Description of potential exemplars: Usability

WP leader and WP9 members

Date: 30th November 2004
<table>
<thead>
<tr>
<th><strong>IST project contract no.</strong></th>
<th>507422</th>
</tr>
</thead>
</table>
| **Project title**           | **HUMAINE**  
Human-Machine Interaction Network on Emotions |
| **Contractual date of delivery** | **November 30, 2004** |
| **Actual date of delivery**  | **November 30, 2004** |
| **Deliverable number**      | D9c    |
| **Deliverable title**       | Description of potential exemplars: Usability |
| **Type**                    | Report |
| **Number of pages**         | 63     |
| **WP contributing to the deliverable** | WP9 |
| **Task responsible**        | Kristina Höök |
| **Author(s)**               | See page 7 |
| **EC Project Officer**      | Philippe Gelin |

Address of lead author:

Kristina Höök  
Department of Computer and Systems Sciences,  
Forum 100,  
164 40 Kista,  
Sweden
Table of Contents

1 THE PLACE OF THIS REPORT WITHIN HUMAINE ...........................................5

2 BRIEF OVERVIEW OF THE WORKPACKAGE, THE EXEMPLAR PROPOSAL, AND RELEVANT RESOURCES .................................................................7

3 RATIONALE FOR THE EXEMPLAR PROPOSAL ...........................................11

4 GENERAL CONTEXT AND BACKGROUND ....................................................15

5 DESCRIPTION OF THE PROGRAM OF RESEARCH .........................................24

6 PROPOSED ALLOCATION OF TASKS .............................................................40

7 PROVISIONAL TIMELINE ..............................................................................43

8 PROPOSED USE-CASES FOR DESIGN AND EVALUATION METHODS........46

9 REFERENCES .................................................................................................53
1 The place of this report within HUMAINE

The HUMAINE Technical Annex identifies a common pattern that is followed by most of the project’s workpackages:

The measure of success will be the ability to generate a piece of work in each of the areas which exemplifies how a key problem in the area can be solved in a principled way; and which also demonstrates how work focused on that area can integrate with work focused on the other areas. We call these pieces of work exemplars. The exact form of an exemplar is not prespecified: it may be a working system, but it might also be a well-developed design, or a representational system, or a method for user-centred design. (p 4)

To that end, each thematic group will work out a proposal for common action, embodied in one or more exemplars to be built during the second half of the funding period (p.16)

The process will begin with production by each thematic group of a review of key concepts achievements and problems in its thematic area; and drawn from the review, an assessment of the key development goals in the area. This review and assessment will be circulated to the whole network for discussion and comment, aimed both at building understanding of basic issues across areas, and at identifying the choices of goal that would be most likely let the different groups achieve complementary developments. That consultation phase will provide the basis for deliverables in month 11, which describe in some detail a few alternatives that might realistically be chosen as exemplars in each area, and their linkages to issues in other thematic areas. A decision and planning period will follow, involving consultation within and between thematic areas, leading to presentations at the second plenary conference, which will describe a single exemplar that has been chosen for development in each area, and the way work on the exemplar will be divided across institutions. The remainder of the project will be absorbed in developing the chosen exemplar. (p. 21)

The review and assessment documents were delivered in May, and the consultation phase has been ongoing since, using several channels, notably e-mail exchanges moderated by the co-ordinator; meetings of workpackage leaders by teleconferencing; meetings between workpackage representatives attending the WP4 workshop and the Summer School; and a consultation meeting of WP leaders in Paris on October 29th & 30th.

This deliverable is one of the group arising from the consultation phase, whose function is defined as to ‘describe in some detail a few alternatives that might realistically be chosen as exemplars in each area’. In general, we believe that we have progressed more quickly than we have expected, and that the alternatives described here are close to the ones that should be pursued. What remains to be completed is largely detailed planning. Given the intricacy of the network, that is not a trivial task.

The following persons have contributed to the deliverable:

Kristina Höök, Jarmo Laaksolahti, Fiorella de Rosis, Brigitte Krenn, Katherine Isbister, Benoit Morel, Asimina Vasalou, Catherine Pelachaud, Markus Ballegoooy and WP9 members
The institutions that have contributed are:

KTH, OFAI, Imperial College of Science, Technology and Medicine, University of Paris 8, Università Degli Studi di Bari, T-Systems Nova GmbH, La Cantoche Production
2 Brief overview of the workpackage, the exemplar proposal, and relevant resources

While the rest of HUMAINE will produce knowledge and theories of many different aspects of emotional systems (ECA behaviours, emotion recognition, interaction principles, exemplar databases, and emotion theory) the overall goal of WP9 is to use that knowledge to produce design and evaluation methods able to deliver functioning affective end-user applications. Such applications allow users to become affectively involved in the course of interaction with the system.

Our strategy for working within an end-user application framework to evolve the usability of affective systems can be divided into two focus areas:

1. Ends: Determining qualities and criteria that demark emotional systems as usable and evoking the desired experiences for users.

2. Means: Forging process—design methods, project goals and evaluation strategies — that will steer a project towards producing a ‘successful’ affective application.

The WP9 exemplar is designed to develop, explore and determine key criteria and methods for designing and evaluating affective systems.

1.1 The field covered by the workpackage

The area covered by this workpackage is described in the Technical Annex, particularly in Section 6.2, and in more depth in the review and assessment document for the workpackage. We summarise the area here so that the deliverable can be read as a stand-alone document.

Regarding Ends, we do not believe there are simple, replicable correlations between system properties and all end users’ experiences. The experience of using an emotional system is not a property of the system itself, but rather is something that arises in the interaction between user and system. To quote Sengers and colleagues (2004):

“Rather than experience as something to be poured into passive users, we argue that users actively and individually construct meaningful human experiences around technology. They do so through a complex process of interpretation, in which users make sense of the system in the full context of their everyday experience.”

Along the same lines, Suchman has criticized the cognitivistic reductionist basis for the field of affective computing (Suchman, 2002), reducing human emotional responses to discrete, universal component parts, arising from some underlying state, and offers examples of alternative designs that put emotional interaction at the core but where: “we might differently conceptualise affective encounters at the interface: as irreducibly contingent meetings of particularly situated persons with equally particular, dynamic, and culturally inflected things.” As we discuss below, usability of emotional systems and the process of arriving at functioning emotional systems need to be
sensitive to how end-users are part in co-constructing the experience and be part of this dynamic and cultural process.

Usability traditionally focuses on goals such as effectiveness, efficiency, safety, utility, learnability, and memorability. These objective usability goals contrast with user experience goals, which cover subjective qualities such as being fun, rewarding, motivating, satisfying, enjoyable, and helpful. Usability goals and user experience goals often stand in complex relationships, involving tradeoffs such as safety vs. fun or efficiency vs. enjoyability (Preece et al., 2002). Introducing emotion thus raises many new dimensions for research on usability to address. It becomes, for instance, a serious issue whether users feel a system is ‘sympathetic’ or morally acceptable, whether it engages them emotionally. Furthermore, emotions address directly inherently adaptive faculties of humans, posing challenges for methods of user studies and artefact design.

Work Package 9 will take a primarily qualitative, situated/contextual approach to measuring Ends. A sub-group of Work Package 9 may also focus on evaluating and developing measures aimed at isolating meaningful evaluative variables and testing various components of systems against these variables.

Regarding Means, we believe that evaluation of affective systems is vital not just at the end of the design process, but as an integral part of the design process from the beginning. Having the ability to bounce early intuitions and design sketches off of real users can make key contributions to the evolution of a truly engaging end application, and may even inform the affective theory that led to the application itself.

Work in the field of HCI has made great strides toward merging design and evaluation of productivity-oriented systems (see the goals section of this document for some relevant citations); we seek to extend and adapt participatory design methods and approaches to the needs and desired outcomes of affective system projects. For example, encouraging nonverbal participation and evaluation, and using physical prototypes and fake systems operated by humans. We believe this holistic and integrated approach will not only lead to better system designs, but may also provide important ‘in situ’ insights back to affective theory makers.

Thus, it is not the primary intent of WP9 to test computational emotion theories or evaluate how particular parameters for e.g. raising an eyebrow in an ECA should work. Instead, the task of WP9 will be to produce design process and evaluation theory, methods, and measures needed to take all that knowledge into the design process and through it, to produce emotional systems that will engage with end-users to create compelling experiences.

2.1 The exemplar proposal

Following the consultation period, the exemplar proposed for WP9 is titled A Framework for Design and Evaluation of Usable Affective Interaction Applications. Similar to how HCI has benefited from user-centred design processes, our aim is to test and further develop such methods for the area of affective interaction. This in turn will hopefully push the research frontier forwards and serve as an important exemplar to others in the field. The framework entails an assortment of techniques and methods for various aspects of the design process including generating initial ideas for affective applications, refining ideas into systems/products or validating the affective interaction
loop in finished systems/products. As a whole the framework is intended to be a broad resource for designing and evaluating affective systems – representing first and foremost a user-centred perspective on affective interaction applications. The tools and methods of the framework will be applied and evaluated in a variety of situations/applications/domains. Records from the sessions will provide valuable guidance for future users of the tools and methods regarding their proper usage and expected results.

2.1.1 The elements of the exemplar

The exemplar consists of four main elements.

2.1.1.1 Criteria for usable affective interaction systems

Develop an understanding for what makes affective interaction systems successful and formulate some criteria. These criteria will not be objective, independently measurable entities, but will make sense relative to the specific application domain, aim to capture subjective experiences of the application, and foremost, be related to the designer’s intention for the application.

2.1.1.2 Evaluation metrics for criteria

Translate criteria for affective interaction systems into some metrics and suggested methods.

2.1.1.3 Existing user-centred methods for design and evaluation

Condense experiences from applying existing user-centred design methods to the design and evaluation of affective interaction systems – which methods work and which do not? Examples of such methods include body-storming (Oulasvirta et al, 2003), interaction re-labelling (Djajadiningrat, et al., 2000) and personas (Cooper, 1999).

2.1.1.4 New methods for design and evaluation

Develop new methods for capturing unique aspects of affective interaction that can be used during an iterative design-evaluate-redesign process. Methods that will be investigated include: a sensual method for non-verbal mediation of affective state, a Wizard-of-Oz environment for multimodal emotional interaction, and an extended think-aloud protocol designed to capture emotional interactions.

2.1.2 The ways in which the exemplar depends on other workpackages

WP3: Methods need to be grounded in theory. Absorb and build upon current affective theory to create appropriate evaluation instruments.

WP5: The annotated databases are expected to provide input to design processes in WP9, and will also serve as material in evaluation processes.

WP6: By working with Work Package 6 in the design of an Affective Interactive Embodied Conversational Agent system, we hope to test and refine our evaluation techniques and instruments, which will serve as our final exemplars. Moreover we plan
to apply the representation language defined in WP6 for the WoZ-environment and also adopt it to the “sensual” method.

WP8: We may start a similar collaboration as with WP6 on specific application examples where we can apply our framework for design and evaluation.

2.1.3 Proposed final output

The expected output from this WP will be on the one hand the hands-on experiences of applying the framework to specific design cases, and on the other hand, the framework itself including methods, criteria and measurements. Both will be described in reports.
3 Rationale for the exemplar proposal

The exemplar proposal represents a choice to follow a particular line of development rather than others that are possible (or might seem possible to the outsider). The key reasons for making this particular choice are as follows.

3.1 Distinctive features of the approach proposed

The aim of the exemplar is to find user-centred methods for design and evaluation that ensures usable affective interaction applications. Through applying methods that are developed to different systems in various stages of development we also aim to arrive at knowledge of what actually makes up a usable affective interaction system. Main features of the proposed approach are that

- **Usability** of affective applications is considered
- Usability is seen as an affective interaction problem (an affective loop)
- A user-centred design approach is adopted
- The whole design, evaluation, redesign cycle is addressed
- Developed designs and methods are to a great extent informed by emotion theory, but not always limited by it.

An important strategy that is adopted by the exemplar work is to start from existing methods and then look at how well they address the specific problem of getting feedback on users’ affective interactions with the system. However developing novel types of methods and tools that specifically address affective interaction problems is an equally important part.

