Modelling "user understanding" in simple communication tasks

Heimo Müller
Medical University Graz
Stiftingtalstrasse 24, A-8010 Graz, Austria
+43-316-385-72916
heimo.mueller@meduni-graz.at

Fritz Wiesinger
University of Applied Science Burgenland
Thomas A. Edison Strasse 2, A-7000 Eisenstadt
+43-5-9010-60350
fritz.wiesinger@fh-burgenland.at

Abstract

We present an architectural model for adaptive interfaces based on eye-gaze patterns and facial expression analysis. In our approach, each basic visual sign can adapt its appearance and level of detail during the communication process. Atomic Communication Units (ACUs) – analogous to graphical output primitives – encapsulate the intended denotation, the encoding of the message and a method for the judgment of the communication goal. We have analyzed feedback cycles in human-human communication tasks, and propose applications scenarios for ACUs.

Categories and Subject Descriptors

H.5.2 [Information Interfaces and Presentation (e.g., HCI)]: User Interfaces – Theory and methods, User-centered design.

H.1.2 [Models and Principles]: user-machine systems – human factors.

Keywords

adaptive interfaces, eye-gaze patterns, visual language, mental models

1. INTRODUCTION

Today, usability engineers have quite a number of methods to evaluate the quality of an interface. If usability tests produce good results the design process is completed and the users and designers are happy. When, however, this is not the case, we have two choices:

(A) The user interface is redesigned according to the input from the usability tests and the evaluation is repeated, or

(B) users adapt their behaviour to the imperfect user interface (see Figure 1.)

Usually, alternative (B) is chosen, not because the user interface designers and the usability engineers are lazy, but because of the long delay in the feedback cycle (A) and the orientation of the user interface toward the lowest common denominator of all user requirements. However, this leads to a mismatch between the designer's intention and the users mental model.

The solution to this problem seems to be obvious: let the user interface adapt its visual appearance and functionality to the needs of the user. The difficult part in this feedback cycle is not the adaptation of the user interface itself, but the assessment of user satisfaction. Such an assessment – a usability test carried out by a machine – can be done either indirectly by analyzing interaction behavior (number of wrong clicks, search time, etc.) or directly by observing the user with a number of sensors.

We believe that most real world applications do not provide good adaptation because they are based on an indirect assessment which:

- reacts slowly (it needs a lot of user interface actions in order to give appropriate results),
- is not reliable or stable enough compared to a human usability test, and
- gives no results if the user is inactive.

We therefore propose a model of "user understanding" based on eye-gaze patterns and facial expression analysis.

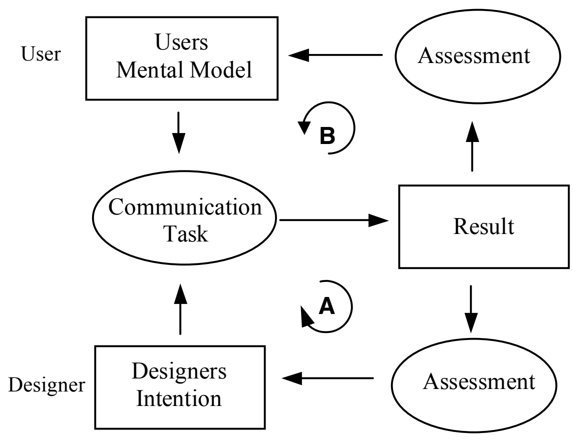

Figure 1. Feedback cycles in user interface design

With the help of an indicator for the understanding of the users, designers should be able to develop new types of visual communication objects which can adapt their visual appearance and level of detail according to the understanding of the user, in line with the statement of Heinz von Foerster "the hearer and not the speaker determines the meaning of an utterance" [1].

2. RELATED WORK

2.1 Intelligent Interfaces

A good overview of intelligent and adaptive interfaces can be found in [2] and [3]. Many applications of adaptive interfaces focus on information filtering and recommendation tasks, e.g. in content-based filtering and collaborative filtering applications, see [4]. Our approach focuses on intermediary interfaces because the interface itself will not be extended, but active communication objects will adapt their semantic depth (the level of detail, which is presented to the user) according to the cognitive requirements of the user.

All intelligent interfaces have, as central part, a user model [4] [5]. Extensions to the simple model of stereotypical user models are programmable user models [6], user models for demonstrational user interfaces [7] and comprehension based user models [8].

In a programmable users model the mental representations and the user behavior results in a cognitive model of the user. Using an Instruction Language (IL) the user interface designer describes the knowledge which a user needs to perform a specific task. The Instruction Language can be seen as programming, which is translated into a useable cognitive model. [6]

With the help of demonstrational user interfaces the user provides examples in the direct manipulation of interfaces from which the application generalizes and creates parameterized procedures or relationships. In addition to providing programming features, demonstrational interfaces can also improve the usability of direct manipulation interfaces, e.g. if the system predicts the operation that the user will perform next based on previous actions so that the user might not have to perform it. [7]

Comprehension based user models are based on the construction-integration (C-I) theory of comprehension [8]. The C-I theory was developed to explain how we use contextual information to assign a single meaning to words. A comprehension based user model uses knowledge representation about current tasks (world knowledge), about context independent declarative facts (general knowledge) and possible plans of action (plans element knowledge). Comprehension based user models, more specifically a Unix tutor system, a model for aviation pilot planning and a user model for army commander intelligence planning, are described and evaluated in [9].

