the State of the Art. International Journal of Artificial Intelligence in Education **10** (1999) 98-129
5. Ntuen, C.A., Balogun, O., Boyle, E., Turner, A.: Supporting Command and Control Training Functions in the Emergency Management Domain Using Cognitive Systems Engineering. Ergonomics **49** (2006) 1415-1436
6. Murray, W.R.: Intelligent Tutoring Systems for Commercial Games: The Virtual Combat Training Center Tutor and Simulation. AIIDE (2006) 66-71
7. Livak, T., Heffernan, N.T., Moyer, D.: Using Cognitive Models for Computer Generated Forces and Human Tutoring. 13th Annual Conference on Behavior Representation in Modeling and Simulation(BRIMS), Simulation Interoperability Standards Organization, Arlington, VA. (2004)
8. Lewis Johnson, W.: Agents that Learn to Explain Themselves. Proceedings of the Twelfth National Conference on Artifcial Intelligence, Seattle, Washington, United States, American Association for Artificial Intelligence **2** (1994) 1257-1263
9. Gomboc, D., Solomon, S., Core, M.G., Lane, H.C., van Lent, M.: Design Recommendations to Support Automated Explanation and Tutoring. Proceedings of the Fourteenth Conference on Behavior Representation in Modeling and Simulation (2005)
10. Koedinger, K.R., Anderson, J.R.: Illustrating Principled Design: The Early Evolution of a Cognitive Tutor for Algebra Symbolization. Interactive Learning Environments **5** (1998) 161-180
11. Anderson, J.R., Conrad, F.G., Corbett, A.T.: Skill Acquisition and the LISP Tutor. Cognitive Science **13** (1989) 467-506
12. Core, M.G., Traum, T., Lane, H.C., Swartout, W., Gratch, J., van Lent, M.: Teaching Negotiation Skills Through Practice and Reflection with Virtual Humans. Simulation **82**:11 (2006) 685-701
13. Psotka, J., Massey, D., Mutter, S. (eds.): Intelligent Tutoring Systems: Lessons Learned. Hillsdale, NJ: Lawrence Erlbaum Associates (1988)
14. Polson, M.C., Richardson, J.J.: Foundations of Intelligent Tutoring Systems. Lawrence Erlbaum Associates, Inc., Mahwah, NJ (1988)
15. Jensen, R., Nolan, M., Chen, D.Y.: Automatic Causal Explanation Analysis in Combined Arms Training AAR. Proceedings of the Interservice / Industry Training, Simulation, and Educational Conference (I/ITSEC) (2005)
16. Bosch, K. van den: Durable Competence in Procedural Tasks Through Appropriate Instruction and Training. In: Harris, D. (eds.): Engineering Psychology and Cognitive Ergonomics, Aldershot, Ashgate **3** (1999) 431-438
17. Ramberg, R.: Construing and Testing Explanations in a Complex Domain. Computers in Human Behavior **12**:1 (1996) 29-47

18. Teichert, M.A., Stacy, A.M.: Promoting Understanding of Chemical Bonding and Spontaneity through Student Explanation and Integration of Ideas. Journal of Research in Science Teaching **39**:6 (2002) 464-496
19. Morrison, P., Barlow M.: Child's Play? Coercing a COTS Game into a Military Experimentation Tool. Proceedings of SimTecT 2004, Canberra (2004) 72-77
20. Nakatsu, R.T.: Explanatory Power of Intelligent Systems: A Research Framework. Proceedings of the IFIP International Conference on Decision Support Systems, Prato, Italy (2004) 568-577
21. Herlocker, J.L.: Position Statement - Explanations in Recommender Systems. Proceedings of the CHI' 99 Workshop, Pittsburgh, USA (1999)

Towards Natural Interaction by Enabling Technologies : A Near Field Communication Approach

José Bravo[1], Ramón Hervás[1], Salvador W. Nava[2], Gabriel Chavira[2] and Carlos Sánchez[3]

[1] Castilla-La Mancha University, Paseo de la Universidad, 13071 Ciudad Real, Spain
{jose.bravo, ramon.hlucas}@uclm.es.
[2] Autonomous University of Tamaulipas, Tampico-Madero, México
{snava, gchavira}@uat.edu.mx
[3] CCNT: Calidad Concertada Nuevas Tecnologías,
28037 Madrid, Spain.
carlos.barba@ccnt-spain.com

Abstract. The recent NFC technology is not only valid for payment and ticketing, a new interaction form is also available. With the only action being to put a mobile phone close to another NFC device (reader, tag or mobile phone) it is possible to get a contactless communication supported by a server or not. Some services such as open doors, location, access, presence, and so on, can be performed with a simple touch. In this work we analyze the interaction through the adaptability of two technologies: RFID and NFC. Challenges in models have been considered. In addition, we are trying to put in practice the idea of tagging context and the awareness only with a single interaction by touch. For that, we consider the benefits to adapting NFC technology as a good approach to model contexts.

Keywords: Ubiquitous Computing, Context Aware, RFID, NFC

Introduction

Ubiquitous Computing paradigm and, most recently, Ambient Intelligence (AmI), are the visions in which technology becomes invisible, embedded, present whenever we need it, enabled by simple interactions, attuned to all our senses and adaptive to users and contexts [1]. A further definition of AmI is as "a*n exciting new paradigm of information technol-*

ogy, in which people are empowered through a digital environment that is aware of their presence and context sensitive, adaptive and responsive to their needs, habits, gestures and emotions."

These visions promote the goals of: 1) embedding technology into everyday objects, 2) attaining effortless and closer interaction and finally 3) supporting a means for getting needed information anywhere and at anytime. This is our particular idea about Ubiquitous Computing, Natural Interfaces and Ubiquitous Communication. With them, the idea of creating intelligent environments requires unobtrusive hardware, wireless communications, massively distributed devices, natural interfaces and security. To attain this vision it is fundamental to analyze some definitions of the context. A. Dey defines this concept as "*any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and application themselves*." [2]. Also, this author defines context awareness as the "*context to provide relevant information and/or services to the user, where relevancy depends on the user's task*".

In order to design context aware applications it is necessary to observe certain types of context-aware information as being more relevant than others [3]. The user profile and situation, that is, the identity-awareness (IAw), are essential. The relative location of people is location-awareness (LAw). Time-awareness (TAw) is another main type of context-awareness that should be taken into account. The task which the user carries out and everything he wants to do is transformed into Activity-awareness (AAw). Finally, Objective-Awareness (OAw) looks at why the user wants to carry out a task in a certain place. All these types of awareness answer the five basic questions ("Who", "Where", "What", "When" and "Why") which provide the guidelines for context modeling. This kind of information allows us to adapt or build the needed technology to disperse throughout the environment and to model the human behavioral support.

Once the context and its important features are defined, it is time to study new interaction forms proposing the approach to the user by means of more natural interfaces. On this point, Albrecht Schmidt proposes the Implicit Human Computer Interaction (iHCI) concept [4][5]. He defines it as "*the interaction of a human with the environment and with artifacts, which is aimed to accomplish a goal. Within this process the system acquires implicit input from the user and may present implicit output to the user*". Schmidt defines implicit input as user perceptions interacting with the physical environment, allowing the system to anticipate the user by offering implicit outputs. In this sense the user can concentrate on the task, not on the tool as Ubiquitous Computing Paradigm proposes. As a next step this author defines Embedded Interaction in two terms: Embedding technologies into artifacts, devices and environments and embedding interactions in the user activities (task or actions) [6].

This paper addresses the identification process as an implicit and embedded input to the system, perceiving the user's identity, his profile and other kinds of dynamic data. Then, a new technology and the context adaptability, the Near Field Communication technology (NFC), are presented. With it, proposed changes of the architecture, the model, the visualization of information and the interaction are exposed.

Under the next heading, we present the identification technologies as inputs, RFID and NFC, both with their corresponding models. In the following section, a way of tagging context and interaction by touching can be seen. The paper finishes with related works and conclusions.

2 Identification Technologies

With the ideas of context and their mentioned characteristics, we have considered some awareness features through two different technologies. We propose these approaches by the identification process. It is a specialized input to by means of identification process. This means that we have the knowledge about the user profile, the context information and the task.

In the next section, we show RFID and NFC technologies with models that we propose while keeping in mind the aforementioned w's.

2.1 RFID (Radiofrequency Identification)

To create context-aware applications it is necessary to adapt sensorial capabilities to provide implicit inputs to the system in order to achieve natural interfaces closer to the users. With this the proactive aspect of the system is guaranteed.

In RFID systems there are basically two elements:

- Tags or transponders, which consist of a microchip that stores data, and an antenna (coupling element), that are packaged in such a way that they can be installed in an object. They also have a unique series number.
- Readers or interrogators have one or more antennas, which emit radio waves and receive signals back from the tag. The "interrogation" signal activates all the tags that are within its reach.

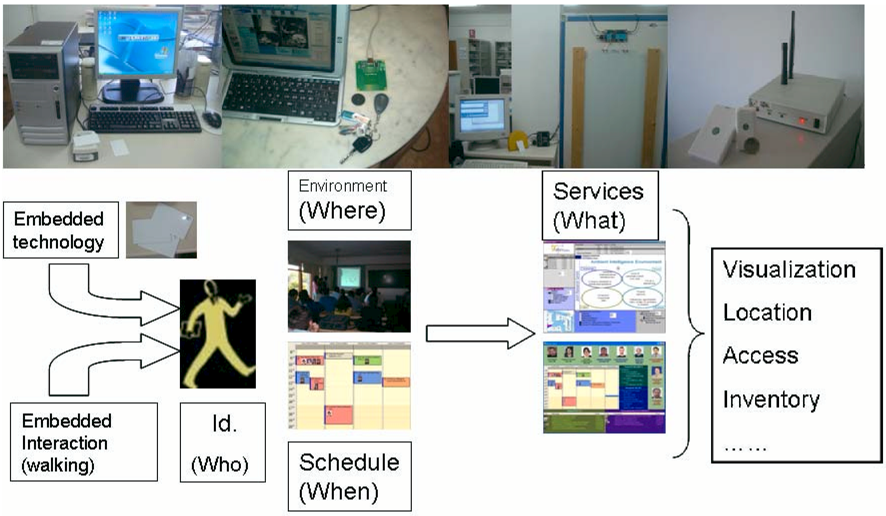


Fig. 1. RFID set and the "who" model.

RFID systems are classified as active and according to the kind of tag used. In this work we use only active tags: they are those that have their own power supply (battery) with a reach of up to 100 m.
Readers have a frequency of operation and are usually divided into three basic ranges: Low (125Khz.), High (13.56 MHz), which is the standard one, and Ultra High Frequency (UHF).
At the top of Figure 1, some types of RFID devices that we have placed in different contexts can be seen. The first one on the left presents a contact reader which is an antenna with a reach of only 10 cm. A model of the tag is also shown. With it, a project in a hospital context has been developed. The second one is another contact RFID set only for identification. We have developed access control for an office, classroom and research laboratory. These systems are ideal for individual use. The next one is a reader and an antenna with a read-and-write capability reaching over 75 cm. This has been specially designed to be placed on doors, or near displays. It can read several labels simultaneously, when identifying people entering the room. Finally, we use another kind of RFID set, offering more distance between reader and tags (up to 88 meters) working with UHF. The bottom of Figure 1 shows the model for this kind of sensorial specialized input. Embedding technology into daily objects and embedding interaction into daily actions, we perceive the context-awareness by the "who", "when" and

“where” aspect obtaining “what”. Typical services by identification are Access, Location, Inventory and so on, but we are interested in the Visualization service. With it, we try to offer information to the user in some parts of a mosaic. This way requires no user interaction according with the Ubiquitous Computing idea about not intrusion.