3.2 Rationale for emphasising this approach

Affective interfaces include those that express emotions, those that attempt to understand user emotions and those that use affect as part of the system’s “mind”. User evaluations in this new field are still infrequent, and lacking an established theory providing similarly well defined targets to assess as in traditional human computer interaction. At the same time, most researchers have been focussing on the natural phenomenon of affect and its expression, communication, perception, and hypothesised processing, without being overly concerned with whether qualities such as believability or naturalness of synthetic characters do in fact contribute to the overall success of a system. Neglecting the relationship between natural source of inspiration and actual engineered systems, designers and researchers are all too easily misled into not questioning whether inter-human communication really forms the best model for interaction between human and machine. Results from natural language interfaces, adaptive interfaces, or intelligent user interfaces indicate that there exist principles and idiosyncrasies that are specific to the design of human machine interaction (Dahlbäck et al., 1993, Höök 2000). Among what little results exist today,
there already are a number of findings underlining the importance of empirical verification and questioning of intuitions and working hypotheses of the appropriateness of the deployment of affective functionalities for specific purposes, e.g. (Issroff and del Soldato, 1996).

On the other hand, the well-established goals of usability research, including effectiveness, efficiency, or safety either are probably not adequate for scenarios centred on subjective user experiences of being satisfied, rewarded, pleased or motivated; at least there is no clear understanding of what should be measured in which way so as to assess the standard usability qualities with respect to these new targets. Many other difficulties ensue from the social and biological grounding of the phenomenon to the investigated. Affective interfaces address directly inherently adaptive faculties of humans; the methods for user studies therefore have to take into account that not only test users, but also “emotionally enabled” artefacts will alter their behaviour within and across evaluation runs. As also mentioned in many other parts of this document, in different contexts, the manifestations of greater or actual interest are not those of extreme states, but rather the subtle emotions; the challenges posed by these difficult to detect and highly susceptible targets are complemented by those of rarely occurring emotional states and of those that take significant time to develop. The latter are only two examples making it rather evident that usability assessments concerning affective interfaces hardly can be sped up at will or carried out in a conclusive fashion within a limited timeframe using standard approaches. To mention just one last example, the simple circumstance that users are always emotional immediately points to the difficulties in clearly identifying affect that is due to and related to the application at hand, excluding reliably exogenous causes; conversely, the scope of usability assessments is difficult to contain when consequences outside the intended reach of the application are hard to rule out or prevent.

The situation is further complicated by the fact that studies of affective interfaces are being carried out against a background of theory and technology that is still embryonic. Appropriate utilisation of reported usability results therefore needs to be carefully considered among the alternatives of performing tuning or adaptation of the currently employed system components, altering the architectural design introducing different components that replace or complement the current ones, or initiating a new research thread defined by the intersection of open issues and indications from the usability study.

Taken jointly, these three views on usability of emotion-related systems document vividly how usability research forms a particularly well suited hub that not only invites but downright requires collaboration and exchange of information across all of the multifarious research community. The circumstance that every discipline can both contribute as well as gain from the interactions is an important facilitating aspect.

The aim of emotional systems is to get users affectively touched by the interaction with the system and even to maintain this relationship or illusion over a longer interaction session. This is, obviously, a hard and very difficult goal to achieve. We know for sure that movies, novels, television shows, arts, music, are indeed able to get people affectively involved. But we want to make end-users affectively touched by interacting with systems modelling emotions, reasoning using emotions and expressing emotions.

There has been quite some research in Europe as well as in the US on how to recognise users’ emotional states through singular, one-off, readings of biosensor data, facial
expressions, body posture, interaction with devices, such as mouse or keypad, or props, such as plush toys, but repeatedly there seems to be the same conclusion: while this may recognise some basic emotions (fear, stress, and arousal), they fail to get the whole picture and oftentimes contradictory results between users’ self-reports of what they think and feel and their physical expressions arise (e.g. Höök et al., 1999). They also fail to understand any more complex and interesting emotional states that users might be in – such as shame, guilt, positive arousal, or flow.

**Affective interaction as an interaction problem**

In WP9 our aim is instead to see usability of emotional systems as an interaction problem that concerns the whole interaction cycle and where emotions arise from an active act of interpretation and participation from the end-user side. We can compare this to, for example, how an empathic successful teacher works in a classroom (Cooper et al., 2000). An affective teacher will not only recognise the affective state of the pupil, but also rather try to influence the pupil’s state through communicating the state that is most desirable in order to solve a problem. An emphatic teacher will illustrate that a math problem is difficult and requires lots of attention through a strong physical facial expression showing of effort. The pupil imitates this expression – imitation is strong, innate, ability – which in turn induces the right kind of starting point for working hard. The imitation of the facial expression together with the cognitive understanding that the problem is hard achieves the right state for working hard. This is an interaction between the pupil and the teacher. It is not an attempt at recognising the pupils state solely. They interact and the goal is to achieve the desired state – to learn math and solve the problem at hand – not to recognise the pupil’s initial state.

The other half of the problem of the affective loop lies in producing the right output from the system to maintain end-users involvement, and to do this over a whole interaction. The proposal pf WP9 is to tackle this part through the user-centred approach to design. The field of affective computing has mainly been driven by AI (artificial intelligence) researchers with little or no knowledge of how to work with end-user applications. The field of HCI (Human Computer Interaction) on the other hand, have addressed this problem for a couple of decades. Many important lessons can be learnt from their work. One such lesson is that human-computer interaction is not, and cannot be, modelled from human-human interaction. A computer system is a designed artefact – not a “natural” thing. While the field of HCI certainly recognises that there are design considerations that should be built from knowledge of human abilities and limitations (see e.g. concepts as affordances by Norman, 1990), they also recognise that computers are part of human culture, and thus subject to change. Over and over, artefacts are designed that users then take into use in ways that are quite different from what the designer expected (Suchman, 1987, 1997). A design process that fails to involve end-users in the design loop, will fail to recognise the particular quirks and problems of how to design these artefacts.

A main part of WP9 will aim to take the lessons learnt and methods and theories of HCI into account in the design of emotional systems. Through involving users in the design process at several stages of the development of a showcase, the aim is to find ways to ensure the development of emotional systems that users find usable and able to create experiences that are meaningful and engaging to them.
3.3 Rationale for subdividing the task

Design and evaluation is a broad area, participants have differing interests and backgrounds, different use-cases require different methods, resources, and study angles. We are interested in the whole design-evaluation process (see Figure 1) and the subtasks of WP9 address different aspects of the process. The exemplar can still be viewed as a single piece though, since it explores one and the same idea: that a user-centred perspective will help focus the design process and create final application systems that involve users affectively. This exemplar will be a touchstone for integrating this perspective throughout the whole life-cycle of a system development process. While we strongly believe that this is the best kind of exemplar for WP9, there is still a lot of white spaces on the map. We do not know how to capture subjective experiences and affective involvement in such a way that it can provide feedback into the design process. A user-centred design perspective and development philosophy is most probably a good path to explore, but little is, as of yet, known in what to design for (flow, pleasure, fun, experiences, excitement, fear), and how to best make users involved.

![User-centred design framework](image)

Figure 1 User-centred design framework

3.4 Measures taken to ensure coherence across subtasks

Positioning the subtasks within the user-centred design framework above ensures that subtasks are working towards a common goal. The nature of the toolbox with its different methods and tools all sharing the same conceptual framework also helps in this respect. The four subtasks follow on one-another. Once criteria for the success of a specific application project are established, there has to be corresponding measurements to know whether the design goals are met according to the same criteria. The process to ensure such a design and evaluation cycle in turn needs methods. All these tasks together will help shape the framework produced by this WP.
4 General Context and Background

Ends:

- Existing quantitative measures of user responses include: Likert scale response sets given in questionnaire form (e.g. those used in the research of Reeves and Nass, 1996), quantitative coding of video-taped data with inter-coder reliability checks (e.g. work like that referenced in Work Package 5), and the use of biometric sensor data (e.g. brain waves, skin conductance, facial muscle contraction, heart rate) to measure affective response to systems (e.g. Hazlett, 2003, and work like that referenced in Work Package 4). These methods usually involve brief, single-episode interactions with prototypes or systems. The use of video coding and bio-sensors reflects the understanding that affective states may be difficult for users to self-evaluate using traditional questionnaire-style measures.

- Existing qualitative measures include: qualitative/ethnographic summaries of use contexts (e.g. Beyer and Holzblatt, 1999); summaries of user preferences for product qualities based on observational settings such as focus groups, and/or user engagement with prototype systems (e.g. Jordan, 2001); thick description of user engagement with prototypes and systems, based upon observation of interactions (e.g. Höök, Sengers, and Andersson, 2003); collations of verbal comments made during open-ended question sessions (Ebling and John, 2000).

Means:

- Classical controlled experiment design has been used with fully functioning systems to isolate and analyze specific system variables which impact the user’s affective response, such as the use of flattery or humor (e.g. Reeves and Nass, 1996; Morkes, Kernal, and Nass, 1998); or to test the application of emotion theory to system design (Zhou and Conati, 2003).

- The HCI community has engaged in quantitative evaluation of fully functioning systems through controlled observation of user interaction, typically using questionnaires and occasionally video-coding, to assess user responses (see Helander, 1988 and Preece et al., 1994 for an overview of current techniques). There is also application of post-use surveys (e.g. Mentis and Gay, 2003).

- There is an emerging movement within the HCI community toward more qualitative evaluation of user affect and engagement (e.g. Höök, Sengers, and Andersson, 2003).

- Participatory design can help shape affectively successful systems through engagement of end users early in the process (see Muller and Kuhn, 1993, for a description of this process and types of interventions).

General Context and Background:

From previous studies on usability of emotional systems, some lessons have been learnt. While we do not want to claim to have an exhaustive literature survey here, we are providing some insights into the kinds of problems that have been encountered. We divide our survey into one part concerning ECAs, since most work so far has focused on ECA evaluation (ECA
Research Summary) and one part on other kinds of interactive emotional systems and strategies (General Affective System Research Summary), for example: development and assessment of affective objects, music, gestures; recognising affect from biosensors or other tangible interaction devices; recognition of affect from speech input; and affective interactive narratives.

**ECA Research Summary**

The following usage of interface agents, all well represented in literature and existing systems, are possible domains for agents that possess emotions.

- **Agents for help and learning.** Agents can ‘sit at the side’ of an application, providing help and guidance on its usage (e.g. Horvitz et al. 1998). Agents can also act as tutors or co-learners in a learning application (e.g. Lester et al. 1997). Possessing emotions or exhibiting emotions might make these agents act more timely (not disturb) and motivating.

- **Delegation.** Users can delegate tasks to agents, such as activities that should be done when the user is away from the computer, monitoring events from other sources, or performing activities at distant locations in the network (e.g. Lieberman 1995). In these situations, the agents may use emotions to direct their efforts to the most important goals/paths to follow.

- **The subjective focus.** Agents provide an anchor for subjective evaluations that follow a presentation (e.g. who it is that gives a helpful suggestion) (e.g. Schank, 1991). An agent that possesses emotions and personality will be able to provide a more interesting, compelling, subjective focus.

- **The dialogue partner.** Agents provide a counterpart in natural language dialogue (e.g. Bretan 1995, Beskow et al. 1997). Again, expressing emotions, personality, and attitude, for example through affective speech, might contribute to the naturalness of the dialogue, but it might also raise too high expectations of anthropomorphism – thus a careful balance between expression and expectation needs to be employed.

- **Emotional behaviour.** Agents can both show emotions and arouse emotions in the user for general entertainment purposes (e.g Elliott, 1997). It can be in MUDs, interactive drama, games, or in advertising and branding. It can also be used for the user to explore emotions.

- **Agents as user representatives.** Agents can be instructed to behave in certain ways and then go out into a world and represent us – avatars are simple examples of this (e.g. Vilhjálmssson and Cassell, 1998). Emotion modelling will contribute to creating avatar behaviours that need not be consciously controlled by the user. Instead the agent can take on the responsibility for generating the right facial expression and body behaviour at the right time.

**Studies of interactive characters**

Interface characters have been much criticised and debated in the HCI community (Shneiderman, 1997, Lanier, 1996, Suchman, 1997). They are said to violate good usability principles, to obscure the line of responsibility between human and machine, and to confuse both designers’ and users’ understanding of the computer’s abilities and inner models of events. The proponents, on the other hand, regard these parameters as opportunities rather than reasons to avoid characters in the interface (Höök, 2000, Waern and Höök, 2000).
Motivational effects

A number of studies have examined the ways in which characters enhance engagement and encourages exploration of a given information space, mostly in relation to learning and creativity. Such motivational effects were studied by van Mulken and colleagues (1998), who compared two versions of the presentation system PPP Persona: one with and one without a character. The study showed effects neither on recall of the presentation, nor on how the presentation was understood (objective measurements). However, it revealed a positive effect on the subjective estimation of whether the explanation was difficult or not. Subjects experienced the explanation as simpler with the PPP Persona character than without it. van Mulken and colleagues named this ‘the persona effect’.