2.2 NFC (Near Field Communication)

It is obvious that we need a great variety of devices placed in the environment around us with wireless connection capabilities. Therefore, a new short-range wireless connectivity technology “Near Field Communications" (NFC), has appeared. Philips and Sony developed this technology in 2002. It is a combination of RFID and interconnection technologies.
NFC uses a high frequency band of up to 13.56 MHz, with a data transmission speed of 424 Kbits/s and a reach of 10 cm. It was deliberately designed to be compatible with the RFID tags operating in this band (ISO 14443), but incompatible with the standards of EPC global [10]. In 2002, international ECMA published the open standard 340 "NFC Interface and Protocol”, which was adopted by ISO/IEC with the number 18092 one year later.
NFC systems consist of two elements:

- The Initiator- as its name indicates it begins and controls the information exchange (called reader in RFID); and
- The Target-the device that responds to the requirement of the initiator (called tag in RFID).

In an NFC system, there are two modes of operation: Active and Passive. In the active one, both devices generate their own field of radio frequency to transmit data (peer to peer). In the passive one, only one of these devices generates the radiofrequency field, while the other is used to load modulation for data transfers.
It is important to mention that, although the NFC protocol can be installed in any electronic device, our interest will be centered on NFC-enabled cell phones.

In this case, the model changes absolutely. While in the RFID one, the model combines the “who”, “when” and “where” aspects, now there are some distinctions. The “when” aspect is not needed because is the user who decides when the action begins. In addition, another important feature is that users store the information needed to obtain services in their mobile phone. So, we can talk about a mode from a function of “who” and “what” (information) applying “where” to obtain “what” (services). This information in “what” is varied: presentation, note to comment, file to upload, etc.

3 Contextual Tags

When distributing context information throughout a building it is important to think about the specific place for every tag. A classification of everyone's position is needed to get services. So, the idea of context proposed by some authors has been considered when adapting input technologies.

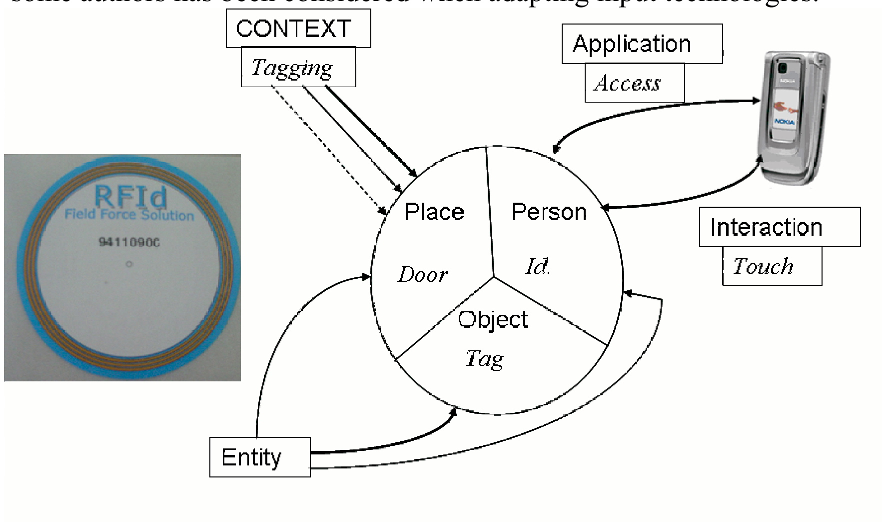


Fig. 2. Example of tagging context.

Keeping in mind the definition of context by Dey, we propose a tag structure. Every tag contains information according to this definition and the three important parts of the context: the object (tag and mobile phones), the user and the place. Figure 2 shows an example of tagging a particular place: the door. Here, the information is distributed labeling the exact place (MAmI Lab door), the objects are mobile phones and, finally, the user that is identified by mobile phones. In a second step, the concept of context awareness by tagged context is accomplished. We seek to translate the context awareness capabilities with only the interaction of two devices. In this sense, we convert the implicit interaction offered by the RFID technology and users wearing tags, to another kind of interaction by touch. It is obvious that the first one is more natural, however, there are some important advantages to using the NFC technology:

- The action dependability.- It is included in the user action by putting a mobile phone near a context tag. The change of the model is expected. The "when" aspect is not needed. The user decides when with a simple touch. This fact saves on awareness capabilities by including them into the explicit user interaction.

- The cost.- It is more expensive to place RFID antennas and readers.
- The readings/writing.- With NFC the process is mobile and more flexible. With RFID it is located in fixed points.
- Storage.- There is more capacity with mobile phones (over two gigabytes in SD cards).
- The server.- In NFC some actions can be managed by the mobile phone. With RFID the server is always needed.
- Privacy.- The individual information is on the mobile phone. This fact is an important requirement for privacy aspects.
- Security.- The distance of readings and writing are about two or three cm. This fact could be considered a disadvantage because the RFID distance is from a few centimeters to hundreds of meters. However, security is easier to control over short distances.

3.1 NFC Architecture

With NFC technology, the information of context is included in tags. Therefore, we can place a group of these around a building in some places: door, board and PC display. The architecture is shown in Figure 3. The mobile phone can interact with the information contained on every tag (about 1Kb). With it and the capabilities of the mobile phone, process, communication and storage, it is possible to offer a context-awareness in these places. Moreover, three more connections are present: phone and phone, phone and reader, tag and reader.

With NFC technology, some services can gather without server connection. For example, we can open a door to allow people to send the correspondent signal to the open door device by Bluetooth. In addition, other services are possible: location, note to comment and visualization, but all of them need a server to be processed. Finally, other services, such as storage in tag notices for people looking for you if you are absent from your office, are possible.

All the applications are managed by the JSR 257 in a 6131 Nokia phone and the JSR 82 for Bluetooth.

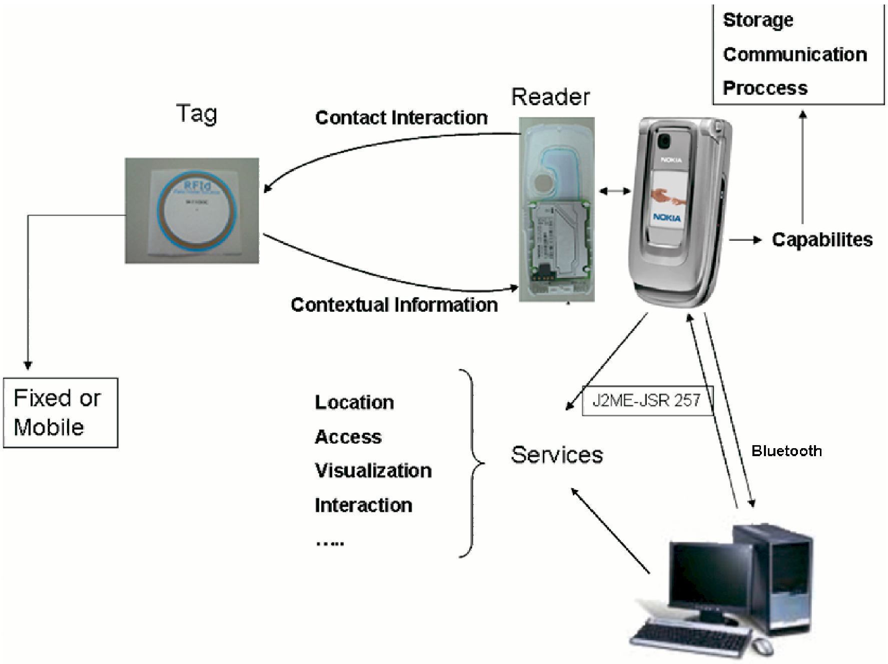


Fig. 3. NFC architecture

4 Tagging context and places

Our first approach for tagging context is studied in an academic environment. We selected two spaces: a research lab and a professor's office. To do so, we arranged the tag structure in three places: the door, the board and the PC display. In them, it is possible to promote interaction with a simple tag and multiple services can be gathered: location, visualization, downloading information and so on.

Figure 4 shows a distribution of information at the mentioned places. All of them have the context identification. This information is crucial in order to check the security interaction by touch. Services on the door are varied: location, access, notices about people looking at you, etc. are achievable. In addition, unexpected situations are provided. For instance, lay down the mobile phone on silent if it is necessary. Furthermore, the user can recover the normal sound status when entering another environment. Therefore, this kind of interaction in a contactless action is embedded, meaning, awareness by contact.

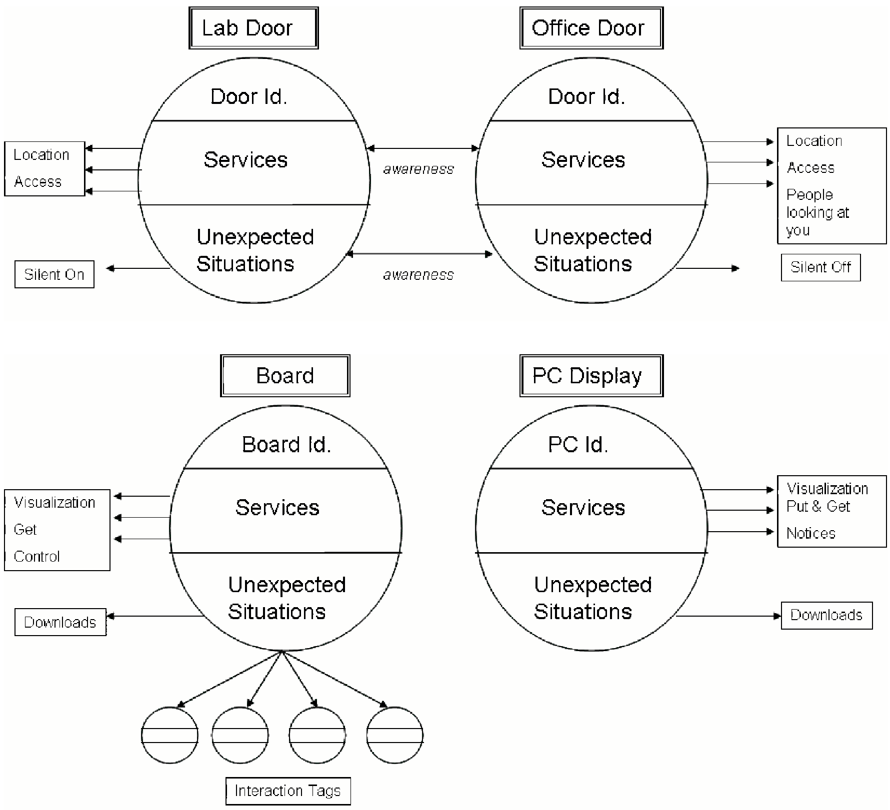


Fig. 4. Information in tags

An important service for us is visualization of information. In order to achieve a single interaction, we have placed a set of tags next to the board. One of these is the control tag to identify people allowed to use the visualization device. In addition, it is possible to initiate interaction with the board, show contents and interact with others tags. With them, it is possible to manage the information shown (select, scrolling, etc.), entering and retrieving documents or notes to comment on. This functionality is similar to a pen-drive. With in and out boxes, it is possible to manage information particularized for every user easily. The PC display offers similar services. In addition, notices informing if the user is not in near the PC, are considered.

4.1 Context-Awareness by Tagging & Touching

As we mentioned before, context-awareness is possible through the contactless user actions by means a simple touch. Therefore, we can talk about embedding interaction in the user actions by means of the information stored in tags. In order to do so, it is first necessary to have an application included into the mobile phone. This generic application solves the download process the first time that people interact with this context. Then, the user only has to touch different tags to execute every process. All interactions, with the proper information in tag and phone, make it possible to safe match both.
A tasks classification on context for tagging is expected:

- Find the place for tags.
- Look for needed information.
- Classify the profile Id., Services and Unexpected Situation (Awareness).
- Develop applications.
- Interaction Features (Automatic running applications, discover nearly computers, download, etc.).

People that usually interact with the context can access needed services, but others need to download applications to interact with it. For this reason, we have developed a “start point” application. It contains a process to download needed applications by touching. To do so, it is necessary for people from the organization to authorize other users. This happens by simply putting two mobile phones near each other and inserting an authorization code giving consent. Finally, every application has to reduce the time for readings. For this reason, we introduce a formalism allowing every application to obtain all the data at the first user tag contact. It is possible through the every tag structure according with the profile context and place.

Another important feature is the possibility to call up applications on the mobile phone contacting tags. This fact reinforces our idea of only a touch being needed. In addition, communication capability, meaning Bluetooth service, are activated automatically when calling application, thus saving phone battery.

It is worth highlighting the effect of the change of interaction that comes about by touching. With traditional RFID technology, interaction is closer to the user because it only requires wearing a tag and placing readers and antennas throughout the building. This case promotes complete control by the server, however. We all know that our daily routine is unpredictable and for this reason, the task responsibility for tasks could be managed by the user- this is possible with NFC technology. With just a simple contact, it is possible to decide when the task can be managed. This fact saves in modeling effort and bears in mind many kinds of everyday interruptions or unexpected situations. (****) Besides all this, other characteristics have to be considered. While in the RFID case the server is defined, in the second one (NFC), the idea of server could "disappear" for some services. That means that every mobile phone should discover a nearly computer for supporting the required service. This fact allows us to get context reaction in places where there is no server. For instance, when I go to a colleague's office, the location service can be managed through his/her computer. It is just the discovering process between mobile phone and the computer that has to be handled. The download service can be administered, too. The idea of disperse information in tags thorough the context, could be an interesting one, bearing in mind aspects such as infrastructure and server responsibilities.

In the next point, we present an approach of interaction with the visualization service that we called "Visualization Mosaic".

5 Tagging Board Interaction

By tagging boards and PC displays users can obtain visualization services to support their daily activities. We propose boards and displays with a "control tag", i. e. a tag that allows specific users to take control of the display/board. The contextual information stored in a tag determines the display behavior according the user profile. Additional tags are placed on the public displays in order to make the interaction with visualization services possible.

When a user or a set of users are placed in front of a display, the traditional human-computer interaction is not suitable because it requires frequent user interventions in order to indicate every action to the computer. Consequently, it is important to reach a more natural interaction making it implicit versus the traditional explicit dialogue with the computer.

First of all, we need to describe ViMos, an architecture that applies context-awareness providing adapted information to the user through embedded devices in the environment.

A mosaic of information is a set of information pieces that form a user interface. Initially we have a whole set of information pieces. A match between the situation and the context model makes the selection of the best pieces possible. Each piece has several associated characteristics (e.g. Optimum size, elasticity, etc.). These characteristics make the final generation process of the mosaic possible.

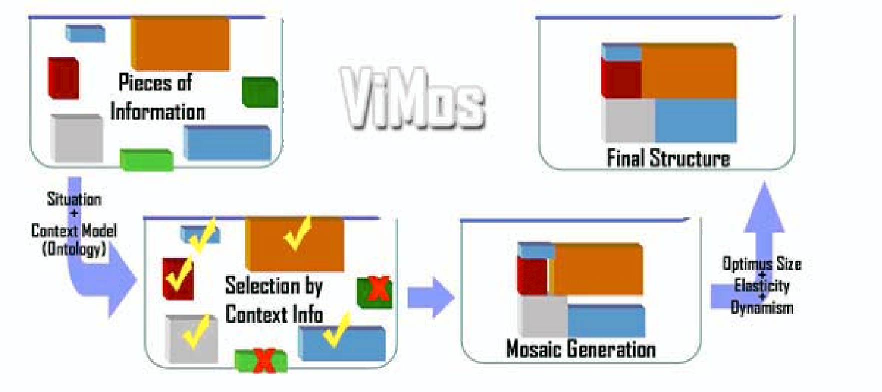


Fig. 5. ViMos Architecture

The generation of information mosaics is carried out by adapting the user interface according to the situation. In this way, we can improve the dynamism, the adaptability and the content quality.
The initial schema of interaction with ViMos was an implicit interaction. The visualization services were offered using the inferred information by a context model (represented by an ontology describing parts of the real world encompassing ViMos) and sensor information (e. g. user presence obtained by RFID antennas). However, there were several limitations, mainly when unexpired situations appeared or in the case of a user needing to navigate among the different parts of the mosaic.
By touching interaction, we can reduce the infrastructure requirements. Whenever a user takes the control of a display, implicit process is activated, such as localization and identification process. The display recognizes the user offering personal visualization services (schedule, timetable, documents in process, etc.). The main visualization services are about collaborative tasks.
In our particular context, the research lab, we have placed two virtual boards. The first one shows important information for the research group. It can be seen in Figure 6. On the right hand side of the board there are some tags for interaction with the mosaic. The second one is the works in progress for supporting collaborative tasks and discussion about new proposals.

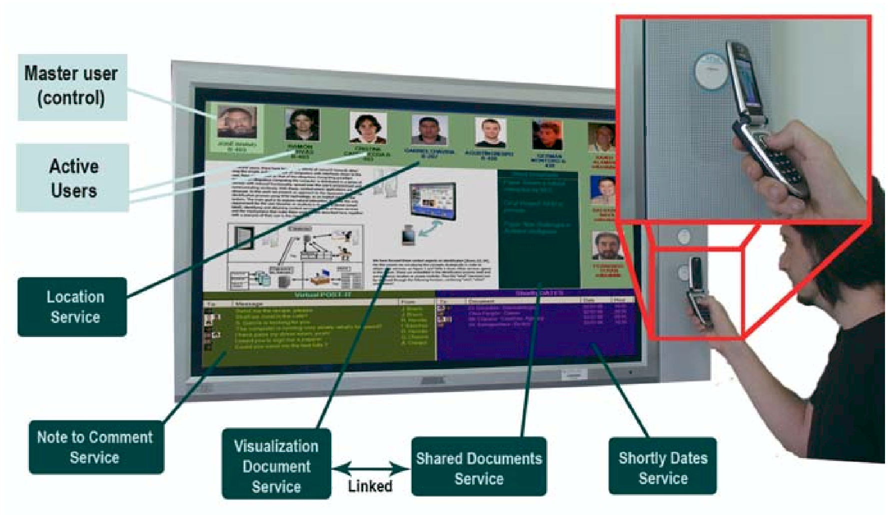


Fig. 6. Interacting with ViMos

The mosaic may be in two states. (a) When there are users in the room but nobody is near the display. ViMos should show information in order to support individual work as well as group work. (b) If one or more users are closer to the display, ViMos should show information on common work of these users. This situation requires previous planning. It offers information about member location, the group schedule, to-do lists, etc. The rest of the mosaic contains specific information regarding the research group, e.g. academic event deadlines, news, notices, member's messages and shared documents. All information changes according to the situation.

The user can take control of the visualization device simply by putting the mobile phone near the control tag. In addition, other users can interact with it. There are some important aspects to consider in order including the varied functionality for tags, visualization services by means of user profile or the interaction control. In Table 1 an example of functionalities can be seen

Active Service	Take control?	Interaction Tag 1	Interaction Tag 2	Interaction Tag 3	Interaction Tag 4	Control Tag
Any	Yes	Next service	Previous service	Activate service	-------	Get control
Any	No	-------------	-----------	-----------	--------	Get control
Shared Document	Yes	Next page	Previous Page	Next document	Previous document	Get control

Shared Document	No	Get document	-----------	----------	----------	Get Control
Note to comment	Yes	Next set of comments	Previous set of comments	Expand comment	Clear comment	Get Control
Note to comment	No	Get note	Get all user notes			Get Control
Presentation	Yes	Next slide	Previous slide	Beginning	End	Get control
Presentation	No	Get presentation				Get Control

Table 1.- Example of tag functionalities

6 Related Works

Although NFC technology is recent, there are some interesting papers for discussion. Anokwa [10] presents an NFC model. With it, when a mobile device scans an item, the item returns the relevant information needed to describe itself. In Kostakos [11] users with mobile phones encrypt and exchange address book information. An interesting tool for practice can be seen in Toumisto [12]. In it an emulator for supporting touching, pointing and scanning is presented. This is used to study the feasibility and usability of the context of physical browsing. In order to support a visual for touch interaction a graphics language is needed. Timo Arnall defines his particular graphic language for touch-based interactions [13].

Public displays have different physical characteristics than PCs. Consequently new interaction paradigms are necessary. This is a well-known challenge and several works have researched this problem. A large number of projects emphasize the use of large interactive/tangible displays (e g. the interactive workspace project [14]). Another trend is interaction using notebook computers or PDS and techniques like pick-and-drop, i. e. to allow a user to pick up a digital object from one screen, and drop it in a different place, typically a different computer screen or public display [15]. In general, Vogl established taxonomy of methods for user input to large displays.

A good method for obtaining input to the systems comes from the devices that identify and locate people. Want and Hopper locate individuals by means of active badges which transmit signals that are picked up by sensors situated in the building [16]. IBM developed the Blueboard experiment with a display and RFID tag for the user’s collaboration [17]. A similar project is IntelliBadge [18] developed for the academic conference context. There are several works on interaction with public displays by mobile devices [19,20]. However, mobile devices have limitations in their input and output capabilities [21]. NFC can take the advantages of mobile devices and can become a valid technology to interact with devices in the environment, such as displays and boards

7 Conclusions

We have adapted a new technology that is supported by devices well known by users and offering some advantages over the traditional RFID. The idea of changing the implicit or embedded interaction that supposes to wear a tag, in front of, the explicit interaction by touch, produces a series of benefits in contexts: architecture, costs, mobile capabilities, and so on.
Our proposal of tagging context reduces the awareness responsibility of the system by means of a single interaction. Tag structures, context profile and services make it easy to interact with the context. In addition, the system allows new services to be implemented as needs appear. This conclusion is possible due the decentralization of services, produced at the time that the user chooses and with the program included in a mobile phone, supported by a server or not.

Acknowledgments

This work has been financed by a 2005-2006 SERVIDOR project (PBI-05-034) from the Junta de Comunidades de Castilla-La Mancha and the Ministerio de Ciencia y Tecnología in the Mosaic Learning project 2005-2007 (TSI2005-08225-C07-07).

References

1. ISTAG, Scenarios for Ambient Intelligence in 2010. Feb. 2001. http://www.cordis.lu/ist/istag.htm.
2. Dey, A. (2001). "Understanding and Using Context". Personal and Ubiquitous Computing 5(1), 2001, pp. 4-7.

3. Kevin Brooks (2003). “The Context Quintet: narrative elements applied to Context Awareness”. In Human Computer Interaction International Proceedings, 2003, (Crete, Greece) by Erlbaum Associates, Inc.
4. Schmidt, A. (2000). “Implicit Human Computer Interaction Through Context”. Personal Technologies Volume 4(2&3) 191-199
5. Schmidt, A. (2005). “Interactive Context-Aware Systems. Intering with Ambient Intelligence”. In Ambient Intelligence. G. Riva, F. Vatalaro, F. Davide & M. Alcañiz (Eds.).
6. Schmidt, A., M. Kranz, and P. Holleis. Interacting with the Ubiquitous Computing – Towards Embedding Interaction. in Smart Objects & Ambient Intelligence (sOc-EuSAI 2005). 2005. Grenoble, Francia.
7. Bravo, J., R. Hervás, and G. Chavira, Ubiquitous computing at classroom: An approach through identification process. Journal of Universal Computer. Special Issue on Computers and Education: Research and Experiences in eLearning Technology. 2005. 11(9): p. 1494-1504.
8. Bravo, J., Hervás, R., Chavira, G., Nava, S. & Sanz, J. (2005). “Display-based services through identification: An approach in a conference context”. Ubiquitous Computing & Ambient Intelligence (UCAmI’05). Thomson. ISBN:84-9732-442-0. pp.3-10.
9. Bravo, J., Hervás, R., Chavira, G. & Nava, S. “Modeling Contexts by RFID-Sensor Fusion”. Accepted paper from 3rd Workshop on Context Modeling and Reasoning (CoMoRea 2006). Pisa (Italy).
10. Yaw Anokwa, Gaetano Borriello, Trevor Pering & Roy Want. “A User Interaction Model for NFC Enabled Applications”. PERTED Worshop at Percom 2007.
11. Kostakos, V., O'Neill, E., Shahi, A. (2006). “Building common ground for face to face interactions by sharing mobile device context”. Workshop on Location and Context Awareness (LOCA 2006), Dublin, Ireland. Lecture Notes in Computer Science 3987, Springer, pp. 222-238.
12. Timo Tuomisto, Pasi Välkkynen, Arto Ylisaukko-oja. “RFID Tag Reader System Emulator to Support Touching, Pointing and Scanning”. In proceedings of Mobile Interaction with the Real World (MIRW 2006)
13. Timo Arnall. “A graphic language for touch-based interactions”. In Proceedings of Mobile Interaction with the Real World MIRW 2006).
14. Johanson, B., Fox, A, Winograd, T. (2002) “The Interactive Workspace Project: Experiences with Ubiquitous Computing Rooms”. Pervasive Computing 2(1) pp 71-78
15. Vogl, S.: Coordination of Users and Services via Wall Interfaces. PhD Thesis. Johannes Kepler. University Linz. 2002.
16. [Want, 92] Want, R. & Hopper, A. (1992). “The Active Badge Location System”. ACM Transactions on Information Systems, 10(1):91–102, Jan 1992.
17. Daniel M. Russell, Jay P. Trimble, Andreas Dieberger. “The use patterns of large, interactive display surfaces: Case studies of media design and use for BlueBoard and MERBoard”. In Proceedings of the 37th Hawaii International Conference on System Sciences – 2004.