Another similar study looked at the persona effect for ‘Herman the Bug’, a pedagogical agent that helps students to create an ecological micro world system with plants, light and air (Lester et al., 1997). Here five different clones of the agent were compared, and the study revealed a persona effect – a strong positive effect on the students’ perception of their learning experience. The animated character also had an effect on learning.

In a study by Wright et al. (1998) a plain textual explanation of a medicine was compared to one with the same text but with an animated dragon illustrating the different threats to the blood system. Here a negative effect on how much was remembered afterwards appeared; the dragon disturbed subjects, rather than aided them.

These conflicting results (PPP Persona and Herman the Bug, versus the dragon studies), point to the need for a better understanding of the design of synthetic characters in order to make use of their potential to encourage learning and exploration, and at the same time avoid the scenario in which the character distracts and disturbs the learning process. This involves, we think, a better understanding of the features of and relationship between wayfinding and exploration activities.

As pointed out by Andrew Stern (Hayes-Roth et al. 1998) (designer of the Catz and Dogz system) the artistic design and practical understanding of the creating of synthetic characters is crucial in determining the success of a system. A similar point is made by Elliott and Brzeinski (1998) when they cite Lester et al. (1997):

“Lester gives the examples of, on the one hand, a humorous, lifelike, joke-cracking, character that ultimately impedes problem solving through his distracting presence; and on the other, a dull assistant that always operates appropriately but yet fails to engage the student. When communications from an agent must be coordinated to be both engaging and purposeful issues in timing, and the multi-layering of actions arise.”

Anthropomorphic effects and believability

Another effect of synthetic characters is the ways in which they tend to raise expectations of anthropomorphism of the system (Reeves and Nass, 1996). Such anthropomorphic effects seem to have many dimensions. On the one hand the user may expect the system to be intelligent and cognitively potent. Brennan and Ohaeri, (1994) showed that users talked more to the anthropomorphic interface. King and Ohya, (1995) showed that users attributed more intelligence to anthropomorphic interfaces. Koda and Maes, (1996) showed that realistic faces are liked and rated as more intelligent than abstract faces. Opponents of synthetic characters argue that raised anthropomorphic expectations may lead to frustration in the user when the system cannot meet the expectations (Schneiderman, 1997). For instance, the presence of a
talking face might influence the user to expect the system to possess natural language and dialogue competence, which no system of today can live up to. The general conclusion is that the more ‘natural’ the interface, the higher expectations on intelligence in the system.

Believability and Emotions

Besides anthropomorphism, which refers to any human-like characteristics in an interface that makes the user think that the system has some form of intentionality and human-like capacity for reasoning, the issue of character believability is sometimes mentioned in the literature. Exactly what is meant by believability is somewhat unclear. Some use the concept to refer to the facial expressions and body language of the character – the idea is that the more human-like and naturalistic, the more believable. Our view is somewhat broader, including the personality and attitude of the character, perhaps less focused on the bodily expressions. To our knowledge there has been no studies focusing on believability, in this sense, in the field of interface characters.

At its most basic, believability involves the assumption that not all users will approach intelligent agents with the explicit aim to pick them apart to check their intelligence level. Instead they might be willing to indulge in an experience with the interactive character where they are amused, intrigued, and stimulated. It is therefore the designer’s task to make the character behave in such a manner that the illusion lasts – at least until the aim of it has been fulfilled. By aim, we mean anything from a task-oriented goal to more entertainment applications.

In making the illusion last, we can be inspired by many things; animated movie characters is one source of inspiration, literature another, but one important source is of course how human beings behave. Many researchers have therefore turned to the psychological literature on how human cognition and emotion works, as described earlier in this deliverable. But an important issue for believable agents is whether human emotions need to be replicated in this exact way.

As pointed out by Damasio, Descartes made an error when he divided human cognition into body and soul where the soul is some magical way conveys its intentions into the body. Instead, cognition is embedded in the body, in the brain, in the nervous system, in a feedback loop between internal and external events and bodily behaviours, and cannot be separated from it. When we attempt to build systems that either 1) imitate this or 2) tries to “fool” users into believing that the systems has these kinds of properties, we are of course lacking the human body. But it is only in case 1 that we really need it. In case 2 we can equally well make simplistic models that act as if there had been a body, a body language, a social context, an inner life, etc. Whether it is possible to go through with 1 is a matter of philosophical distinctions: if we (spending years and years) build something that in every respect models human behaviour (including emotions, social behaviour, etc.) but that is realised in “hardware” instead of “wetware” can we really claim that it possess intelligence? Or must there be a subjective, conscious, self in order for this to happen? For our purpose here, we can leave this issue aside since we are aiming to establish how we can evaluate and assess interface characters. On the surface we might assume that there is not going to be huge differences between characters that fake emotions and cognition, and characters that do indeed possess complete cognitive and emotion models.

In case 2, when we deceive the user to believe that the created character/agent/system they interact with does indeed have emotions and intelligence, the problem is different. In that case, simple models with partly ready-made stories or simple mechanisms to couple external events to emotional expressions might be enough. Sometimes people do like to be deceived: they will
go to see a movie, they will read a novel or a comic strip, or watch soap operas on TV – just to be deceived for a short while to believe that these characters do exist.

**Active users**

Even if synthetic personal character assistants are intentional interfaces that are ‘deceptive’ in some way, the role of the user is not a passive one. The user will interact with the character and infer a lot of the characteristics of it: its personality, intelligence, or emotional state. S/he will also fill in the gaps in the story/dialogue that develops between user and system. Here we can be inspired by theories from the humanities on narrative construction. One of the basic presumptions within narrative theory in general, and discourse psychology in particular, is the notion of a reader/spectator constantly striving for *coherence* in his or her understanding and experience of a given text (e.g. Bordwell, 1985). Coherence is, on this account, *accomplished* or *constructed* by the reader through a huge battery of tacit, and hence non-conscious, everyday assumptions or knowledge about the perceptual, physical and socio-cultural world, influenced of course by pragmatic parameters (e.g. the purpose of the reading).

In the entertainment business when, for example, creating animated characters, it is well-understood that not everything should be explicitly stated but that the spectator must be made involved and curious about the next turn in the story. A basic assumption here is that characters – whether they appear in computers, fiction or cinema – are not only made sense of through the features of face and body, but through the ways in which character act within situations, display humour and share values and moral perspective of the user (Tan, 1996). It is, for instance, not enough to place characters in a film in order to evoke emotions (and thereby commercial success). These characters must move, act, feel, think and talk in concrete situations in ways that raise the spectator’s sympathy or antipathy. As pointed out by Porter and Susman (2000) when discussing how to create life-like characters in cartoons (‘Toy Story’):

“They (animators) understand that ‘life-like’ does not mean ‘has movement’; lifelike means ‘has a brain’. The underlying notion of Pixar and Disney animation is that action is driven by the character’s cognitive processes – that it reflects intelligence, personality and emotions.”

These dimensions, however, are much more difficult to design for and evaluate.

**Interactive Characters in Computer Games**

Current trends in computer games suggest that in the future, they will become populated, not just by people, but also by (possibly large numbers of) autonomous and intelligent agents and characters. In this context, such agent’s behaviours and expressions will be of particular importance for inducing to the user a sense of presence and lead to the illusion conveyed by the game. Classical animators are masters at conveying intentionality and emotions through the physical movements of their characters. In The Illusion of Life, Thomas and Johnston (Thomas *et al.*, 1981) argue that simple changes in bodily movement can convey radically different emotions to human observers.

Similarly, graphically represented autonomous agents could express a great deal of emotion and personality based on the way in which they move their bodies -- if it were not for the fact that these highly expressive bodily movements are surprisingly difficult to design and to control in a way that is dynamically responsive to user activity. Expressive actions of synthetic characters in an interesting research problem for affective system design which has only begun to be explored.
Design problems

Crucial design problems that needs to be addressed in this area are the following:

- Agents need to display behaviour and affective expressions in such a manner that the user understands them. This means that they cannot always act in the most efficient rational way but instead they might have to act in ways that conveys to the user what is going on (Sengers 1998). This also holds for agents that work in multi-agent systems but where their results have to be communicated to a user in the end.

- Agents need to be timely. When an emotion is displayed to the user it has to come at the right point in time, and last for an appropriate length (Ruttkay et al. 2000). If an affective response from the user is the aim, then the interaction has to be carefully paced so that the user can follow it without being bored or puzzled.

- Agents sometimes need to have interesting personalities. Only then will their emotional behaviour be comprehensible and interesting to the user. Conveying the personality might be difficult if the interaction with the user is limited. This is where idle behaviour or interaction between several agents can come into play. When several agents interact, they can take the opportunity to show more of their personality traits.

- For some affective agent situations, it is necessary to create a narrative context (a situation, an interaction history) in order to understand the emotional behaviour (Isbister and Doyle, 2003).

- If the agent is used for a longer time span with a user, different personalities and attitudes might be needed in order to fit the needs of different users (Boyce 2000, Nass and Gong, 2000).

Evaluation of ECAs

When designing an Embodied Conversational Agent to act as interface with the user, one should answer a few questions. First of all, is the ECA the best interface? (Where ‘best’ needs to be defined in the context of the interface: the most convenient, the most natural, the easiest way to interact with the system, the most enjoyable, etc?) If the interface does require the use of ECA, how to design it? Should it be 2D, 3D? Should it be a realistic or a cartoon-like model? What should be its gender and its age, a young male or an older woman? Should it have a humanoid aspect or not? Should it be a puppy or a famous cartoon character? Should the agent look more like a French or like a Japanese? Once details on the geometry of the agent has been decided for a given application targeted toward given users’ profiles, once has to wander not only how the agent should communicate information, but also how should the agent interact with the user(s)? One has to ponder if the agent would be extrovert or introvert? Would the agent be emphatic, or expressionless? How should the agent communicate, with which modality: with the face only, the gaze, gesture? Should the agent be sparing in its communication style and use one modality to communicate an information; or should it be very redundant and use as much modalities as possible when talking? Does redundancy affect the style of the agent, or does it also affect the interaction quality, and if so along which dimensions (learning, maintaining attention, attracting attention...)? Which role should play the agent in the interaction? What should be its goals? Should it persuade the user, or simply provide information, or may be show empathy? From this long list of questions one can see the diversity of choices that need to be made in order to design an ECA. Evaluation of these choices and on their effect on the design of ECA is a necessity.
Evaluation should happen at different levels: the micro level in which one aspect of the agent is observed; the user level in which the reaction of the user is evaluated; the application level in which the agent within an application is studied.

Evaluation of agents and of interactive systems using agents is still a relatively new area. No benchmarks or standard evaluation methodologies exist, though several attempts to remedy such a lack have happened (Ruttkay et al., 2004; Christoph, 2004; Catrambone et al., 2004; Isbister and Doyle, 2004). In particular, Ruttkay et al. (2004) propose a common definition of terms that are largely used in the agent community, but with different meanings associated with them. They propose a taxonomy of the various dimensions upon which evaluation should be made (e.g., agent’s appearance, function, believability). On the other hand, Christoph (2004) proposes a methodology to evaluate user’s interaction with the system. The author proposes ways to collect data and analyze them, to select the appropriate subjects of a study as well as to define the research strategy in order to set up properly the evaluation studies. She bases her guidelines on evaluation study from empirical research.

Evaluations at the micro-level of the design of an ECA involve the evaluation of aspects such as the muscular computational model of the face (muscle contraction, skin model, wrinkle), the lip model, and/or emotion model. Several muscular and skin models have been developed (Lee, 1995; Wu, 1995; Guenter, 1998; Pighin, 1998). Some are physically-based and propose an accurate model of the muscle and the skin; while others offer a rendering-based method which give realistic visual results. Researchers have also developed computation models for lip shape and co-articulation effects (Cosi, 2002; Bevacqua and Pelachaud, 2003; Cohen, 1993; Beskow, 1995; Reveret et al., 2000; LeGoff, 1997). Evaluation methodologies have been proposed by Massaro (1997). Computation models are evaluated at the syllable level, the word level, or even the sentence level (Massaro, 1997; Siciliano et al., 2003). Comparisons between models can be done using an objective evaluation: for example by comparing between real targets of lip parameters and predicted parameters of the computational model. On the other hand, intelligibility evaluation works at the perception level: by performing perceptual tests on subjects. Studies have shown (Beskow, Ph.D. thesis) that results using objective evaluation and intelligibility evaluation may not be correlated. Indeed models with hyper-articulation give better perceptual results even though it shows much less natural movements.