18. Donna Cox, Volodymyr Kindratenko, David Pointer In Proc. of 1st International Workshop on Ubiquitous Systems for Supporting Social Interaction and Face-to-Face Communication in Public Spaces, 5th Annual Conference on Ubiquitous Computing, October 12- 4, 2003, Seattle, WA, pp. 41-47.
19. Huang, E.; Mynatt, E. (2003) "Semi-public displays for small, co-located groups". In: Proceedings of the conference on Human factors in computing systems. , (Ft. Lauderdale, Florida, USA.).
20. Aizawa, K.; Kakami, K.; Nakahira, K. "Ubiquitous Displays for Cellular Phone Based Personal Information Environments". In: IEEE Pacific Rim Conference on Multimedia (PCM2002).
21. Rukzio, E., Schmidt, a., Hussmann, H. (2004) "An Analysis of the Usage of Mobile Phones for Personalized Interactions with Ubiquitous Public Displays" In: Workshop on Ubiquitous Display Environments in conjunction with UbiComp (Nottingham, UK).

Model-Based Reasoning Methods within an Ambient Intelligent Agent Model

Tibor Bosse, Fiemke Both, Charlotte Gerritsen, Mark Hoogendoorn, and Jan Treur

Vrije Universiteit Amsterdam, Department of Artificial Intelligence
De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands
{tbosse, fboth, cg, mhoogen, treur}@few.vu.nl
http://www.few.vu.nl/~{tbosse, fboth, cg, mhoogen, treur}

Abstract. Ambient agents react on humans on the basis of their information obtained by sensoring and their knowledge about human functioning. Appropriate types of reactions depend on in how far an ambient agent understands the human. On the one hand, such an understanding requires that the agent has knowledge to a certain depth about the human's physiological and mental processes in the form of an explicitly represented model of the causal and dynamic relations describing these processes. On the other hand, given such a model representation, the agent needs reasoning methods to derive conclusions from the model and the information available by sensoring. This paper presents a number of such model-based reasoning methods. They have been formally specified in an executable temporal format, which allows for simulation of reasoning traces and automated verification in a dedicated software environment. A number of such simulation experiments and their formal analysis are described.

1 Introduction

Recent developments within Ambient Intelligence provide technological possibilities to contribute to personal care; cf. [1, 2, 18]. Such applications can be based on possibilities to acquire sensor information about humans and their functioning, but more substantial applications depend on the availability of adequate knowledge for analysis of information about human functioning. If knowledge about human functioning is explicitly represented in the form of computational models in ambient agents, these agents can show more understanding, and (re)act accordingly by undertaking actions in a knowledgeable manner that improve the human's wellbeing and performance. In recent years, human-directed scientific areas such as cognitive science, psychology, neuroscience and biomedical sciences have made substantial progress in providing an increased insight in the various physical and mental aspects involved in human functioning. Although much work still remains to be done, dynamic models have been developed and formalised for a variety of such aspects and the way in which humans (try to) manage or regulate them. From a biomedical angle, examples of such aspects are (management of) heart functioning, diabetes, eating regulation disorders, and HIV-infection; e.g., [5, 15]. From a psychological and social angle, examples are emotion regulation, attention regulation, addiction management, trust management, stress management, and criminal behaviour management; e.g., [6, 11, 16]. Such models can be the basis for dedicated model-based reasoning methods that allow an agent to derive relevant conclusions from these models and available sensor information.

This paper addresses the design of ambient agents that have knowledge about human behaviours and states over time in the form of explicitly represented models of the causal and dynamical relations involved. First it is shown how such models can be formally represented in a logical format that also integrates numerical aspects; cf. [9]. Next a number of logical reasoning methods are presented that are based on such models. These reasoning methods are represented in a temporal logical format according to the approach put forward in [14]. A number of simulation experiments to obtain reasoning traces are described. These traces have been formally analysed by a dedicated verification tool. The types of reasoning methods addressed cover a variety of phenomena such as causal and numerical simulation, qualitative reasoning and simulation, abductive reasoning [17], and explanation generation. The reasoning methods provide a conceptual and logical foundation for these phenomena. Moreover, they provide a solid basis for conceptual and detailed design of model-based ambient agents that need such capabilities.

Section 2 describes the formal modelling approach that is used throughout this paper. Next, in Section 3 and 4 the reasoning methods themselves are presented. Section 3 addresses uncontrolled methods for belief generation, and Section 4 addresses controlled methods for belief generation. Section 5 illustrates how these reasoning methods can be used, by performing simulation experiments in two example case studies. Section 6 provides a number of basic properties that may hold for model-based reasoning methods within ambient agents. Section 7 addresses verification of basic properties as introduced in Section 3 against simulation traces, and interlevel relations between properties at different aggregation levels. Section 8 concludes the paper with a discussion.

2 Modelling Approach

This section introduces the formal modelling approach that is used throughout this paper. Section 2.1 briefly describes the Temporal Trace Language (TTL) for specification of dynamic properties (and its executable sublanguage LEADSTO), and Section 2.2 briefly explains how reasoning methods are formalised in this paper.

2.1 The Temporal Trace Language TTL

In order to execute and verify human-like ambience models, the expressive language TTL is used [7]. This predicate logical language supports formal specification and analysis of dynamic properties, covering both qualitative and quantitative aspects. TTL is built on atoms referring to states, time points and traces. A *state* of a process for (state) ontology Ont is an assignment of truth values to the set of ground atoms in the ontology. The set of all possible states for ontology Ont is denoted by STATES(Ont). To describe sequences of states, a fixed *time frame* T is assumed which is linearly ordered. A *trace* γ over state ontology Ont and time frame T is a mapping γ : T $\rightarrow$ STATES(Ont), i.e., a sequence of states γ_t ($t \in T$) in STATES(Ont). The set of *dynamic properties* DYNPROP(Ont) is the set of temporal statements that can be formulated with respect to traces based on the state ontology Ont in the following manner. Given a trace γ over state ontology Ont, the state in γ at time point t is denoted by state(γ, t). These states can be related to state properties via the formally defined satisfaction relation |=, comparable to the Holds-predicate in the Situation Calculus: state(γ, t) |= p denotes that state property p holds in trace γ at time t. Based on these statements, dynamic properties can be formulated in a sorted first-order predicate logic, using quantifiers over time and traces and the usual first-order logical connectives such as $\neg$, $\wedge$, $\vee$, $\Rightarrow$, $\forall$, $\exists$. A special software environment has been developed for TTL, featuring both a Property Editor for building and editing TTL properties and a Checking Tool that enables formal verification of such properties against a set of (simulated or empirical) traces.

Executable Format To specify simulation models and to execute these models, the language LEADSTO [8], an executable sublanguage of TTL, is used. The basic building blocks of this language are causal relations of the format $\alpha \twoheadrightarrow_{e, f, g, h} \beta$, which means:

if	state property α holds for a certain time interval with duration g,
then	after some delay (between e and f) state property β will hold for a certain time interval of length h.

where α and β are state properties of the form 'conjunction of literals' (where a literal is an atom or the negation of an atom), and e, f, g, h non-negative real numbers.

2.2 Temporal Specification of Reasoning Methods

In this paper a dynamic perspective on reasoning is taken, following, e.g.. [14]. In practical reasoning situations usually different lines of reasoning can be generated, each leading to a distinct set of conclusions. In logic semantics is usually expressed in terms of models that represent descriptions of conclusions about the world and in terms of entailment relations based on a specific class of this type of models. In the (sound) classical case each line (trace) of reasoning leads to a set of conclusions that are true in all of these models: each reasoning trace fits to each model. However, for non-classical reasoning methods the picture is different. For example, in default reasoning or abductive reasoning methods a variety of mutually contradictory conclusion sets may be possible. It depends on the chosen line of reasoning which one of these sets fits.

The general idea underlying the approach followed here, and inspired by [14] is that a particular reasoning trace can be formalised by a sequence of *information states* $M_0, M_1, \ldots\ldots$. Here any M_t is a description of the (partial) information that has been derived up to time point t. From a dynamic perspective, an inference step, performed in time duration D is viewed as a transition $M_t \rightarrow M_{t+D}$ of a current information state M_t to a next information state M_{t+D}. Such a transition is usually described by application of a deduction rule or proof rule, which in the dynamic perspective on reasoning gets a temporal aspect. A particular reasoning line is formalised by a sequence $(M_t)_{t \in T}$ of subsequent information states labelled by elements of a flow of time T, which may be discrete, based on natural numbers, or continuous, based on real numbers.

An information state can be formalised by a set of statements, or as a three-valued (false, true, undefined) truth assignment to ground atoms, i.e., a partial model. In the latter case, which is followed here (as in [14]), a sequence of such information states or reasoning trace can be interpreted as a partial temporal model. A transition relating a next information state to a current one can be formalised by temporal formulae the partial temporal model has to satisfy. For example, a modus ponens deduction rule can be specified in temporal format as:

derived(I) ∧ derived(implies(I, J)) ↠ derived(J)

So, inference rules are translated into temporal rules thus obtaining a temporal theory describing the reasoning behaviour. Each possible reasoning trace can be described by a linear time model of this theory (in temporal partial logic).

In this paper, this dynamic perspective on reasoning is applied in combination with facts that are labelled with temporal information, and models based on causal or temporal relationships that relate such facts. To express the information involved in an agent's internal reasoning processes, the ontology shown in Table 1 is used.

Predicate	**Description**
belief(I:INFO_EL)	information I is believed
world_fact(I:INFO_EL)	I is a world fact
has_effect(A:ACTION, I:INFO_EL)	action A has effect I
Function to INFO_EL	**Description**
leads_to_after(I:INFO_EL, J:INFO_EL, D:REAL)	state property I leads to state property J after duration D
at(I:INFO_EL, T:TIME)	state property I holds at time T

Table 1. Generic Ontology used within the Ambient Agent Model

As an example belief(leads_to_after(I:INFO_EL, J:INFO_EL, D:REAL)) is an expression based on this ontology which represents that the agent has the knowledge that state property I leads to state property J with a certain time delay specified by D. An example of a kind of dynamic modus ponens rule can be specified as

belief(at(I, T)) ∧ belief(leads_to_after(I, J, D)) → belief(at(J, T+D))

This temporal rule states that if it is believed (by the agent) that I holds at T and that I leads to J after duration D, then it will be believed that J holds at T + D. This representation format will be used to formalise this and other types of model-based reasoning methods, as will be shown more extensively in Sections 3 and 4.