Evaluation at the micro level looks also at behavior along one channel. For example, Krahmer and Swerts (2004) found that eyebrow movement on pitch accent does not play the same role for Dutch language as for Italian language. Gaze behavior has also been studied (Kraemer et al., 2004; Garau et al., 2003). This behavior has a great impact on the way the user perceives the agent. Buisine and Martin (2004) investigate if the strategies of using multimodalities in communicating have effect on the perception of ECAs as well as on the learning dimension. The authors designed a study in which three different looking agents use just one modality (say gaze OR gesture) to communicate an information or use redundant behaviors (say gaze AND gesture) to communicate. Moreover Cassell and Thorisson (1999) found that users prefer interacting with agents communicating multi-modally: users found the agent more natural if this one shows gaze, gesture and facial behaviors. The study reported below looks at how verbal and nonverbal behaviors may persuade users or not.

Other evaluation studies look more particularly at the user standpoint. Höök’s research is a good example of such studies (Höök et al., 2000; Höök et al., 2003; Höök, 2004). Höök works at the user’s perception level, investigating how agent design contributes to the users’ experience of the interface. Using several applications, she examines how role, behavior, speech, and emotions of agents act on the users’ emotions. She argues that developing an
agent should be done in a loop of design and evaluation; that is to say, the design of an agent should be interwoven with evaluation studies looking at the effect the agent has on the user’s experience. Nass and his colleagues have performed important studies on different aspects of agent’s design: culture (Lee and Nass, 1998), personality (Isbister and Nass), and appearance (Nass et al., 2004). They based all their research on psychological studies.

Given an application, the impact agents may have on the user has been reported (McBreen et al., 2001; Morton et al., 2004; Krenn et al., 2004; Darves and Oviatt, 2004). Studies have also looked at the effect an agent’s clothing (McBreen et al., 2001), intonation (Darves and Oviatt, 2004), and gender (Buisine et al., 2004) may have on dimensions such as trust.

**General Affective System Research Summary**

Some work has also been done to create systems that incorporate affective components without explicitly evoking anthropomorphic relationships with ECAs. Rosalind Picard and her research team at MIT’s Media Lab have focused on both reading and generating affective cues in a wide range of projects (Picard, 1997). The Influencing Machine is an art-based exploration of how users will interpret and engage with a machine’s internal affective model (Sengers et al., 2002, Höök et al., 2003).

Work has also been done to incorporate non-mouse-based interaction with systems, towards richer affective engagement. SenToy — a doll-controlled game (Höök et al., 2003, Paiva et al., 2003), and eMoto — a gesture-controlled device for colouring mobile phone messages with emotional cues (Fagerberg et al., 2003) are two examples of this approach. Cowie’s work on analyzing affective cues from speech (Cowie and Cornelius, 2003) should help further efforts to allow use of one of the most natural affective ways of interacting with a system: human speech.

Non-character-based emotion synthesis and analysis research has also been carried out in the study of affect and music. Bresin and his colleagues (e.g. Bresin and Friberg, 2000) have isolated and parameterized dimensions of musical performance that have been incorporated into a system which can generate performances of the same piece of music with very different affective qualities.

Finally, work has been done to bring techniques from film and other narrative media to enhance affective engagement with interactive systems (Laaksolahti et al. 2003).

**Novel types of affective systems: affective communication media**

Apart from ECAs there are many attempts to produce alternative designs, such as particular artefacts mediating presence and awareness of others to which you have a close relationship (Gaver, 2002). There are also attempts to find novel ways of expressing emotional communication through interactive media. Besides the smilies and variants of smilies, such as emoticons, we can distinguish between three classes of solutions: communicating through specifically designed artefacts, communicating through composing multimedia messages (MMS, Instant messaging with photos and emoticons, etc), and communication of affect through adding colours or animations to textual media of different kinds.

The first set of solutions, the *designed communicating artefacts*, typically involves equipping two people who have a close relationship with some kind of artefacts that can communicate wirelessly (e.g. van der Hoog et al., 2004, Strong and Gaver, 1996, Brave and Dahley, 1997, Goodman and Misilim, 2003). These solutions can often communicate presence or indicate
that the other party is showing some kind of concern. In *kiss communicator* designed by Heather Martin and Duncan Kerr of IDEO (Barley 1999) a kiss is blown into an oval shaped product, which in turn creates a red, blinking light in the person’s device. With *interactive stones* created by the interactive institute in Sweden, concern for the other is shown by holding the stone in your hand, which in turn makes the other person’s stone warm, thereby communicating affect. These solutions have obvious constraints when it comes to communicating different and complex emotions. They are also dependent on specific devices for the communication. But what is interesting about them, is the simplicity of expression and the use of physical interaction means.

Our second category of communicating affect, is through various multimedia possibilities. Many communication media allows for emoticons, photographs, avatar expressions or other ways of communicating affect. Here the emotional expression is entirely within the hands of the user. They choose which smiley, emoticon or avatar expression to pick, and it is subsequently shown to the other party. Mel Slater and colleagues (Slater et al. 2000) have designed a system where the user expresses her emotions by changing characteristics of a drawn face. The user can influence the eye brows and the mouth of the face. ExMS is another system where the user explicitly states her emotions (Persson, 2003). ExMS is an avatar-based messaging system where users can create short pieces of animated film to send to each other.

Visualisation of affect, through colour, shape and animation, has been used in several systems. For example, colour coding has been used to show the affective structure of a text document (Liu et al., 2003). The aim of that system was not really to create a media for emotional communication between people, but rather as a way of structuring an existing text from an emotional perspective. A combination of shape and animation of text in a message to convey emotional content is used in the Kinedit system (Forlizzi et al, 2002).

Apart from adding colour and animations, some have also attempted to add haptics to the multimedia communication channels, see e.g the system ContactIM ([http://www.mle.ie/~ian/contact/default.shtml](http://www.mle.ie/~ian/contact/default.shtml)) from the Palpable Machines Group at Medialab Europe.

Unfortunately, very few of these emotional communication systems have been evaluated from an affective involvement point of view. In many cases, there is only mentioning of early end-user study results, such as for the system the Hug (Gemperle et al., 2003).
5 Description of the program of research

5.1 Criteria for usable affective interaction systems

Designs for affective interaction systems may take on any number of forms and be intended for any number of uses. Hence, similarly to design in general, there are seldom any commonly agreed upon criteria for what constitutes a good or successful affective interaction system. Instead success is often determined by the degree to which a system achieves the designer’s intentions. For instance, it is wrong to say that creating a mood of “well being” is a general usability criterion for affective interaction systems. Sometimes, in a game for instance, the purpose may be to make the user afraid or even revolted. If the game fails in doing so the design of the game has been unsuccessful and the system may be considered a failure. Despite the application and domain specific nature of many criteria there may also be some general criteria, or properties, that make interactions more or less likely to pull users into an affective loop, for instance flow and ambiguity.

Csikszentmihalyi (Csikszentmihalyi 1996) has defined flow as the state people get in when they become so involved in doing something that they lose control of time and space. To get to a state of flow people need to feel that something is complicated enough for them to feel proud of themselves when they make achievements and get to enter a new level. To reach a state of flow it is important that there always are new levels to reach. When people feel that they can master it all they will loose interest. However, they should never feel that it is impossible for them to make improvements or to get anywhere at all. That will have the opposite affect to flow. That will make them feel stupid and incapable and they will instead loose interest for that reason.

Csikszentmihalyi for example uses climbing, computer programming and gaming as good examples that can take people to a state of flow. He also mentions reading as something that can have the same effect on people. Reading might sound like a totally different task but a good book can keep people’s attention for hours. As long as people do not feel that they have understood it all they want to continue reading. If the book is too hard for them or if it is too easy for them to see how it is going to end reading will not bring people to a state of flow.

We do not wish to challenge users in the sense of climbing or gaming instead we want to reach some of the characteristics of reading. We want users to stay interested and feel that there are new things to discover. We do not want them to find the interaction simplistic and easy to comprehend and thereafter boring, but on the other hand we do not want them to feel that the application is uncontrollable and too hard to understand. To have users emotionally involved we want the interaction to be somewhat ambiguous and open for interpretation.

Most designers would probably see ambiguity as a dilemma for design. Gaver, however, looks upon it as “a resource for design that can be used to encourage close personal engagement” (Gaver et al. 2003). He argues that in an ambiguous situation people are forced to get involved and decide upon their own interpretation of what is happening. As affective interaction oftentimes is an invented, on-going process inside ourselves or between partners and close friends, taking on different shades and expressions in each relationship we have with others, ambiguity of the designed expressions will allow for interpretation that is personal to our needs. For example, if a system was to have buttons where each was labeled with a concrete
emotion, users might feel extremely limited since they would not be able to convey the subtleties of their emotional communication to others.

Ambiguity may also follow from the ideas of embodiment, which sees meaning as arising from social practice and use of systems – not from what designers intended originally. An open-ended ambiguous design might allow for interpretation and for taking expressions into use based on individual and collective interpretations – both by sender and receiver of affective messages. Ambiguity in a system will perhaps also create a certain amount of mystery that will keep users interested. However, there needs to be a balance, since too much ambiguity might make it hard to understand the interaction and might make users frustrated (Höök et al. 2003).

While Gaver wants to provoke people with his design so that they get into a process where they create their own meaning of the artifact it is not necessary to go that far. Sometimes it is enough for systems to have a little bit of ambiguity to them so that users can explore the interaction and find new alternative ways of interpreting the results and become emotionally involved in the process.

We intend to investigate the abovementioned and other criteria through a series of workshops, brainstorming sessions and practical application to use cases (see below). Our aim is that this will lead to an iterative refinement of useful criteria for developing affective interaction systems.

### 5.1.1 Criteria for Affective ECA Systems

So far, evaluations of affective interactive ECAs systems have mainly looked at factors such as: learning curve, performance level, degree of realism etc, but not at the quality of the interaction such as pleasantness of use, fun or engagement. These factors are important in video games and other recreation software; but they are also a major concern for other types of applications such as tutoring system, information booths and e-commerce systems.

Most evaluation studies done so far have been micro-evaluations of an ECA (does the ECA have lip-readable movements? Should an ECA display facial expressions of emotions, gaze behaviors or not?); fewer studies have looked at what is gained by having an affective ECA in an interaction. We believe that an important criterion for evaluation of an ECA is related to the notion of affective bonds created between users and agents. As pointed out by Lester, ECAs, by their mere presence, brought a touch of sociality. Within an interaction ECAs should perceive user’s emotion and react accordingly. But they can also affect user’s emotion by e.g. creating a positive affective state that favors learning. Another criterion of evaluation is related to the capacity of the agent to engage and maintain the user in an interaction. An affective interaction systems ought to address at least one such aspect.

Moreover, to allow for a smoother interaction, users should correctly perceive and interpret the emotions displayed by ECAs. Expression of emotions is done not only through a particular set of muscular contractions, hand shape or body movement but also through the “manner” of movement. While evaluation of the quality of expressivity and the quality of expressions of emotion will be done within WP6, we will look at the effect of expression of emotion and of its expressivity on interaction along the dimensions of enjoyment and engagement. We wish to know how it feels to interact with a bland ECA (Droopy-like) or an exaggerated one (more like a joker)? Should the ECA display a polite smile or a big laugh?
Several studies have mentioned the difficulty an ECA has to pass the Turing test. Such a result may be due to poor:

- graphics representation of the ECA,
- dialog capability,
- emotion model,
- personality model,
- culture model,
- social model,

There are loads of reason an ECA may not fool a user to make her believes it is like a human. But this might not be the proper question to ask. Rather, we pose as criteria of success to pass not the Turing test but to pass the believability test. An ECA should behave coherently during the course of the interaction (Ortony, 2000) with its personality and emotional models. Particular attention should also be given to check that the interactive ECA system affords to the user what the ECA can do so as not to raise users’ expectations too high.

### 5.2 Evaluation metrics (for criteria)

As discussed above, we aim to explore success criteria for affective interaction application wrt specific design aims and what the designer intended to achieve. Only if the designer’s intentions are experienced by the end-users as intended is it possible to talk about a successful system. Measuring the success of the designer’s intentions is oftentimes best done using subjective evaluation metrics.

A subjective evaluation metric can be quite open-ended – any sources of information may be collected and interpreted vis-a-vis the goals of the system. Anecdotal evidence, informal chats between users and system-builders, tiny study sizes, forms structured to influence user interpretation, no discussion or analysis of results: this may sound like a to-do list for bad evaluation. But these choices are deliberately and thoughtfully made, highlighting underlying conceptual problems in using standard so-called ‘scientific’ evaluation techniques to evaluate emotionally involving systems. Grossly speaking, there is a major conflict between scientific and design-oriented perspectives on user interaction, where a more design-oriented perspective strives to be subjective, while a science and engineering evaluation strives to be objective.