3 Model-Based Reasoning Methods for Belief Generation

Two types of reasoning methods to generate beliefs can be distinguished:

- *Forward reasoning methods for belief generation*

These are reasoning methods that follow the direction of time and causality, deriving from beliefs about properties at certain time points, new beliefs about properties at later time points.

- *Backward reasoning methods for belief generation*
  These are reasoning methods that follow the opposite direction of time and causality, deriving from beliefs about properties at certain time points, new beliefs about properties at earlier time points.

In Section 3.1 the forward reasoning methods for belief generation are discussed, in Section 3.2 the backward reasoning methods.

3.1 Forward reasoning methods for belief generation

Forward reasoning methods are often used to make predictions on future states, or on making an estimation of the current state based on information acquired in the past. The first reasoning method is one that occurs in the literature in many variants, in different contexts and under different names, varying from, for example, computational (numerical) simulation based on difference or differential equations, qualitative simulation, causal reasoning, execution of executable temporal logic formulae, and forward chaining in rule-based reasoning, to generation of traces by transition systems and finite automata. The basic specification of this reasoning method can be expressed as follows.

Belief Generation based on Positive Forward Simulation

If it is believed that I holds at T and that I leads to J after duration D, then it is believed that J holds after D.
∀I,J:INFO_EL ∀D:REAL ∀T:TIME
belief(at(I, T)) ∧ belief(leads_to_after(I, J, D)) ↠ belief(at(J, T+D))
If it is believed that I1 holds at T and that I2 holds at T, then it is believed that I1 and I2 holds at T.
belief(at(X1,T)) ∧ belief(at(X2, T)) ↠ belief(at(and(X1, X2), T))

Note that, if the initial beliefs are assumed correct, belief correctness holds for leads to beliefs, and positive forward correctness of leads to relationships holds, then all beliefs generated in this way are correct. A second way of belief generation by forward simulation addresses the propagation of negations. This is expressed as follows.

Belief Generation based on Single Source Negative Forward Simulation

If it is believed that I does not hold at T and that I leads to J after duration D, then it is believed that J does not hold after D.
∀I,J:INFO_EL ∀D:REAL ∀T:TIME
belief(at(not(I), T)) ∧ belief(leads_to_after(I, J, D)) ↠ belief(at(not(J), T+D)))
If it is believed that I1 (resp. I2) does not hold at T, then it is believed that I1 and I2 does not hold at T.
belief((at(not(I1),T))) ↠ belief(at(not(and(I1, I2)), T))
belief(at(not(I2),T)) ↠ belief(at(not(and(I1, I2)), T))

Note that this only provides correct beliefs when the initial beliefs are assumed correct, belief correctness holds for leads to beliefs, and single source negative forward correctness holds for the leads to relationships.

Belief Generation based on Multiple Source Negative Forward Simulation

If for any J and time T, for every I that is believed to lead to J after some duration D, it is believed that I does not hold before duration D, then it is believed that J does not hold.
∀I,J:INFO_EL ∀D:REAL ∀T:TIME
∀I, D [belief(leads_to_after(I, J, D)) → belief(at(not(I), t-D)] ↠ belief(at(not(J), T))

If it is believed that I1 (resp. I2) does not hold at T, then it is believed that I1 and I2 does not hold at T.
belief(at(not(I1),T)) ↠ belief(at(not(and(I1, I2)), T))
belief(at(not(I2),T)) ↠ belief(at(not(and(I1, I2)), T))

This provides correct beliefs when the initial beliefs are assumed correct, belief correctness holds for leads to beliefs, and multiple source negative forward correctness holds for the leads to relationships.

3.2 Backward reasoning methods for belief generation

The basic specification of a backward reasoning method is specified as follows.

Belief Generation based on Modus Tollens Inverse Simulation

If it is believed that J does not hold at T and that I leads to J after duration D, then it is believed that I does not hold before duration D.
∀I,J:INFO_EL ∀D:REAL ∀T:TIME
belief(at(not(J), T)) ∧ belief(leads_to_after(I, J, D)) ↠ belief(at(not(I), T-D))
If it is believed that not I1 and I2 holds at T and that I2 (resp. I1) holds at T, then it is believed that I1 (resp. I2) does not hold at T.
belief(at(not(and(I1, I2), T)) ∧ belief(at(I2, T)) ↠ belief(at(not(I1), T))
belief(at(not(and(I1, I2), T)) ∧ belief(at(I1, T)) ↠ belief(at(not(I2), T))

Belief Generation based on Simple Abduction

If it is believed that J holds at T and that I leads to J after duration D, then it is believed that I holds before duration D.
∀I,J:INFO_EL ∀D:REAL ∀T:TIME
belief(at(J, T)) ∧ belief(leads_to_after(I, J, D)) ↠ belief(at(I, T-D))
If it is believed that I1 and I2 holds at T, then it is believed that I1 holds at T and that I2 holds at T.
belief(at(and(I1, I2), T)) ↠ belief(at(I1,T)) ∧ belief(at(I2, T))

As another option, an abductive causal reasoning method can be internally represented in a simplified form as follows.

Belief Generation based on Multiple Effect Abduction

If for any I and time T, for every J for which it is believed that I leads to J after some duration D, it is believed that J holds after duration D, then it is believed that I holds at T.
∀I:INFO_EL ∀T:TIME
∀J [belief(leads_to_after(I, J, D)) → belief(at(J, T+D))] ↠ belief(at(I, T))
If it is believed that I1 and I2 holds at T, then it is believed that I1 holds at T and that I2 holds at T.
belief(at(and(I1, I2), T)) ↠ belief(at(I1,T)) ∧ belief(at(I2, T))

Belief Generation based on Context-Supported Abduction

If it is believed that J holds at T and that I2 holds at T and that I1 and I2 leads to J after duration D, then it is believed that I1 holds before duration D.
∀I,J:INFO_EL ∀D:REAL ∀T:TIME
belief(at(J, T)) ∧ belief(at(I2, T-D)) ∧ belief(leads_to_after(and(I1, I2), J, D)) ↠ belief(at(I1, T-D))
If it is believed that I1 and I2 holds at T, then it is believed that I1 holds at T and that I2 holds at T.
belief(at(and(I1, I2), T)) ↠ belief(at(I1,T)) ∧ belief(at(I2, T))

4 Controlling Belief Generation

An uncontrolled belief generation approach may easily lead to a combinatorial explosion of generated beliefs, for example, based on all conjunctions that can be formed. Therefore, controlled approaches where selection is done in some stage of the process are usually more effective. Often more specific knowledge is available based on which belief generation can leave out of consideration some (or most) of the possible beliefs that can be generated. To incorporate such selections, the following three approaches are possible: *selection afterwards overall, selection afterwards step by step, selection before*. Each of these options is discussed in more detail. Furthermore, it is discussed what selection criteria can be used to make such a selection.

Belief Generation Selection

Selection Afterwards Overall

In this approach first (candidate) beliefs are generated in an uncontrolled manner, and after that a selection process is performed based on some selection criterion. Two examples, one for a forward belief generation form and one for a backward belief generation form are as follows.

Controlled Belief Generation based on Positive Forward Simulation by Selection Afterwards Overall

If it is believed that I holds at T and that I leads to J after duration D, then it is believed that J holds after D.
∀I,J:INFO_EL ∀D:REAL ∀T:TIME
belief(at(I, T)) ∧ belief(leads_to_after(I, J, D)) ↠ belief(at(J, T+D))
If it is believed that I1 holds at T and that I2 holds at T, then it is believed that I1 and I2 holds at T.
belief(at(I1,T)) ∧ belief(at(I2, T)) ↠ belief(at(and(I1, I2), T))
If I is a belief and selection criterion s is fulfilled, then I is a selected belief.
belief(at(I, T)) ∧ s ↠ selected_belief(at(I, T))

Controlled Belief Generation based on Multiple Effect Abduction by Selection Afterwards Overall

If for any I and time T, for every J for which it is believed that I leads to J after some duration D, it is believed that J holds after duration D, then it is believed that I holds at T.
∀I:INFO_EL ∀T:TIME
∀J [belief(leads_to_after(I, J, D)) → belief(at(J, T+D))] ↠ belief(at(I, T))
If it is believed that I1 and I2 holds at T, then it is believed that I1 holds at T and that I2 holds at T.
belief(at(and(I1, I2), T)) ↠ belief(at(I1,T)) ∧ belief(at(I2, T))
If I is a belief and selection criterion s is fulfilled, then I is a selected belief.
belief(at(I, T)) ∧ s ↠ selected_belief(at(I, T))

This approach to control can only be applied when the number of beliefs that is generated in an uncontrolled manner is small. Otherwise more local approaches are better candidates to consider.

Selection Afterwards Step by Step

The step by step variant of selection afterwards performs the selection immediately after a belief has been generated. By such a local selection it is achieved that beliefs that are not selected can not be used in further belief generation processes, thus limiting these processes. The approach uses the temporal selection rule given above together with a slightly adapted form of specification to generate beliefs. Again two examples, one for a forward belief generation form and one for a backward belief generation form are as follows.

Controlled Bel. Generation based on Positive Forward Simulation by Selection Aft. Step by Step

If it is believed that I holds at T and that I leads to J after duration D, then it is believed that J holds after D.
∀I,J:INFO_EL ∀D:REAL ∀T:TIME
selected_belief(at(I, T)) ∧ belief(leads_to_after(I, J, D)) ↠ belief(at(J, T+D))
If it is believed that I1 holds at T and that I2 holds at T, then it is believed that I1 and I2 holds at T.
selected_belief(at(I1,T)) ∧ selected_belief(at(I2, T)) ↠ belief(at(and(I1, I2), T))
If I is a belief and selection criterion s is fulfilled, then I is a selected belief.
belief(at(I, T)) ∧ s ↠ selected_belief(at(I, T))

Controlled Belief Generation based on Multiple Effect Abduction by Selection Aft. Step by Step

If for any I and time T, for every J for which it is believed that I leads to J after some duration D, it is believed that J holds after duration D, then it is believed that I holds at T.

∀I:INFO_EL ∀T:TIME

∀J [belief(leads_to_after(I, J, D)) → selected_belief(at(J, T+D))] ↠ belief(at(I, T))

If it is believed that I1 and I2 holds at T, then it is believed that I1 holds at T and that I2 holds at T.

selected_belief(at(and(I1, I2), T)) ↠ belief(at(I1,T)) ∧ belief(at(I2, T))

If I is a belief and selection criterion s is fulfilled, then I is a selected belief.

belief(at(I, T)) ∧ s ↠ selected_belief(at(I, T))

This selection approach may be much more efficient than the approach based on selection afterwards overall.

Selection Before

The approach of selection afterwards step by step can be slightly modified by not selecting the belief just after its generation, but just before. This allows fo
r a still more economic process of focus generation. Again two examples, one for a forward belief generation form and one for a backward belief generation form are as follows.

Controlled Belief Generation based on Positive Forward Simulation by Selection Before

If it the belief that I holds at T was selected and it is believed that I leads to J after duration D, and selection criterion s1 holds, then the belief that J holds after D is selected.

∀I,J:INFO_EL ∀D:REAL ∀T:TIME

selected_belief(at(I, T)) ∧ belief(leads_to_after(I, J, D)) ∧ s1 ↠ selected_belief(at(J, T+D))

If the beliefs that I1 holds at T and that I2 holds at T were selected, and selection criterion s2 holds, then the conjunction of I1 and I2 at T is a selected belief.

selected_belief(at(I1,T)) ∧ selected_belief(at(I2, T)) ∧ s2 ↠ selected_belief(at(and(I1, I2), T))

Controlled Belief Generation based on Multiple Effect Abduction by Selection Before

If for any I and time T, for every J for which it is believed that I leads to J after some duration D, the belief that J holds after duration D was selected, and selection criterion s1 holds, then it the belief that I holds at T is a selected belief.