The subjective approach means that we might not be building systems for ‘normal’ or ‘average’ users. Instead we are interested in the richness and complexity of unique, individual users, cultural contexts, and resulting variety of interpretations and experiences of our systems. It is not appropriate to summarize the results of a study into a few statements that are said to hold for everyone. Also, the statistical averaging and laboratory simplifications necessary for reliable scientific statements may wash out all the details that interest us in a user-centred design process where we want input for the next cycle of design. We may prefer a rich, narrative, and singular understanding to a simpler but rigorous and generalisable understanding in the early stages of design.

Subjective measurements will be complemented by more generalisable and objective measures where appropriate and needed in order to further our knowledge of what constitutes a successful design.
5.3 Existing user-centred methods for design and evaluation

A set of user-centred methods for brainstorming, early evaluation, etc. will be applied in various settings and experiences will be gained regarding their suitability for affective interaction design. Several of the partners in WP9 have previous experiences of user-centred methods that will be used as a starting point.

T-Systems Focus groups

T-Systems has developed a phone-based personal digital assistant with an emotional component. The system can be controlled by the user via speech and provides two assistants (a serious female character and a funny male character) who can show emotions and react on the user’s emotions as well. The user’s emotions are concluded from a lexicographic analysis of the verbal/textual user input. The two characters were designed to react on the user’s emotion by giving direct feedback on the current emotional behaviour (“You seem to be angry, maybe I can help you!”) before going on with fixing the user’s dates and making appointments with the user’s business partners.

While designing the system, T-Systems ran several focus groups in order to find out the user’s attitude towards emotionality in interactive systems used in a professional context. The aim was also to collect some ideas for the concrete persona and dialogue design before investing too much effort into the implementation.

The groups were held with 6-8 participants and moderated by two psychologists who had some years of experience in the HCI and human factors area. A prototype of the emotional system was presented by playing sound files (pieces of the planned dialogue) and a “wizard of oz”-like demonstration of the system’s reaction to the user’s emotional behaviour. The main functionality of the system (scheduling and appointing) was also presented by sound files and verbal scenario descriptions.

The participants were then asked to imagine themselves spending a day with those two agents. They were given the task to produce application scenarios that were desirable for themselves and asked to write them down on moderation cards. The scenarios were visualized on a moderation board and discussed within the group by focussing on the following questions:

- Personal reasons for/against the usage of the emotional system.
- Positive/negative effects of the usage of the emotional system.
- Ideas for the concrete emotional behaviour of the system in certain dialogue situations.

The results of the discussion touched different levels of design issues/problems that had to be solved in the implementation stage of the system. Without asking them directly, answers to following questions were found.

- Should the user be asked explicitly by an emotional system to express his emotional status verbally or not?
- How should the system react on a negative user’s emotion? (by mirroring, joking, consolation etc.)?
- What kind of character is appropriate for an application like a PDA? Should it be e.g. the user’s best friend, buddy or a polite secretary?
- How can an emotional character show genuine emotions without having a personality that determines his reactions to the events in his “life”?
- What kind of persona/personality the users would like to meet in an office application?

The results were extremely useful to solve the detailed qualitative problems in designing believable and pleasant characters that seem authentic to the user. The results indicate that the method can be used in the early stages of the design process.

The method will be investigated and described in detail during the next 18 months.

**Brainstorming using user-centred methods**

In order to explore some of the existing methods for user-centred brainstorming and design, we plan to hold a brainstorming session together with several of the WP9 partners. The aim is to arrive at an application idea that will be relevant to some user group (using the persona method (Cooper, 1999), designing for extreme characters (Djajadingigrat et al., 2000), or the user-driven innovation method (Holmquist, 2004) to focus the design process on some particular user group with some particular need). Databases with how people behave in some scenario relevant to the application area will be used as an inspiration source, for example the data from the Bad Mouth Study performed already in WP9, collecting data from how a group of young women talk about movie starts in Sweden, the US and France respectively. Interestingly, this kind of approach requires something different from the databases than does testing of for example an affective speech system. In order to design an interactive application, we need inspiration to issues such as floor-grabbing: how come that one person can grab the floor and make the others listen? Can we use this as inspiration for how to make a character in the interface entertaining at the right moment? In general, timing issues are quite difficult (as discussed above, see also (Höök, 2004)).

The brainstorming session might also make use of methods such as bodystorming (Oulasvirta et al., 2003), where a user and a designer brainstorm together “in situ” where the application would latter be used as part of that person’s life. It is a compact method for doing a light-weight ethnography and brainstorming – all in one go.

Once ideas are generated, we may use methods such as “six thinking hats” for early evaluation of ideas (deBono, 1985).

The three-day workshop will also include early prototyping using “tiny fingers” methods (Rettig, 1994), modified to work with affective interaction applications. In a “tiny fingers” session, or so called low-fi prototyping, a system is created on paper and then tested with an end-user through simulating interaction with the low-fi prototype. A modern version of low-fi prototyping and testing with end users is what Iaccuci names “staged lived happenings” (Iaccuci and Isomursu, 2004). The aim is to, within three days, arrive at a viable application idea, already tested with some outside recruited end-users, that may subsequently be developed into a test application for WP6 and WP9.

The idea with this whole exercise is both to teach user-centred design methods to other partners in the HUMAINE network, but also, simultaneously get feedback on how well these methods can capture the difficult issue of emotional involvement with an artefact or system. Can these kinds of systems provide input that helps steer towards an application that will indeed make users emotionally involved? If not, how can we modify them to fit with this particular problem?

**Cantoche Avatar Tool**
During the first 18 months, Cantoche will work on creating a tool that allows researchers to experiment with emotional avatars capable of facial expressions and gestures. The tool consists of three main components

- An ECA. Probably, a woman.
- An editor in order to work on the facial expressions and to prepare a character with different expressions
- A scripting tool that allows one to create several scenarios the user can play.

Using the tool researchers can easily set-up and evaluate the effect of varying e.g. the intensity (exaggeration) of the facial expressions and the added value of gestures. The tool can among other things be used to rapidly try out new application ideas generated during brainstorming sessions as described above.

5.4 New methods for design and evaluation

5.4.1 Sensual Evaluation Instrument

Neurophysiologists and psychologists have in recent years proposed that our brains, rather than operating in a wholly logical, conscious verbal manner, actually process information and make decisions using various layers working in parallel, complementary ways. They have demonstrated, for example, that we can learn something new and ‘intuitively’ put it into action before we are able to consciously verbalize it (Myers, 2002).

Affective processing, in particular, has been shown to occur at levels other than the cognitive/word-oriented level of the brain (e.g. the primal nature of fear: http://www.nimh.nih.gov/events/ledoux.htm).

Yet we rely heavily on verbal reports from users for evaluation of the affective effects of systems. Likert scale items (e.g. How enjoyable was this interaction? Circle the number from 1-7 that best fits) are still the most common subjective quantitative measures used.

Of course it is possible to supplement verbal reports with some form of observation of affective reaction—videotaping or physiological tracking. However, we wonder if it is possible to develop a form of subjective report mechanism that ‘end-runs’ the verbal part of a person. The exploration of this terrain is the goal of our project.

Such an instrument could provide several key benefits:

- Preserves benefits of subjective measures (easy to use, portable, and empowering the user).
- Transcends language and cultural barriers.
- Provides results more in line with current research about affective processing—potentially less distortion through the ‘lens’ of the verbal mind.
- More fun for the user!
Direction for exploring solutions to this problem:

We would like to pursue the possibilities for subjective, self-report of affective state in nonverbal ways. We are especially interested in developing an instrument that could be calibrated and re-used to measure affective reaction to systems and interactions, across cultures, something that could engage the senses in alternate ways—a sensual evaluation instrument.

Related work:

There is some history of the standardization and use of nonverbal scales in psychology (e.g. PONS (the Profile of Nonverbal Sensitivity), from which we can draw lessons.

To date, most work done on nonverbal systems of evaluation has involved anthropomorphic imagery (likenesses of human faces and/or bodies). For example, researchers who work with children have established the value and reliability of face-based Likert scales for determining emotional responses to systems and situations (e.g. Wong and Baker’s work on children’s subjective evaluation of pain—see Figure 1)

![Figure 1. A pain scale created to help children report on their pain feelings.](image1.png)

There are also popular uses of nonverbal affective scales (thumbs-up and down movie ratings, the ‘little man’ on the San Francisco Chronicle movie review page [http://www.sfgate.com/eguide/movies/reviews/](http://www.sfgate.com/eguide/movies/reviews/) see Figure 2), indicating that calibration and use of nonverbal scales is possible and appealing in everyday contexts.

![Figure 2. The San Francisco Chronicle’s movie review system uses facial expression and body posture to indicate a movie’s quality.](image2.png)

Finally, there has been some work in the product design community on mapping product qualities to affective reactions, for example this facial wheel used by Wensveen and colleagues in developing an affectively appropriate alarm clock (Figure 3, see Wensveen, 1999)
Our wish in this project is to move away from faces and figures, toward some form of nonverbal code that still evokes emotion without explicitly representing the human form. We also hope to extend the sensory experience of this instrument beyond the purely visual.

There is some nonrepresentational work on sensing emotion in alternate sensory channels. For example in his book *Sentics*, Clynert describes characteristic movement patterns on a touch pad when users are asked to ‘perform’ a particular emotion with their finger (Clynes, 1989). Höök et al.’s work on eMoto, a non-representational system for adding affective content to emails, is an example of the use of gesture and touch to generate an emotional response (Fagerberg et al., 2003, 2004). Product designers know that surface materials and their tactile qualities can profoundly impact users’ emotional response to products (Green and Jordon, 2002), but there has been little systematic work done up to now to assess the specific emotional effects of various materials.

**Design Process:**

**Design criteria**

As initial design criteria, we proposed that the sensual evaluation instrument should be:

- Intuitive to use
- Nonverbal (make use of visual/tactile/auditory and other nonverbal cues instead)
- Provides reliable results

*Figure 3. Facial expressions arranged along the axes of arousal and valence, used to help product designers gauge user emotions about designs.*
- Easy to calibrate for different user groups
- Portable, durable.

Phase one: User research and early prototyping

We began the design process with a series of brainstorming sessions, and a cycle of user research and rapid prototyping of a trial group of objects.

Our first brainstorming sessions led us to the following suggestion: we would create (in partnership with product design experts) a kit of non-representational objects that make use of affective cues (e.g. color, texture, shape). Taking advantage of visual and tactile cues.

We began by exploring ways that everyday objects provoke emotional response. We made use of the Russell circle, which arrays emotions along the axes of arousal and valence (see Höök et al. for further description of this taxonomy). Each researcher brought small household objects that had emotional valence for them to the lab, and arrayed these objects upon the Russell circle, projected on a floor carpet (see Figure 2).

![Figure 2 Objects arrayed on the Russell circle](image)

Based upon these initial explorations, we decided to craft a simple set of objects that varied only upon one sensual dimension—color—to test whether users could ‘calibrate’ these objects in an emotional taxonomy and then use them to evaluate an interaction with an interface.

We constructed our objects from heavy candles and felt (see Figure 4), and conducted an informal user session in which participants first arrayed the objects according to their emotional valence, then used them to evaluate their experience of a computer game. Participants would simply move the objects closer to them, that reflected their current emotional state, as they were playing. They would shift the mixture of objects in front of them as their emotions changed during the course of the interaction.
Users seemed to enjoy the calibration exercise, and displayed a variety of approaches to arraying the objects. Some users took a very narrative-based approach to mapping the objects (e.g. green and red remind me of Christmas). Those using this approach had a much harder time using the calibrated objects to evaluate their engagement with the computer game. In both cases, users had a difficult time remembering the mapping they had made between the objects and emotions.

In the sessions, we also tried out a screen-based mapping, allowing users to indicate by pointing to the circle, what emotion they were feeling. We noticed some advantages to having physical objects. Users were able to easily indicate multiple emotions at once (I feel confused but also excited and a little scared), and it seemed that they could manipulate the physical objects without taking as much attention away from working with the game itself. There is some evidence from the study of attention that dissimilar tasks are easier to perform in parallel (Pashler, 1998).

We also found that the set of emotions that was desirable to communicate the experience of the game was different than the traditional set of primary emotions from category-based approaches to emotion (e.g. Ekman, 1972). For example, representing emotions such as boredom, confusion, neutral state, and frustration was important.

Next steps:

Based upon these initial investigations, we decided to create a next-generation set of objects to further explore whether physical objects are an intuitive and reliable way to give emotional evaluation of interactive experiences. Rather than using color, we plan to use the form of the object itself to evoke the various emotions. We are drawing upon the work by Disney animators showing how emotions can be evoked through physical forms (see Figure 5). By indirectly referencing the human form, we hope to gain the advantage of the universality of expression of emotion through body position, without relying on explicitly anthropomorphic representations.
Figure 5. The famous Disney flour sack, an illustration of emotions despite the absence of a clearly articulated human figure.

We are working with a sculptor who has experience crafting evocative, handheld objects to draft a set of emotional objects. (See Figure 6 for examples of her prior work; see also http://www.raineystraus.com).

Figure 6. Sample objects created by the sculptor who will help us generate the sensual evaluation object prototypes.

We plan to create an initial set of objects representing the following emotions (descriptions in parentheses clarify the user context for the emotion), which are uniquely tailored to evaluation of emotionally active systems:

- Confusion (I don't get what's going on here)
- Frustration (what the system just did drove me nuts; or, I can't solve this level and I hate this right now)
- Fear (the game is making me anxious; or, I think I might've erased the wrong files)
- Happiness at success (I just cracked a level; or, I just figured out how to do a new thing)
- Surprise (positive--something good happened I wasn't expecting)
- Surprise (negative--something bad happened I wasn't expecting)
- Satisfaction (something happened that I like)
- Contentment (all is okay, going smoothly)
- Frantic stress (things are out of my control, too much going on)
- Flow (I'm in my groove right now, really enjoying working with the system, we feel as one)
- Neutral (not feeling emotions right now, just working...)

We plan to conduct further user evaluations with this set of objects, beginning with small-scale tests, and moving toward cross-cultural validation of the object set, should we find that it supports evaluation as we hope.

5.4.2 WoZ environment for multimodal emotional interaction

Although a number of studies have been produced, which describe how users see Embodied Animated Agents and how their vision is influenced by variations in the agent characteristics, the exact nature of the relationship which is established in conversations between human and artificial agents is still unclear. Several studies at Stanford, based on the always cited media equation hypothesis, supported the idea of applying social science theories in this domain. However, the need to better specify the applicability conditions of this hypothesis and its rationale begins to emerge. Some studies proved that human interaction with technology is not exactly the same as the human-human one, and that humans tend to automatically adapt their dialog style when they are aware of interacting with a tool. These findings brought to propose organizing Wizard of Oz studies to investigate interaction issues in natural language and with artificial agents, from which to collect evaluations of the agent behaviour in different conditions and, at the same time, qualitative and deeper data from analysis of the resulting corpus of dialogs.

Affective interaction is a generic term which may have several nuances. One of the open questions about human relationships with ECAs is whether empathy may really occur in these relationships: that is, whether the user may feel (and show) empathy for an ECA, and conversely how the simulation of affective ECA behaviour is interpreted (and appreciated) by the human. Empathy is a quite fuzzy concept: it implies listening skill and emotional intelligence, with the ability to identify with and understand another’s situation, feelings and motives. It therefore implies an ‘active’ attitude, requires some kind of cognitive evaluation of the interlocutors situation and of their emotional state, may occur even in absence of any expression of emotion by the ‘empathizing interlocutor’ and may be either sincere or pretended. A broader meaning is attributed by others to this concept, by claiming that: “Empathy depends not only on one’s ability to identify someone else’s emotions but also on
one’s capacity to put oneself in the other person’s place and to experience an appropriate emotional response … The empathor empathizes not only with the empathee’s emotions but also with his physical state and other parameters of existence”. In this broader sense, empathy may be seen as the process of entering into a warm social relationship with someone else, of being in a way involved in her goals and feelings.

Attempting to provide a (provisional!) answer to the question of whether empathy may occur in human-ECA interactions therefore requires reflecting on the kind of social relationship that may be established between an ‘appropriately designed’ ECA and a user and on the factors which influence this relationship. Cassel and Bickmore are among those who most deeply reflected on this topic. Social relationships may facilitate, in their view, trust and collaboration between users and ECAs and this may obtained by endowing agents with the ability to apply some of the strategies which are applied by humans to this aim: in particular, increase intimacy and common ground over the course of the conversation, decrease interpersonal distance, use non explicit ways of achieving conversational goals and display expertise. With their Perceiving and Experiencing Fictional Characters (PEFiC) model, Hoorn and Konijn proposed a key for interpreting these phenomena, by claiming that the appraisal of characters by an observer occurs primarily along ‘ethical’, ‘aesthetic’ and ‘epistemic’ dimensions; positive and negative values of these dimensions influence, respectively, the ‘involvement’ and ‘distance’ of observers towards the agent, and a good blending of positive and negative features seems to be ideal to produce appreciation of the agent by the user.

Involvement, ‘friendship’ and empathy are closely related concepts, although they are not synonyms. Friendship may involve varying types and degrees of companionship, intimacy, affection and mutual assistance. It is influenced, again, by interpersonal attraction but also by ‘rewards’, which should outweigh costs such as irritation or disappointment. In the case of information-giving dialogs, rewards are affected by the subject’s expectations and goals: therefore, even if subjects are pre-informed that the ECA with which they are going to interact is still in a prototypical stage, their involvement will be affected by the degree of satisfaction in the information received and by how pleasant they will find interacting with it. How may involvement of users be increased or, in other terms, empathy be induced by the agent on the user? Several authors strive for avoiding too much realism or too much ‘positive’ features, to rather employ agents whose features are ‘a little bad, ugly and unrealistic and that arouse some negative valence and dissimilarity with their daily practice’. On the contrary, relevance should be insured at its best, by providing features that ‘tune into the goals and concerns of the user’.

By building on the body of the cited theories, in the scope of WP9 we aim at discussing whether and how empathy, in the broad sense of ‘entering into a warm social relationship with someone’, may be induced in interactions with ECAs and how it may be measured. A major issue here is also to account for the fact that human emotion perception (as well as display) is inherently multimodal, and situationally and socio-culturally mediated, cf. (Scherer, Ceschi 2000). Therefore it is not only important to identify and assess design criteria for modelling the multimodal affective behaviour simulated by the ECA, but also to take into account contextual criteria such the interaction scenario, the type of dialogue, the roles of the communication partners, etc., as well as the animation technology and the speech technology employed which strongly influence the look and sound of the animated characters.

As our interest is focused on methods rather than on evaluation of specific products, we wish to build a ‘testbed’ for evaluation of multimodal affective interaction integrating a tool for Wizard of Oz studies, and a system for generating multimodal dialogue. This concept of an integrated Wizard of Oz (WoZ) and multimodal generation tool offers us high flexibility in
the collection of data on affective human-ECA interaction, as well as in the design of evaluation and usability studies of various kinds.

As regards the WoZ environment, we build upon existing work at the University of Bari (Lit). For the generation system, we draw upon the NECA platform developed in the FP5 IST project NECA (http://www.ai.univei.ac.at/NECA/) which has been coordinated by OFAI and on the dialog simulator which was developed in the scope of MAGICSTER.

Other than in NECA, where full dialogues are generated in one go, the generation system envisaged for the HUMAINE WP9 exemplar will also allow for the generation of smaller units e.g. single turns. This is a preconditions for true interactivity. Accordingly, the architecture will be designed such that the system can be extended to a full interaction system, where multimodal communication is viewed as a sequence of action-react processes.

The work on the generation system is done in close cooperation with WP6 where an architecture for interactive ECA systems is developed, and with WP8, where an advice-giving dialog will be prototyped. Similar as suggested in WP6, we will also rely on file-based interfaces, which allows us to work with dummy system components where only in- and output representations are specified.

The two major assets of the proposed testbed are

1. high flexibility, in particular
   the WoZ environment employed allows quick customization to new application domains, dialogue scenarios and languages;
   the testbed can work with different player technologies, i.e., due to the use of a common high level scripting language (developed in WP6) the components in the testbed are largely independent of the player technologies employed;

2. creation of design-evaluation loops, i.e.,
   while the WoZ environment allows us to deliberately collect data on affective communication behaviour which in turn are analysed to support theory building, the generation environment enables us to model these insights and render the display of emotional behaviour in ECAs. The effect on the user can then be tested in various ways, either by playing animated scenes to the user (the NECA approach) or by integrating the user in the communication process with the ECA (the Magicster approach)

Considerations concerning the design of affective ECAs will be based on data acquired in a set of Wizard of Oz studies, that we will design and perform with several purposes: as an iterative design method to refine conversational characters, as a tool to reflect and test hypotheses about relationships between humans and ECAs and the factors which affect them, and as a source of a corpus of data to employ in recognizing and modelling the ‘affective’ state of the user during interaction.

In addition, replacing the wizard by an automatic system (which in the base version is realized as a suite of scripts, which can be partially or, if appropriate in the course of developments in HUMAINE, fully replaced by dedicated modules) will also help us to identify possible differences in user involvement and appreciation arising from differences in the dialogue behaviour of the wizard-driven ECA and the computer-driven version of the ECA. This is important to find out in how far the findings from WoZ studies can be transformed into
computer generated ECAs. Of course it is not the goal of HUMAINE to develop and implement fully-fletched ECA systems, and this is definitely not what we are aiming at with the proposed test bed. The overall goal of the proposed test bed is to facilitate studies that will apply criteria for defining whether an affective interaction system is 'usable' by translating them into several, complementary kinds of measures. On one side, the classical, 'quantitative' evaluations of various aspects of the agent and the message, by means of 'Likert scales', to assess the user's subjective feeling about overall interaction and on the other side, evaluation of individual modes by means of 'sensual' methods described in Section ...

However, these methods will not be the main and unique evaluation criteria. Another strand of evaluation methods will be based on the analysis of the user behavior in terms of their 'level of initiative' in the dialog, 'level of involvement', 'level and type of social relationship' with the ECA. Although being qualitative, these evaluation criteria will be translated into quantitative terms (evaluation metrics), to enable comparison of various experimental conditions and to enable decision of how to revise the prototype after every iterative design step.

Side-effects of these studies will be the following.

a) collection of a corpus of interactions of users with ECAs, in various contexts and conditions: the kind of data collected will go from pure text to speech and possibly also to videos;

b) labelling of this corpus with a mark-up language which is defined according to goals of analysis but is, at the same time, compatible with the reference language which will be proposed in WP5;

c) methods of analysis of this corpus, to extract signs of 'subtle' emotions which are typical of human-computer interactions (irritation, approval, satisfaction etc) but also of the 'social' emotions which are typical of interactions with ECAs in particular (sympathy, antipathy, sense of intimacy, etc).

To have this testbed at hand is crucial for advancing the research in evaluation and usability of ECAs, especially as the area of research is in its early stage and there are up to date no well established design criteria for the evaluation and usability testing of ECA systems, but a variety of loose ends. This also holds for theory building in the display and perception of multimodal affective behaviour, an area which will be much further advanced in the course of HUMAINE. Thus it is crucial to have a flexible testbed which can be easily adapted to the requirements in testing novel insights gained from affective theory.

5.4.3 Extended think-aloud protocol

Ericsson and Simon (1993) presented the think aloud protocol as a viable method that accurately reveals the cognitive processes supporting problem-solving. Since then, among other applications, usability testing has leveraged the think aloud protocol to provide insight on users' cognitive workings during their interaction with an interface. In a typical usability session, human subjects are instructed to verbalize their thought process while conducting a certain task. Therefore, think-aloud accounts in the domain of usability testing, provide the reasons behind users’ decisions or behaviours. Interestingly, the think aloud protocol in its original form did not suffice and was readapted to fit the special requirements of usability testing (Boren & Ramey, 2000).
Departing from its cognitive facility, we propose an emotional extension of the think aloud protocol that will serve as an evaluation tool revealing information about a user’s emotional state. We present two facts to support this proposal. First, vocal expression is characterized by identifiable changes in acoustic cues that map to a number of emotions with above chance accuracy. In fact, the possibility of correlating vocal expression to an emotional state has been the driving force behind many vocal evaluation efforts (Scherer, 2003). Second, voice appears to be spontaneous in expression. For example, voice pitch has been found to persist despite efforts to suppress expression and to deceive others (Ekman, Friesen & Scherer, 1976).

In our short term objectives, we envision the emotional think-aloud extension to evolve around three points:

- **User training.** Designing useful and effective instructions to train users in the extended protocol administration
- **Individual differences.** Determining the cultural universality of vocal emotional expressivity during the application of the extended think aloud
- **Decoding.** Achieving inter-rater agreement and reliability (always in terms of user-rater culture).