∀I:INFO_EL ∀T:TIME

∀J [belief(leads_to_after(I, J, D)) → selected_belief(at(J, T+D))] ∧ s1 ↠ selected_belief(at(I, T))

If the beliefs that I1 and I2 holds at T were selected, and selection criterion s2 holds then the belief that I1 holds at T is a selected belief.

selected_belief(at(and(I1, I2), T)) ∧ s2 ↠ selected_belief(at(I1,T))

If the beliefs that I1 and I2 holds at T were selected, and selection criterion s2 holds then the belief that I2 holds at T is a selected belief

selected_belief(at(and(I1, I2), T)) ∧ s3 ↠ selected_belief(at(I2, T))

4.2 Selection Criteria in Reasoning Methods for Belief Generation

Selection criteria needed for controlled belief generation can be specified in different manners. A simple manner is by assuming that the agent has knowledge which beliefs are relevant, expressed by a predicate in_focus. If this assumption is made, then any selection criterion s can be expressed as in_focus(I), where I is the property for which a belief is considered. The general idea is that if a belief can be generated, it is selected (only) when it is in focus. For example, for the two methods for selection afterwards, the temporal rule will be expressed as:

belief(at(I, T)) ∧ in_focus(I) ↠ selected_belief(at(I, T))

For the method based on selection before, based on focus information the temporal rules will be expressed for the forward example by:

∀I,J:INFO_EL ∀D:REAL ∀T:TIME
selected_belief(at(I, T)) ∧ belief(leads_to_after(I, J, D)) ∧ in_focus(J) →→ selected_belief(at(J, T+D))
selected_belief(at(I1,T)) ∧ selected_belief(at(I2, T)) ∧ in_focus(and(I1, I2)) →→ selected_belief(at(and(I1, I2), T))

For the backward example of the method based on selection before, the temporal rules will be expressed by:

∀I:INFO_EL ∀T:TIME
∀J [belief(leads_to_after(I, J, D)) → selected_belief(at(J, T+D))] ∧ in_focus(I) →→ selected_belief(at(I, T))
selected_belief(at(and(I1, I2), T)) ∧ in_focus(I1) →→ selected_belief(at(I1,T))
selected_belief(at(and(I1, I2), T)) ∧ in_focus(I2) →→ selected_belief(at(I2, T))

It is beyond the scope of this paper whether such foci may be static or dynamic and how they can be determined by an agent. For cases that such general focus information is not available, the selection criteria can be specified in different manners.

5 Simulation

This section illustrates for a number of the reasoning methods provided in the previous sections how they can be used within ambient agents that perform model-based reasoning. This is done by means of two example case studies, each involving an ambient system that uses a causal dynamic model to represent the behaviour of a human, and uses the reasoning methods to determine the state of the human in a particular situation. Section 5.1 focuses on a system that monitors the state of car drivers in order to avoid unsafe driving. Section 5.2 addresses an ergonomic system that monitors the stress level of office employees. Both case studies have been formalised and, using the LEADSTO simulation software [8], have been used to generate a number of simulation traces. In this section, for each model one example simulation trace is shown. More simulation traces can be found in the Appendix on[1].

5.1 Ambient Driver Model

The example model used as an illustration in this section is inspired by a system designed by Toyota which monitors drivers in order to avoid unsafe driving. The system can basically measure drug level in the sweat of a driver (e.g., via a sensor at the steering wheel, or at an ankle belt), and monitor steering operations and the gaze of the driver. Note that the system is still in the experimental phase. The model used in this paper describes how a high drug intake leads to a high drug level in the blood and this leads to physiological and behavioural consequences: (1) physiological: a high drug level (or a substance relating to the drug) in the sweat, (2) behavioural: abnormal steering operation and an unfocused gaze. The dynamic model is represented within the ambient agent by the following beliefs (where D is an arbitrary time delay):

belief(leads_to_after(drug_intake_high, drug_in_blood_high, D))
belief(leads_to_after(drug_in_blood_high, drug_in_sweat_high, D))

[1] http://www.cs.vu.nl/~mhoogen/reasoning/appendix-rm-ami.pdf

belief(leads_to_after(drug_in_blood_high, abnormal_steering_operation, D)
belief(leads_to_after(drug_in_blood_high, unfocused_gaze, D)

Figure 1 shows this dynamical model in a graphical form.

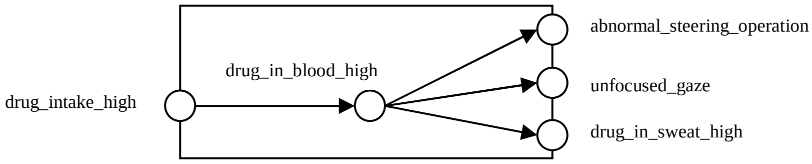


Fig. 1. Graphical representation of the dynamical model

By applying the different reasoning methods specified in Section 3 and 4, the state of the driver and the expected consequences can be derived. In the simulations below the controlled belief generation method has been used based on selection before beliefs are generated; every temporal rule requires that certain selection criteria are met and that the belief to be derived is in focus. In the following simulations, for the sake of simplicity all information is desired, therefore all derivable beliefs are in focus. The selection criteria involve knowledge about the number of effects and sources that are required to draw conclusions. The knowledge used in this model is the following.

sufficient_evidence_for(and(abnormal_steering_operation, unfocused_gaze), drug_in_blood_high)
sufficient_evidence_for(drug_in_sweat_high, drug_in_blood_high)
sufficient_evidence_for(drug_in_blood_high, drug_intake_high)
in_focus(drug_intake_high); in_focus(drug_in_blood_high); in_focus(drug_in_sweat_high);
in_focus(abnormal_steering_operation); in_focus(unfocused_gaze)

Here, the predicate sufficient_evidence_for(P, Q) represents the belief that expression P is sufficient evidence for the system to derive Q. An example simulation trace is shown in Figure 2. In the Figure, the left side shows the atoms that occur during the simulation, whereas the right side represents a time line where a grey box indicates an atom is true at that time point, and a light box indicates false. In this trace, it is known (by observation) that the driver is steering abnormally and that the driver’s gaze is unfocused. Since these two beliefs are sufficient evidence for a high drug level in the blood, using the reasoning method Belief Generation based on Multiple Effect Abduction, at(drug_in_blood_high, 1) becomes a selected belief at time point 3. Given this derived belief, the belief can be deduced that the drug level in the sweat of the driver is high, using Positive Forward Simulation. At the same time (time point 4), the reasoning method Simple Abduction determines the belief that the drug intake of the driver must have been high.

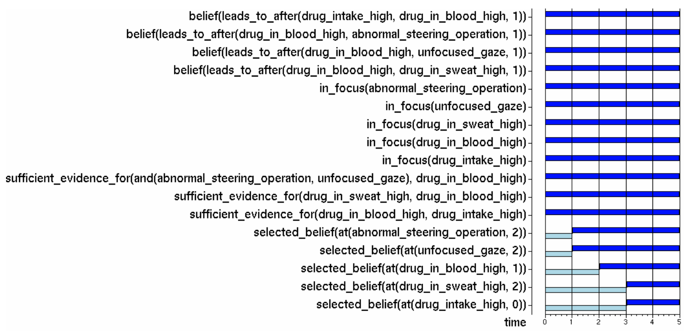


Fig. 2. Simulation Trace: abnormal steering and unfocused gaze detected

5.2 Ambient Stress Model

The example model used in this section is inspired by ergonomic systems that monitor the activities of office employees in their workspace, e.g., in order to avoid RSI (for example, WorkPace, see http://workpace.com/). Such systems may measure various types of information. In this section, three types of measurable (sensor) information are taken into account, namely *actions* (e.g., mouse clicks or key strokes), *biological aspects* (e.g., heart beat, temperature, or skin conductivity), and *activities* (e.g., incoming e-mails, telephone calls, or electronic agenda items). The model considered here describes how (the observation of) a certain activity can lead to a high level of stress and this leads to biological/physiological and behavioural consequences: (1) biological: called here 'high biological aspect' (e.g., increased heart rate) (2) behavioural: changed action (e.g., high number of keystrokes per second). The dynamical model is represented within the ambient agent by the following beliefs:

belief(leads_to_after(activity, observes(activity), D))
belief(leads_to_after(observes(activity), preparedness_to_act, D))
belief(leads_to_after(observes(activity), stress(high), D))
belief(leads_to_after(preparedness_to_act, stress(high), D))
belief(leads_to_after(stress(high), preparedness_to_act, D))
belief(leads_to_after(preparedness_to_act, action, D))
belief(leads_to_after(stress(high), biological_aspect, D))

Figure 3 shows this dynamical model in a graphical form.

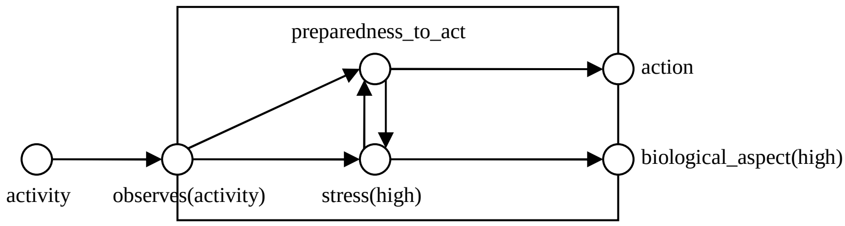


Fig. 3. Graphical representation of the dynamical model

Similar to Section 5.1, by applying the different reasoning methods specified earlier, the expected consequences for the state of the human and can be derived. Below, a number of simulation traces are shown, each with different settings for the selection criteria:

sufficient_evidence_for(biological_aspect(high), stress(high))
sufficient_evidence_for(observes(activity), activity)
sufficient_evidence_for(preparedness_to_act, stress(high))
sufficient_evidence_for(preparedness_to_act, observes(activity))
sufficient_evidence_for(stress(high), preparedness_to_act)
sufficient_evidence_for(stress(high), observes(activity))
sufficient_evidence_for(action, preparedness_to_act)

in_focus(action); in_focus(biological_aspect(high); in_focus(stress(high));
in_focus(observes(activity)); in_focus(activity)

In other words, by selecting different combinations of these criteria, different reasoning steps will be performed. Notice that the model considered here contains a cycle (see Figure 3). Therefore it is possible to derive an infinite number of beliefs for different time points. For example, if at(preparedness_to_act, 8) is believed, then by simple Positive Forward Simulation also at(stress(high), 9) would be derived, after which at(preparedness_to_act, 10) would be derived, and so on. However, it is not conceptually realistic, nor desirable that an agent attempts to derive beliefs about time points very far in the future. Therefore, by means of the in_focus predicate, an indication of a focus time interval has been specified, for example by statements like in_focus(at(preparedness_to_act, 8)).

An example simulation trace is shown in Figure 4. This trace uses as foci all possible information between time point 0 and 10. These foci have been derived using the following rule:

in_focus(I) $\wedge$ $0 \leq T \leq 10$ $\twoheadrightarrow$ in_focus(at(I, T))

The only initially available knowledge that is present in this trace is at(action, 5). As shown in the figure, both Positive Forward Simulation and Simple Abduction are performed several times, eventually leading to all possible derivable information between time point 0 and 10.

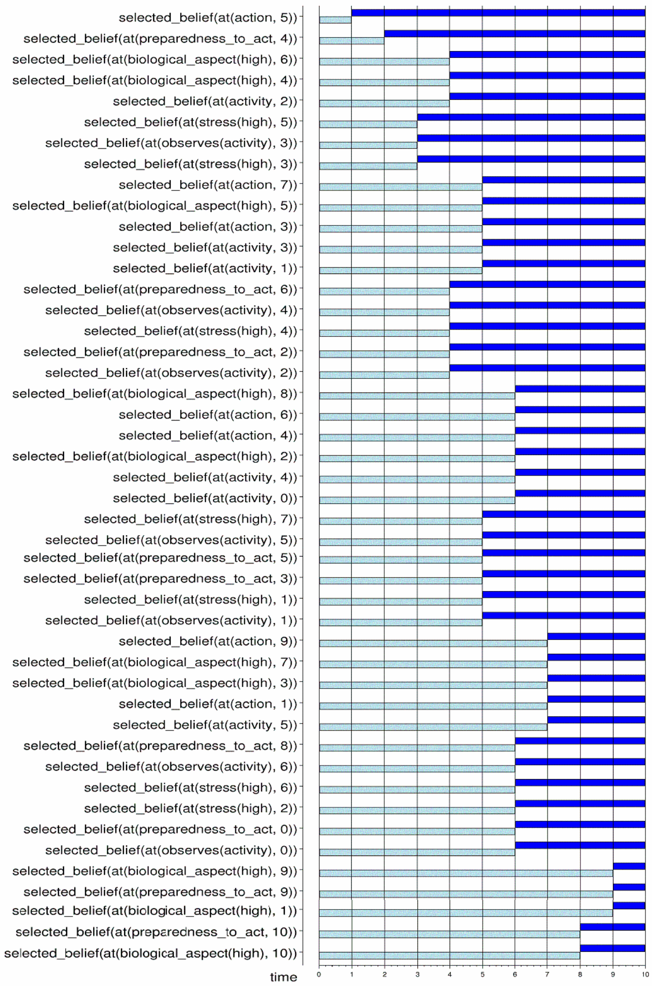


Fig. 4. Simulation Trace: Employee performs active behaviour (to be continued on next page)

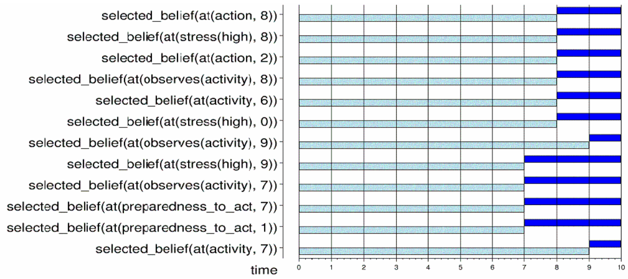


Fig. 4. Simulation Trace: Employee performs active behaviour (continued from previous page)

6 Basic Properties of World Facts, Beliefs and Leads To Relations

This section provides a number of basic properties that may hold for model-based reasoning methods within ambient agents. Section 6.1 addresses properties of world facts and beliefs; Section 6.2 addresses properties of LEADSTO relations.

6.1 Properties of world facts and beliefs

The following basic assumptions concerning two-valued world facts may hold:

Consistency of world facts In any state, it never happens that a world fact and its negation both hold.

$\text{not}\ [\ \text{state}(\gamma, t) \models \text{world\_fact}(I)\ \ \&\ \ \text{state}(\gamma, t) \models \text{world\_fact}(\text{not}(I))\ ]$

Completeness of world facts In any state, for any world fact it holds or its negation holds.

$\text{state}(\gamma, t) \models \text{world\_fact}(I)\ \ |\ \ \text{state}(\gamma, t) \models \text{world\_fact}(\text{not}(I))$

Consistency and completeness of world facts In any state, for any world fact it holds if and only if its negation does not hold

$\text{state}(\gamma, t) \models \text{world\_fact}(I)\ \Leftrightarrow\ \text{not}\ \ \text{state}(\gamma, t) \models \text{world\_fact}(\text{not}(I))$

Belief consistency In any state, it never happens that a fact and its negation are both believed.

$\text{not}\ [\ \text{state}(\gamma, t) \models \text{belief}(I)\ \ \&\ \ \text{state}(\gamma, t) \models \text{belief}(\text{not}(I))\ ]$

Belief correctness In any state, when a fact is believed it holds as a world fact.

$\text{state}(\gamma, t) \models \text{belief}(\text{at}(I, t'))\ \Rightarrow\ \text{state}(\gamma, t') \models \text{world\_fact}(I)$

Belief persistence In any state, if a fact is believed, it will be believed at any later time point, unless its negation is believed at that time point.

$\forall t, t' \geq t\ [\ \text{state}(\gamma, t) \models \text{belief}(I)\ \ \&\ \text{not state}(\gamma, t') \models \text{belief}(\text{not}(I)) \Rightarrow\ \text{state}(\gamma, t') \models \text{belief}(I)\ ]$

$\forall t, t' \geq t\ \ [\ \text{state}(\gamma, t) \models \text{belief}(\text{not}(I))\ \ \&\ \text{not state}(\gamma, t') \models \text{belief}(I) \Rightarrow\ \text{state}(\gamma, t') \models \text{belief}(\text{not}(I))\ ]$

Belief completeness For any state, any fact is believed or its negation is believed.

$\text{state}(\gamma, t) \models \text{belief}(I)\ \ |\ \ \text{state}(\gamma, t) \models \text{belief}(\text{not}(I))$

Belief coverage In any state, any true world fact is believed.

$\text{state}(\gamma, t) \models \text{world\_fact}(I) \Rightarrow\ \text{state}(\gamma, t) \models \text{belief}(I)$

In the general form, where a universal quantifier is assumed over I, belief completeness and belief coverage will usually not hold. However, it may hold for a specific class of information I. For example, sometimes it is assumed that the agent has complete beliefs about leads to relationships.

6.2 Properties of leads to relationships

The leads_to_after relationship expresses the conceptual core of a wide class of dynamic modelling concepts that occur in the literature in different contexts and under different names; see also [10]. Examples of such dynamical modelling concepts are, computational numerical modelling by difference or differential equations, qualitative dynamic modelling, causal relationships, temporal logic specifications, rule-based representations, Petri net representations, transition systems and finite automata. Often, either explicitly or implicitly the general assumption is made that when facts are true in the world, the facts to which they lead are also true in the world. This property is expressed as follows, also formulated by contraposition into a logically equivalent one:

Positive forward correctness If a world fact I holds in a state and it leads to another world fact J after duration D, then in the state after duration D this J will hold
state(γ, t) |= world_fact(I) & state(γ, t) |= world_fact(leads_to_after(I, J, D)) ⇒ state(γ, t+D) |= world_fact(J)
Negative backward correctness If a world fact J does not hold in a state and another world fact I leads to J after duration D, then in the state before duration D this I will not hold
state(γ, t) |= world_fact(not(J)) & state(γ, t) |= world_fact(leads_to_after(I, J, D)) ⇒
state(γ, t-D) |= world_fact(not(I))

Sometimes, also the more specific assumption is made that a world fact can be true *only* when a world fact preceding it via a leads to relation is true. This assumption can be seen as a temporal variant of a Closed World Assumption.

Negative forward correctness (single source) If a world fact I doers not hold in a state and it leads to another world fact J after duration D, then in the state after duration D this J will not hold
state(γ, t) |= world_fact(not(I)) & state(γ, t) |= world_fact(leads_to_after(I, J, D)) ⇒
state(γ, t+D) |= world_fact(not(J))
Positive backward correctness (single source) If a world fact J holds in a state and another world fact I leads to J after duration D, then in the state before duration D this I will hold
state(γ, t) |= world_fact(J) & state(γ, t) |= world_fact(leads_to_after(I, J, D)) ⇒ state(γ, t-D) |= world_fact(I)

The latter property can be formulated by contraposition into a logically equivalent property of the former one. These properties play a role in abductive reasoning methods, and automated explanation generation (in particular for why-explanations: answers on questions such as 'Why does J hold?'). The latter two properties may not be fulfilled in cases that two (or multiple) non-equivalent world facts I1 and I2 exist that each lead to a world fact J. If I1 holds, and it leads to the truth of J, then it may well be the case that I2 never was true. A more complete property to cover such cases is the following.

Negative forward correctness (multiple sources) If for a world fact J, for every world fact I which leads to J after a duration D it does not hold in the state before duration D, then in the state after duration D this J will not hold
∀I, D [state(γ, t-D) |= world_fact(leads_to_after(I, J, D)) ⇒ state(γ, t-D) |= world_fact(not(I))]
⇒ state(γ, t) |= world_fact(not(J))
Positive backward correctness (multiple sources) If a world fact J holds in a state, then there exists a world fact I which leads to J after a duration D which holds in the state before duration D.
state(γ, t) |= world_fact(J)
⇒ ∃I, D [state(γ, t-D) |= world_fact(leads_to_after(I, J, D)) & state(γ, t-D) |= world_fact(I)]

To obtain a logical foundation for a temporal variant of the Closed World Assumption in such situations in the context of executable temporal logic, in [13] the notion of temporal completion was introduced, as a temporal variant of Clark's completion in logic programming.

7 Formal Analysis of Dynamic Properties

This section shows how it can be verified that the reasoning methods introduced in Section 3 and 4 (and simulation traces generated on the basis of these methods) satisfy certain basic properties as introduced in Section 6. This is done by establishing logical (inter-level) relationships between a *global property* (GP) of reasoning methods on the one hand, and the basic reasoning steps (or *local properties*, LP's) on the other hand, in such a way that the combination of reasoning steps (logically) entails the global property. In order to establish such inter-level relationships, also certain *intermediate properties* (IP's) are constructed, which can be used as intermediate steps in the proof. Here, the focus is on one particular property from Section 6, namely the Belief Correctness property. This global property for belief generation is expressed below in GP1 and states that all beliefs should be correct. This should hold for all reasoning intervals within the trace (i.e. starting at an observation interval, and the reasoning period thereafter without new observation input). Note that all variables γ that are not explicitly declared are assumed to be universally quantified. Moreover, E is assumed to be the duration of a reasoning step.

GP1 (Belief Correctness)

For all time points t1 and t2 later than t1 whereby at t1 a observations are observed, and between t1 and t2 no new observations are received, GP1(t1, t2) holds.

$$\begin{array}{l}
\text{GP1} \equiv \\
\forall t1,\ t2 \geq t1 \\
[\text{state}(\gamma, t1) \models \text{observation_interval}\ \& \\
\neg\text{state}(\gamma, t2) \models \text{observation_interval}\ \& \\
\forall t' < t2\ \&\ t' > t1\ [\text{state}(\gamma, t2) \models \neg\text{observation_interval}]\ ] \\
\Rightarrow \text{GP1}(t1, t2)
\end{array}$$

The specification of the global property for an interval is expressed below.

GP1(t1, t2) (Belief Correctness from t1 to t2)

Everything that is believed to hold at T at time point t' between t1 and t2, indeed holds at that time point T.

$$\begin{array}{l}
\text{GP1}(t1, t2) \equiv \\
\forall I, T, t' \geq t1\ \&\ t' \leq t\ \text{state}(\gamma, t') \models \text{belief}(\text{at}(I, T)) \Rightarrow \text{state}(\gamma, T) \models \text{world_fact}(I)
\end{array}$$

In order to prove that property GP1 indeed holds, a proof by means of induction is used. The basis step of this proof is specified in property LP1, whereby the beliefs during the observation interval need to be correct.

LP1(t) (Belief Correctness Induction Basis)

If time point t is part of the observation interval, then everything that at time point t is believed to hold at time point T, indeed holds at time point T.

$$\begin{array}{l}
\text{LP1}(t) \equiv \\
\text{state}(\gamma, t) \models \text{observation_interval} \Rightarrow \\
[\ \forall I,T\ \text{state}(\gamma, t) \models \text{belief}(\text{at}(I, T)) \Rightarrow \text{state}(\gamma, T) \models \text{world_fact}(I)\ ]
\end{array}$$

Furthermore, the induction step includes that if the global property holds from a time point t to the same time point, then the property should also hold between t and t + E.

IP1 (Belief Correctness Induction Step)

For all time points t, if GP1(t, t) holds, then also GP1(t, t+E) holds.

$$\begin{array}{l}
\text{IP1} \equiv \\
\forall t\ \text{GP1}(t, t) \Rightarrow \text{GP1}(t, t+E)
\end{array}$$

In order to prove that this induction step indeed holds, the following three properties are specified: IP2, LP2, and LP3. First of all, the *grounding* of the belief generation (IP2) which states that for all beliefs that have not been generated since the last observation interval, they should either have been derived by means of forward reasoning, or by means of abduction.