Our long term efforts will focus on speech analysis tools that will support evaluators during their assessments.
6 Proposed allocation of tasks

The work in WP9 will be divided into 4 main tasks corresponding to the elements of the exemplar as described above. A description of the tasks and estimated starting and ending dates can be found below.

<table>
<thead>
<tr>
<th>Task</th>
<th>Description</th>
<th>Start</th>
<th>End</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>Criteria for usable affective interaction systems</td>
<td>M0</td>
<td>M48</td>
</tr>
<tr>
<td>T2</td>
<td>Evaluation metrics</td>
<td>M18</td>
<td>M48</td>
</tr>
<tr>
<td>T3</td>
<td>Existing user-centred methods for design and evaluation</td>
<td>M13</td>
<td>M48</td>
</tr>
<tr>
<td>T4</td>
<td>Exploring new methods for design and evaluation</td>
<td>M0</td>
<td>M48</td>
</tr>
</tbody>
</table>

Table 1 Summary of main tasks

6.1 Criteria for usable affective interaction systems

<table>
<thead>
<tr>
<th>Task</th>
<th>Description</th>
<th>Start</th>
<th>End</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>Criteria for usable affective interaction systems</td>
<td>M0</td>
<td>M48</td>
</tr>
</tbody>
</table>

We aim to explore success criteria for affective interaction application wrt the specific design aims and what the designer intended to achieve. Only if the designer’s intentions are experienced by the end-users as intended is it possible to talk about a successful system. In addition, concepts like flow, ambiguity, pleasure and expressivity will be explored as potential criteria for success.

6.2 Evaluation metrics (for criteria)

<table>
<thead>
<tr>
<th>Task</th>
<th>Description</th>
<th>Start</th>
<th>End</th>
</tr>
</thead>
<tbody>
<tr>
<td>T2</td>
<td>Evaluation metrics</td>
<td>M18</td>
<td>M48</td>
</tr>
</tbody>
</table>

Subjective measurements will be complemented by more generalisable and objective measures where appropriate and needed in order to further our knowledge of what constitutes a successful design.

6.3 Existing user-centred methods for design and evaluation

<table>
<thead>
<tr>
<th>Task</th>
<th>Description</th>
<th>Start</th>
<th>End</th>
</tr>
</thead>
<tbody>
<tr>
<td>T3.1</td>
<td>Methods for design</td>
<td>M13</td>
<td>M23</td>
</tr>
<tr>
<td>T3.2</td>
<td>Methods for evaluation</td>
<td>M18</td>
<td>M48</td>
</tr>
</tbody>
</table>
Existing user-centred design and evaluation methods will be exploited and when needed modified to fit with the specific problem of creating affectively involving systems.

### 6.4 New methods for design and evaluation

<table>
<thead>
<tr>
<th>Task</th>
<th>Description</th>
<th>Start</th>
<th>End</th>
</tr>
</thead>
<tbody>
<tr>
<td>T4.1</td>
<td>Sensual method</td>
<td>M0</td>
<td>M48</td>
</tr>
<tr>
<td>T4.1.1</td>
<td>Design of sensual method</td>
<td>M0</td>
<td>M36</td>
</tr>
<tr>
<td>T4.1.2</td>
<td>Application/evaluation of sensual method</td>
<td>M18</td>
<td>M48</td>
</tr>
<tr>
<td>T4.2</td>
<td>WoZ test-bed</td>
<td>M0</td>
<td>M48</td>
</tr>
<tr>
<td>T4.2.1</td>
<td>Assessment of existing evaluation methods and strategies for ECAs</td>
<td>M0</td>
<td>M19</td>
</tr>
<tr>
<td>T4.2.2</td>
<td>Realizations of the design-evaluation loop</td>
<td>M13</td>
<td>M48</td>
</tr>
<tr>
<td>T4.3</td>
<td>Extended think aloud protocol</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T4.3.1</td>
<td>Development of the extended emotional-think aloud protocol</td>
<td>M15</td>
<td>M17</td>
</tr>
<tr>
<td>T4.3.2</td>
<td>Instructions employment and ratings accuracy experiment</td>
<td>M17</td>
<td>M20</td>
</tr>
<tr>
<td>T4.3.3</td>
<td>Experiment considering cultural vs. universal emotion expression and interpretation always in terms of the extended think aloud protocol</td>
<td>M21</td>
<td>M24</td>
</tr>
</tbody>
</table>

T4.1.1: Here we will investigate the possibilities for subjective self-report of affective state in nonverbal ways. As described above this work has already started and resulted in one possible toolset. This toolset will iteratively by refined and expanded during the project.

T4.1.2: The toolset(s) will be applied in “real-world” settings to verify their functionality and to gain a better understanding of possibilities and limitations of the proposed form of evaluation studies. Results from T4.1.1 will feed into this task and vice versa.

T4.2.1: Here we will give a review of the state-of-the art in evaluation studies on ECAs, and draw amongst others upon evaluation studies on empathic ECAs conducted at the University of Bari and at OFAI. The review will start from a reflection on the concept of ‘empathy’, to critically reflect on whether and how ECAs have proved, so far, to be able to establish this kind of relationship with their human users. The study will consider the evaluation criteria and methods applied, and will critically present in a flexible structure and dynamic structure the results produced. Results from this task feeds into T1 “Criteria for affective interaction systems” and T3 “Existing user-centred methods for design and evaluation”.

T4.2.2: Design and conduction of WoZ experiments, in order to hypothesize and test aspects of empathic interactions of users with ECAs, in various application domains, in various contexts and by subjects with various background. These experiments will consist in sets of
small-scale iterative studies, which will be designed so as to provide to users a framework which will enable them to express their requirements with a relatively high degree of freedom. Results expected from these experiments will not only consist in statistics about subjective and objective evaluation parameters, but will provide, as well, a corpus of data upon which to ground the design of artificial dialogs. The insights gained from every WoZ experiment will therefore feed into the redesign of the WoZ experiments, as well as into modelling of fully computer generated empathic ECAs, the effect of which will then be tested in user experiments. Results from this task feeds into T2 “Evaluation metrics for criteria” and T4 “Exploring new methods for designs and evaluation”.

T4.3.1: In this stage we will determine the nuances of the current cognitive think aloud protocol and will adapt the protocol to current knowledge of emotion in speech.

T4.3.2: Participants will take part in a DigitalBlush experiment where during their interaction an emotion eliciting event will take place. We will focus on the use of instructions employed during the protocol application (and given to the participants prior to the experiment) and their effects on rating reliability. Rating reliability will also be considered independently and correlated to participants’ self reports.

T4.3.3: Participants will take part in a DigitalBlush experiment where during their interaction an emotion eliciting event will take place. We will investigate the impact of users’ culture on the evaluators’ interpretation. Evaluators will belong to the same culture as the participant or to a distant cultural group. Three cultural groups will take part in this experimental study (Swedish, English and Czech). The end objective will be to determine whether evaluation accuracy is higher when user and evaluator share the same culture.
# 7 Provisional timeline

## 7.1 Criteria for usable affective interaction systems

<table>
<thead>
<tr>
<th>Key actions</th>
<th>Estimated completion</th>
<th>Type of associated deliverable</th>
<th>Estimated date of delivery</th>
</tr>
</thead>
<tbody>
<tr>
<td>T4.2: A webpage consisting of a commented literature list on ECA evaluation studies and of initial definition criteria for empathic ECAs is available on the HUMAINE portal.</td>
<td>M19</td>
<td>Webpage</td>
<td></td>
</tr>
<tr>
<td>T1: Workshop on design criteria/concepts for affective interaction design</td>
<td>M22</td>
<td>Report</td>
<td></td>
</tr>
</tbody>
</table>

## 7.2 Evaluation metrics (for criteria)

<table>
<thead>
<tr>
<th>Key actions</th>
<th>Estimated completion</th>
<th>Type of associated deliverable</th>
<th>Estimated date of delivery</th>
</tr>
</thead>
<tbody>
<tr>
<td>T4.2: First version of a comprehensive collection of design, test and evaluation strategies and methods for empathic ECAs is available.</td>
<td>M30</td>
<td>A document which specifies aspects to evaluate according to the goal of the study and describes methods to employ in measuring them.</td>
<td></td>
</tr>
<tr>
<td>T2: Assessment of subjective evaluation metrics</td>
<td>M30</td>
<td>Document describing possible subjective evaluation metrics</td>
<td></td>
</tr>
</tbody>
</table>
### 7.3 Existing user-centred methods for design and evaluation

<table>
<thead>
<tr>
<th>Key actions</th>
<th>Estimated completion</th>
<th>Type of associated deliverable</th>
<th>Estimated date of delivery</th>
</tr>
</thead>
<tbody>
<tr>
<td>T4.2: A webpage consisting of a commented literature list on evaluation</td>
<td>M19</td>
<td>Web page</td>
<td></td>
</tr>
<tr>
<td>studies of ECAs, of natural dialog systems and of other interaction systems</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>which employ some form of ‘embodied interaction. Initial definition criteria for empathic ECAs is available on the HUMAINE portal.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T3: Workshop about existing methods for affective interaction design (brainstorming session)</td>
<td>M15</td>
<td>Design of example application from user-centred perspective</td>
<td>M15</td>
</tr>
<tr>
<td>T3: Workshop about existing methods for affective interaction evaluation</td>
<td>M 30</td>
<td>Evaluation of example application from user-centred perspective</td>
<td></td>
</tr>
<tr>
<td>(brainstorming session)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 7.4 New methods for design and evaluation

<table>
<thead>
<tr>
<th>Key actions</th>
<th>Estimated completion</th>
<th>Type of associated deliverable</th>
<th>Estimated date of delivery</th>
</tr>
</thead>
<tbody>
<tr>
<td>T4.1: First design of emotion-evoking sub-symbolic set of objects</td>
<td>M18</td>
<td>Document</td>
<td></td>
</tr>
<tr>
<td>T4.1: First realisation of object set</td>
<td>M24</td>
<td>hardware</td>
<td></td>
</tr>
<tr>
<td>T4.2: First version of a comprehensive collection of hypotheses about human-</td>
<td>M30</td>
<td>A WoZ tool for iterative</td>
<td></td>
</tr>
<tr>
<td>ECAs interaction and how to design and evaluate systems which consider these</td>
<td></td>
<td>design of ECAs in various</td>
<td></td>
</tr>
<tr>
<td>interaction factors.</td>
<td></td>
<td>conditions, with</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>a</td>
<td></td>
</tr>
<tr>
<td>T4.3: First proposition of extended think aloud</td>
<td>M17</td>
<td>Document</td>
<td></td>
</tr>
<tr>
<td>T4.3: Validation study of extended think aloud</td>
<td>M20</td>
<td>Document</td>
<td></td>
</tr>
<tr>
<td>T4.3: Cultural study of extended think aloud</td>
<td>M24</td>
<td>Document</td>
<td></td>
</tr>
</tbody>
</table>
8 Proposed Use-cases for design and evaluation methods

The tools and methods of the framework will be applied and evaluated in a variety of situations/applications/domains. Records from the sessions will provide valuable guidance for future users of the tools and methods regarding their proper usage and expected results. Following is a list of proposed applications and domains and a preliminary account related methods/tools from the framework.

8.1 Dramatic Gaming

Interactive drama focuses on providing a dramatic experience rather than an action packed one in contrast to many existing computer games. Instead of monster bashing players can engage in navigating the socio-emotional web linking characters together. This is something that requires players to be both sensitive and attentive to social and emotional cues displayed by characters. In an interactive drama the motivating factor for playing a game switches from “winning” to exploring the social and emotional relations between characters.

How to design and evaluate interactive drama is yet an open question. So far research has focused mostly on how to create believable characters (discussed elsewhere in this deliverable) and to a lesser extent story sequencing. Creating characters with the necessary attributes to become interesting counterparts in emotional and social interactions is challenging. However, there are also other aspects that can influence a player’s experience and function as emotional boosters e.g. how characters are presented (cinematography), sound effects and music that have received less attention.

We will evaluate and redesign an interactive drama system based on existing technology developed at SICS (Swedish Institute of Computer Science). An important step on the way will be to refine the scenario for the game. The idea is to use methods developed within the WP (e.g. the affective instrument) to incrementally improve on the system as well as the scenario. Results from the evaluations will feed back into the iterative design/evaluation/redesign cycle.

An important aspect of the work will be to evaluate cinematography as a method for enhancing emotional expression. Cinematography refers to how something is filmed – in contrast to what is being filmed – and typically involves three factors:

- **Photographic aspects of a shot**, e.g. how a shot is illuminated. For instance a shot can be very dark and gloomy, light from the sides casting sinister shadows, or it can be bright and happy.

- **Framing of a shot**, i.e. what is included in the camera rectangle and its location within the rectangle. For instance, a shot can be centred on a person talking to someone outside the frame, or both persons can be visible on opposite sides of the frame.

- **The duration of a shot**. A shot can be very long, e.g. showing a person giving a speech, or short, showing the person giving the speech and then rapidly cutting to the audience’s reactions to the speech.

By altering these factors different dramatic and emotional effects can be constructed although the content remains the same. Inspired by comic artists, colour theory and industrial design we
also want to explore how colour and shape of shot frames can boost the emotional content of the drama.

Rizzo (1999) has shown that emotions sometimes require a narrative context in which they can be given their meaning and understood. Hence, another important aspect of the work will be to evaluate how the story sequencing mechanism affects users’ emotions as well as their perception of the constructed story. We intend to use several of the methods developed within WP9 to evaluate these aspects and find ways of designing for interactive drama.

8.2 Designing “bad mouthing” ECAs

ECAs are one output of affective interface research. As of yet, there are not compelling applications for these characters, that truly take advantage of their expressive power and have demonstrably favourable user response using facial expressions and gestures.

We believe this may be in part due to a focus on tactical versus social conversation applications for such faces. Emotional expressions are layered onto essentially functional dialog (e.g. about health care or real estate), without regard for whether the expressions truly provide additional value to users.

We also believe that ECAs may gain in appeal and believability to the extent that we ground their development in situated observation of behavior. That is to say, if we want to design a social application for the ECA, we should look at how people really use their faces and body in such social contexts, and build the expression set and model from this sort of contextualized data. This is in opposition to working from generalized models of emotional expression and then layering them onto a given social situation.

We intend to use methods and tools developed within WP9 to aid WP6 to develop an ECA application that is primarily social in nature, building from observation of real human faces in an appropriate context.

The target communication context is engaging in gossiping about celebrities or other people, and either praising or criticizing them.

We intend to videotape young women engaging in this kind of talk, in several countries (currently, Sweden, the U.S., Japan, France and possibly also Ireland). We plan to analyze this material with the assistance of WP3, to discern patterns of emotional expression.

Together with WP6 we intend to use our results to evolve an engaging application with an ECA provided by Cantoche.

The evaluation of the application will be done in association with work undertaken in WP6 and WP8. Thus the process of the design and evaluation of the considered application will be done having 2 dimensions of interest in mind:

- Emotion in interaction: WP6 is mainly concerned with elaborating affective interactive ECAs. Several capabilities have been defined to be essential for such systems. Within WP9, we will consider some of them, namely: creating affective bonds, displaying multimodal expressive behaviors. The first one refers to the capability to establish and maintain affective connection with the user; the second one regards the behaviors specification of the agent. More detailed descriptions of these capabilities may be found in Deliverable 6c. The evaluation on how well these capabilities have been
elaborated (how well expressivity is modelled; if the users perceive correctly the facial expressions and the gesture of an ECA; …) is part of WP6. Rather, within WP9, we will evaluate what is the added value of the capabilities in the given application.

- Persuasion: One subject of interest of WP8 is related to the different facets of persuasion. Here we are particularly interested in understanding what is the role and effect of an ECA on the user, and in determining which factors of the interaction acts on the persuasion dimensions.

WP9 provides the evaluative expertise to the team as the design moves all the way to trials with real users.

Next steps:

- The collaboration partners (WP9, 6, and 8) will meet to brainstorm based on the results of the initial videotaping, in early June.

- Cantoche, a partner in WP8 and WP9, has agreed to, not only to be an active participant but also to provide material for the evaluation: 3 components will be provided:
  - An ECA. Probably, a woman.
  - An editor in order to work on the facial expressions and to prepare a character with different expressions
  - A scripting tool that allows one to create several scenarios the user can play.
  - We would evaluate the different results depending of the intensity (exaggeration) of the facial expressions and the added value of the gestures.

- WP3 and 9 need to meet to talk about analysis techniques and tools.

### 8.3 DigitalBlush

DigitalBlush is envisioned as a community where collaborative work takes place under the lens of self-evaluative emotions. We believe that the inclusion of “self-aware” and “moral” tools will lead to productive, controlled and prosocial interactions. We propose this DigitalBlush community to be brought together in a distance learning setting. In this setting, participants of the community (i.e. students and professors) will connect with other members via their personal agents termed ‘moral agents’. Participants will conduct in team activities such as project assignments and will carry certain responsibilities towards their fellow members. Among other possibilities, in the context of this community a dishonest exchange of information (i.e. not delivering a promised assignment) will constitute a violation.

A number of exchanges will take place in our platform

- Human participants of a networked community will connect to each other synchronously and asynchronously via their moral agents.

- Moral agents will represent their human counterparts in the agent community.

- Moral agents will not be entirely autonomous. Their role will be one of communicative and moral facilitation. Specifically,
  - A moral agent will carry social cues from its human participant to other participants and in reverse
A moral agent will evaluate the success of its interactions with others and will relate its moral judgment to its human counterpart.

In our evaluation studies, DigitalBlush will serve as our experimental platform. In this platform, the emotions of shame and embarrassment will be induced by a number of events taking place during collaborative work activities (i.e. accusation of a low quality deliverable). The extended think-aloud protocol will be used to assess the presence and intensity of the emotion as experienced in our platform. A correlation study with other toolsets is also a possibility of further study.

8.4 eMoto – Emotionally Engaging Interaction

eMoto is an emotional text messaging service built on top of a SonyEricsson P900 mobile terminal. The goal of this service is to provide users with means to emotionally enhance their SMS-messages. The user first writes the textual content of the message and then adjusts the affective background to fit the emotional expression she wants to achieve. The adjustments are done through affective gestures that will render an animated background acting as an emotional expression to the user’s text message (Figure 3). The P900 terminal is used with a stylus pen. We have equipped this pen with two sensors that will recognize the affective gestures: an accelerometer and a pressure sensor. In a first prototype, the extended stylus is connected to the serial port of a stationary PC, which in turn communicates with the P900 terminal – in the final prototype this will be a direct wireless communication channel between the stylus and P900 terminal.

Through the eMoto-design we intend to emotionally engage users both cognitively and physically using a tangible interface allowing for affective gestures that are mirrored in the expressions produced by the system. A questionnaire sent to 66 potential users showed a need for richer emotional expressiveness in text messaging in mobile phones than available today. Emotions are expressed not only through what is said, but also through body gestures and tone of voice, mediums not available in this context.

In this specific design, our aim is to let users consciously express their emotions. This should not entail a simple one-to-one mapping of emotions to specific expressions. Instead we build the interaction on the fact that emotions should not be seen as singular, discrete states, but instead as processes that blend into one another. Through creating the interaction model based on Russell’s circumplex model of affect (Russell, 1980) we could create a system that allows users to choose emotional expressions that best suit their messages. Without explicitly naming each emotion in the interaction we maintain open interpretations of emotional expressions. In Russell’s model emotions are seen as a combination of arousal and valence. By combining two basic movements that together can render an infinite amount of affective gestures, the user will move around in this circumplex plane (Figure 4). The affective gestures are closely connected to the affective feedback the user gets as visual output. The characteristics of emotional expressions found in the analysis of body movements are represented through colors, shapes and animations in the design of the affective feedback. Colors are used to express arousal, where red represents emotions with high arousal and blue is calm and peaceful (Ittens, 1971). The shapes of the animated objects in the areas containing high arousal are small and can therefore render animations and patterns that are energetic, quick and spreading. Moving around the circle towards less energy and calmer expression, the shapes get bigger and more connected, rendering slower and more billowing animations. Shapes placed on the positive side of the circle are softer and more round, while shapes placed on the negative side are more angular and sharp. The emotional expressions are stronger along
the outer border of the circle while weaker towards the middle; this is represented through less depth in colors and fewer animated elements. (Figure 5). We intend to use novel types of methods, in particular, the sensual instrument, to evaluate and further refine the eMoto concept.

Figure 3 Affective message

Figure 4 The affective gestural plane

Figure 5 The affective background combined with Russel’s model of affect

8.5 I – Shadows

The aim of this use-case is to test a new approach for developing affective interactions. The context is storytelling and we aim to develop an application where affect plays an important
role both in the way it is expressed by the users (children) and in the way it is expressed by the system, thus creating an effective affective loop.

**Stories and children development**

Stories are part of us since early childhood. Young children construct their first stories by imitating their parents' daily routine or even their actions. By using toys as elements of their make-believe activities, children start to explore the novelty that surrounds them, and simultaneously confronting their own fears (Singer, 1994). Under such conditions, children afford to induce some variations in the flow of their imaginative play and through this they acquire knowledge about the external and mysterious world that surrounds them. Many psychologists and educators have studied and investigated the influence of narratives and stories during child development phases. It is now very clear that narrative can be an important vehicle to structure the knowledge and to help in the process of meaning making. From Malone’s perspective (Malone, 1984), narrative plays a central role in memory by providing an organising structure for the knowledge.

'It is therefore one of the goals of I-Shadows to promote story creation, creativity and allow children to develop memory, sense of perspective, and empathy, thus promoting the emergence of emotional intelligence.'

**Narrative theory and Learning Environments**

One of the major influences in narrative theorists of this century is perhaps the work of Aristotle on narrative analysis. In Poetics, Aristotle suggests that all literary works are imitations of the reality (Berger, 1997), and the purpose of any of such literary works would be to mimic what happens in real life, and since life is composed by actions, then all literary works should reproduce such actions. In Poetics, Aristotle defines as: “(...) an imitation (mimesis) of an action that is admirable, complete and possesses magnitude; in language made pleasurable, each of its species separated in different parts; performed by actors, not through narration; effecting through pity and fear the catharsis1 of such emotions”(Aristotle, 1996, p. 10). Inspired by this view of narrative, several theorists (Bal 1997) (Campbell 1993) (Branigan 1992) have developed their own theories which in turn have been used in interactive drama. However, while in traditional drama, the audience plays the passive role of watching a play, in interactive drama, the user plays an important and active role in the play, intervening by acting with the story characters or participating within the achievement of the story. Thus, to have interactivity we need flexibility in the way the story flows allowing to the user to influence such story. On the other and, stories need to keep some structure and some pre-defined flow of narrative, guaranteeing the climax of the story. These two, clearly opposite goals, lead to the Narrative Paradox which is a recurrent trap which most Interactive Narrative Environments (INE) have difficult to avoid. Furthermore, when collaboration is at the centre, guaranteeing some structure in the story seems to be an even harder task to achieve.

So, in I-Shadows we will address the following problem:

‘How can we promote collaborative story construction without restricting the flexibility, creativity and emotional expression of the children, and at the same time, guarantee some coherence in the stories produced?’

1 Katharsis can be interpreted as the purification, i.e., the spectator is exposed to emotional tension experienced by the tragic hero, which has a purifying effect on himself.
Affective Expression and the Affective Loop

Emotions can be expressed through sound, colour, shapes, forms, movement among other. Children usually express their inner states by carrying on make belief activities where the characters act emotionally and in context. Given this, we will build an affective application (the I-Shadows system) which will be used to effectively show some of the ideas and research directions behind this challenging topic. This small system will provide children, ages seven to nine, an interactive experience for collaborative story writting using chinese shadows.

![Figure 6 I-Shadows, creating stories with chinese shadows](image)

The system of I-Shadows will allow for a group of children to create stories, using a projector, a projected screen and a set of shadows previously made with coloured shapes (cuts of transparent paper see Figure 7 and Figure 8).

![Figure 7 A boy and a duck](image)
By physically manipulating the shadows, children can build stories on the screen and portray affective actions by the props and characters in the story. A camera placed above the children, will detect the motion and emotions of the shadows in the screen (made by the children) and will intervene, interactively, adding new elements to the story being created, such as new forms and shadows, affective sounds and even music. The elements created by the computer will allow for a story structure to be maintained and will guide the children in the collaborative story construction.

The audience, whom can be either placed behind the screen or also manipulating shadows (see Figure 4), will be able to see a play on the screen, which will be a result of the collaboration between the several children and the computer, in a collaborative and creative story writing experience.


Berry, D., Butler, L. and de Rosis, F. Evaluating GRETA. The importance of consistency of behaviour in a multimodal animated agent. Submitted.


Elliot Clark and Brzeinski, Jacek (1998) Autonomous Agents as Synthetic Characters, Communications of the ACM.


Haptek website:  http://www.haptek.com


Itten, J (1971) Kunst der Farbe, Otto Maier Verlag, Ravensburg