IP2 (Belief Generation Grounding)

For all time points t+E, if information element J is believed to hold at time point T and J was not believed during the last observation interval, then either this was derived by applying a forward leadsto rule, or by means of abduction.

IP2 ≡
∀t,t0,J,T
[state(γ, t) |= belief(at(J, T)) & last_observation_interval(t, t0) & ¬state(γ, t0) |= belief(at(J, T))
⇒ ∃I,t2, D
[state(γ, t2) |= belief(at(I, T-D)) & state(γ, t2) |= belief(leads_to_after(I, J, D)) |
state(γ, t2) |= belief(at(I, T+D)) & state(γ, t2) |= belief(leads_to_after(J, I, D))]

Property LP2 expresses the correctness of the model believed, that should correspond with the model present in the world.

LP2 (Model Representation Correctness)

For all time points t, if it is believed that I leads to J after duration D, then I indeed leads to J after duration D.

LP2 ≡
∀t,I,J,D
state(γ, t) |= belief(leads_to_after(I, J, D)) ⇒ state(γ, t) |= world_fact(leads_to_after(I, J, D))

The correctness of the derivations within the world is expressed in LP3.

LP3 (Positive Forward Correctness)

For all time points t, if information element I holds and I leads to J after duration D, then at time point t+D information element J holds.

LP3 ≡
∀t,I,J,T,D
state(γ, t) |= world_fact(I) & state(γ, t) |= world_fact(leads_to_after(I, J, D)) ⇒ state(γ, t+D) |= world_fact(J)

The final properties specified (LP4 and LP5) are used to ground property IP2. LP4 expresses that if a certain belief concerning an information element holds, and from this belief another belief concerning an information element can be derived, then this is the case at some time point t2.

LP4 (Belief Generation based on Positive Forward Simulation)

For all time points t, if information element I is believed to hold at time point T and it is believed that I leads to J after duration D, then there exists a time point t2 information element J is believed to hold at time point T+D.

LP4 ≡
∀t1,t2,I,J,T,D
state(γ, t1) |= belief(at(I, T)) & state(γ, t1) |= belief(leads_to_after(I, J, D)) ⇒ state(γ, t2) |= belief(at(J, T+D))

Property LP5 specifies how beliefs can be generated based on abduction.

LP5 (Belief Generation based on Abduction)

For all time points t, if information element J is believed to hold at time point T and it is believed that I leads to J after duration D, then there exists a time point t2 information element I is believed to hold at time point T-D.

LP4 ≡
∀t1,t2,I,J,T,D
state(γ, t1) |= belief(at(J, T)) & state(γ, t1) |= belief(leads_to_after(I, J, D)) ⇒ state(γ, t2) |= belief(at(I, T-D))

Figure 5 depicts the relations between the various properties by means of an AND tree. Here, if a certain property is connected to properties at a lower level, this indicates that the properties at the lower level together logically imply the higher level property. Note: LP4G and LP5G are the *grounding*[2] variant of LP4 and LP5 respectively, which is why they are depicted in grey.

---

[2] The grounding variant of an executable property states that there is no other property with the same consequent. For example, the grounding variant of A ⇒ B states that there is no other property with B in its consequent, thus B ⇒ A can be derived.

Figure 5 shows that global property GP1 can be related (by logical relations, as often used in mathematical proof) to a set of local properties (LPs) of the reasoning methods put forward in Section 3 and 4. Note that it is not claimed here that GP1 holds for all reasoning methods, but that it holds for those methods that satisfy the lower level properties (LP1, LP4G, LP5G, LP2, and LP3). Such inter-level relations can be useful for *diagnosis of dysfunctioning* of a reasoning process. For example, suppose for a given reasoning trace (obtained either by simulation, such as in Section 5, or by other means, e.g. based on empirical material of an existing ambient system) that the dynamic property GP1 does not hold, i.e., not all beliefs are correct. Given the AND-tree structure in Figure 5, at least one of the children nodes of GP1 will not hold, which means that either LP1 or IP1 will not hold. Suppose by further checking it is found that IP1 does not hold. Then the diagnostic process can be continued by focusing on this property. It follows that either IP2, LP2, or LP3 does not hold. This process can be continued until the cause of the error is localised.

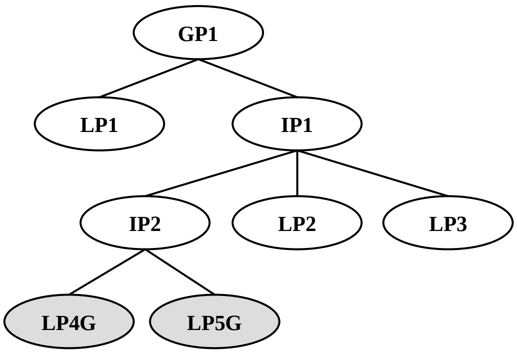


Fig. 5. Proof of GP1 depicted by means of an AND tree

The process mentioned above is based on the assumption that it is possible to (automatically) check any property against a trace. To this end, the TTL Checker Tool [5] can be used (and has indeed been used). For the traces presented in Section 5 all properties shown in Figure 5 were checked, and turned out to hold.

8 Discussion

When ambient agents need to have knowledge about human behaviours and states over time, it is useful when they possess explicitly represented causal and dynamical models about the human's processes. Once an ambient agent has such a model, a number of logical reasoning methods can be based on such a model, and formally specified as part of the agent design, as shown in this paper. The reasoning methods included cover, for example, causal and numerical simulation, qualitative reasoning and simulation, and abductive reasoning. In a number of simulation experiments example reasoning patterns were shown based on this, thus showing reusability of the ambient agent design obtained. These simulation traces have been formally analysed and verified.

In the general abductive reasoning framework, integrity constraints can be specified (see e.g. [3, 12]). Such constraints can also be specified using the approach specified in this paper, namely by incorporating these by means of the focus mechanism specified in Section 4.2. Note that the notion of a focus is not only meant to avoid integrity constraints not being satisfied, but is also meant as a way to direct the reasoning process in an appropriate and efficient way.

In [4] temporal reasoning is combined with an Active Database (ADB) for the detection of complex events in Smart Homes. The focus of that research is the combination of ADB and temporal reasoning. There is no selection mechanism in that paper as in the current work: the focus mechanism. Another example of temporal reasoning in Ambient Intelligence [19] developed a multi-agent system based on a knowledge-goal-plan (KGP) agent for transparent communication between users and an Ambient Intelligence device. They have based their reasoning model on well-known reasoning techniques such as Abductive Logic Programming and Logic Programming with Priorities. In the current work however, the focus is at developing the underlying reasoning methods that are useful in Ambient Intelligence applications.

Although the proposed reasoning methods have been applied successfully in two case studies, the examples addressed were modelled at an abstract, conceptual level. In future work, more complex and realistic case studies will be performed. In these case studies, the possibilities to incorporate the proposed reasoning methods in real artefacts in the environment will be explored. A specific question that will be addressed is to what extent the reasoning methods are able to deal with dynamic learning of new knowledge.

References

1. Aarts, E., Collier, R., van Loenen, E., and Ruyter, B. de (eds.) (2003). *Ambient Intelligence. Proc. of the First European Symposium, EUSAI 2003*. Lecture Notes in Computer Science, vol. 2875. Springer Verlag, 2003, pp. 432.
2. Aarts, E., Harwig, R., and Schuurmans, M. (2001), Ambient Intelligence. In: P. Denning (ed.), *The Invisible Future*, McGraw Hill, New York, pp. 235-250.
3. Abdennadher, S., and Christiansen, H. (2000). An Experimental CLP Platform for Integrity Constraints and Abduction, Fourth International Conference on Flexible Query Answering Systems, FQAS 2000, Warsaw, Poland.
4. Augusto, J., and Nugent, C. (2004). The use of temporal reasoning and management of complex events in smart homes. In de Mantaras, R.L., Saitta, L., eds.: *Proceedings of European Conference on Artificial Intelligence (ECAI 2004)*, IOS Press (Amsterdam, The Netherlands) p. 778–782. August, 22-27.
5. Bosse, T., Delfos, M.F., Jonker, C.M., and Treur, J. (2006). Analysis of Adaptive Dynamical Systems for Eating Regulation Disorders. *Simulation Journal (Transactions of the Society for Modelling and Simulation)*, vol. 82, 2006, pp. 159-171.
6. Bosse, T., Gerritsen, C., and Treur, J. (2007). Integration of Biological, Psychological and Social Aspects in Agent-Based Simulation of a Violent Psychopath. In: Shi, Y., Albada, G.D. van, Dongarra, J., and Sloot, P.M.A. (eds.), *Computational Science II, Proceedings*

of the Seventh International Conference on Computational Science, ICCS'07, Part II. Lecture Notes in Computer Science, vol. 4488. Springer Verlag, 2007, pp. 888-895.

7. Bosse, T., Jonker, C.M., Meij, L. van der, Sharpanskykh, A, and Treur, J. (2006). Specification and Verification of Dynamics in Cognitive Agent Models. In: Nishida, T. et al. (eds.), *Proceedings of the Sixth International Conference on Intelligent Agent Technology, IAT'06*. IEEE Computer Society Press, 2006, pp. 247-254.
8. Bosse, T., Jonker, C.M., Meij, L. van der, and Treur, J. (2007). A Language and Environment for Analysis of Dynamics by Simulation. *International Journal of Artificial Intelligence Tools*, vol. 16, 2007, pp. 435-464.
9. Bosse, T., Sharpanskykh, A., and Treur, J. (2007). Integrating Agent Models and Dynamical Systems. In: Baldoni, M., Son, T.C., Riemsdijk, M.B. van, and Winikoff, M. (eds.), Proceedings of the Fifth International Workshop on Declarative Agent Languages and Technologies, DALT'07, 2007, pp.1-16.
10. Bosse, T., and Treur, J., (2007). Higher-Order Potentialities and their Reducers: A Philosophical Foundation Unifying Dynamic Modelling Methods. In: M.M. Veloso (ed.), *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence, IJCAI'07*. AAAI Press, 2007, pp. 262-267.
11. Brazier, F.M.T., Jonker, C.M., and Treur, J. (2000). Compositional Design and Reuse of a Generic Agent Model. *Applied Artificial Intelligence Journal*, vol. 14, 2000, pp. 491-538.
12. Endriss, U., Mancarella, P., Sadri, F., Terreni, G. and Toni, F. (2004). Abductive logic programming with ciff: Implementation and applications. In Proceedings *of the Convegno Italiano di Logica Computazionale CILC-2004*. University of Parma.
13. Engelfriet, J., Jonker, C.M., and Treur, J. (2002). Compositional Verification of Multi-Agent Systems in Temporal Multi-Epistemic Logic. *Journal of Logic, Language and Information*, vol. 11, 2002, pp. 195-225.
14. Engelfriet J., and Treur J. (2000). Specification of Nonmonotonic Reasoning. *Journal of Applied Non-Classical Logics*, vol. 10, 2000, pp. 7-27.
15. Green D. J. (2005). Realtime Compliance Management Using a Wireless Realtime Pillbottle - A Report on the Pilot Study of SIMPILL. In: *Proc. of the International Conference for eHealth, Telemedicine and Health, Med-e-Tel'05*, 2005, Luxemburg.
16. Gross, J.J. (ed.) (2007). *Handbook of emotion regulation*. New York: Guilford Press.
17. Josephson, J.R., and Josephson, S.G. (eds.) (1996), Abductive Inference: Computation, Philosophy, Technology. New York: Cambridge University Press.
18. Riva, G., F. Vatalaro, F. Davide, M. Alcañiz (eds.) (2005). *Ambient Intelligence*. IOS Press, 2005.
19. Stathis, K., and Toni, F. (2004). Ambient Intelligence using KGP Agents. *Lecture Notes in Artificial Intelligence*, 3295:351–362.