2.2 Cognitive Systems

Almost 20 years ago Luy Suchman wrote “interaction between people and computers requires essentially the same interpretive work that characterizes interaction between people, but with fundamentally different resources available to the participants. People make use of linguistic, nonverbal, and inferential resources in finding the intelligibility of actions and events, which are in most cases not available and not understandable by computers” [10]. The interdisciplinary research field Cognitive Systems puts together theories of perception, communication, knowledge representation and reasoning in order to address the problem described above. Taking a cognitive systems approach the central questions in user models for intelligent user interfaces are [11]:

(A) How to capture the context behind the user interaction.

(B) How to increase the “richness of resources” available for user modelling applications from sensors and how to construct feedback cycles.

Several interdisciplinary projects are currently being carried out in the field of cognitive systems. Among them are e.g. the Network of Excellence (EU's FP6) “HUMAINE” (Human-Machine Interaction Network on Emotion); SIMILAR – The European task force creating human-machine interfaces similar to human-human communication; and the EU Integrated Projects (IPs) “COSY” (Cognitive Systems for Cognitive Assistants) and “AMI” (Augmented Multi-party Interaction).
In our approach, communication objects are adapted according to sensor inputs and primary eye-gaze patterns. Gaze-added interfaces using a probability algorithm und user model to interpret gaze focus are described in [12] and [13]. An interactive system (iTourist) sensing user’s interest based on eye-gaze patterns is presented in [14].

3. REFERENCE MODEL

In human-human communication the information flow is usually symmetrical. In human-computer interaction the situation is different, as the amount of information which is presented to the user is in most cases much higher than information gathered from the user, which typically consists of simple mouse and textual interactions.

In Figure 2, a reference model for the information flow between humans and an artificial system is shown. On the interaction surface, which can be seen as a border between the artificial system (computer) and the outside world, the information flow is maximal. Going “deeper” into the artificial system the information flow decreases and at the same time the semantic depth of the representation objects is increased.

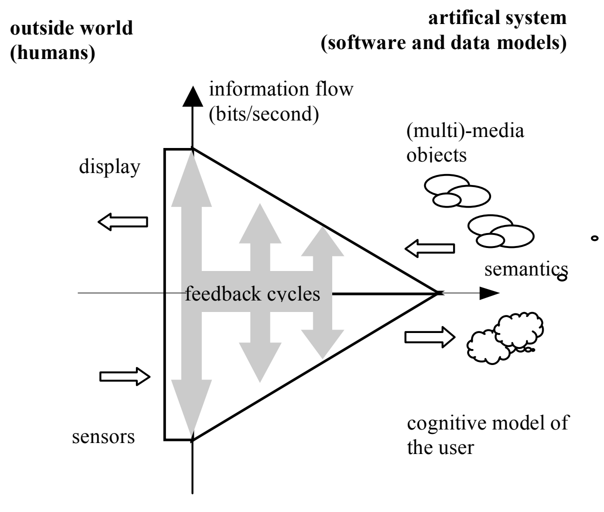

Figure 2. Reference Model

In the upper part of the figure, multimedia primitives are expanded by the rendering process and finally presented on a display or a multimedia device. Parallel to the rendering pipeline, sensor data is captured and analyzed in the lower part of the figure. Only when a rich set of sensor inputs, e.g. cameras, microphones, movement sensors is used, can the information flow reach the magnitude of the rendering process. However, the fusion and interpretation of the sensor data is a very challenging task, due to a high degree of ambiguity and – especially in high level analysis – context dependent semantics.

Following the structure of the human perception apparatus feedback cycles should be introduced at different processing levels. Such hierarchical feedback cycles are useful for the control of sensors, e.g. point of interest of a camera or initial filtering of the sensor data, and for a hypothesis driven analysis through the coupling of feedback loops.

Building on previous studies in the field of static and dynamic visual languages we have developed a model where each basic sign is able to adapt its visual appearance and level of detail. In order to achieve this goal we propose a new type of interface object, the Atomic Communication Unit (ACU). An ACU consists of

Intended Denotation	formal or natural language description of its mission.
Coding	representation of the intended denotation.
Semantic Inspection	method for the analysis of the receiver reactions.

The following examples illustrate the concept of an Atomic Communication Unit:

Visual Sign:	
Intended Denotation:	a car has to stop in front of the traffic sign.
Semantic Inspection:	movement sensor, or camera.

Visual Sign:	"hello"
Intended Denotation:	to greet somebody.
Semantic Inspection:	eye contact, to raise a smile, to reply to the greeting.

An ACU is modelled in an object oriented way, i.e. it has an internal state (data) and autonomous behaviour (methods), (see Figure 3) In order to achieve the overall goal – congruence between the intended denotation and the constructed denotation – the semantic inspection method adapts the presentation process (rendering, level of detail, presentation speed, additional explanations). The fundamental innovation of an ACU lies in the distinction between the semantics of a message and the used visual sign.

A semantic inspection method performs an analysis of the communication process in three constitutive levels:

Level 1: *The communication process was successful*
The receiver has seen the sign and the construction of the denotation has started. A semantic inspection method uses simple eye tracking systems at this level. [16]

Level 2*: The construction of the denotation at the receiver is finished*
The analysis of facial expressions and the body language are used as indicators for the completion of the interpretation task [17], [18] and [19]. This does not mean that the intended denotation is concordant with the constructed denotation.

Level 3: *The constructed denotation is concordant to the intended denotation*
In this case we can distinguish between (i) simple denotations, e.g. a command or question, where the fulfilment of the command deals as direct confirmation, and (ii) complex denotations, e.g. a part of a story. In the case of a complex denotation, concordance can only be measured in a wider context.

4. WORKSHOP INPUT

The following questions are the starting point for discussion during the workshop:

Q1 Is it enough to provide feedback mechanisms at level 1 and 2 in order to implement an adaptive visual communication process? The authors believe the answer is "Yes" for most of communication tasks. If the receiver has a completely wrong constructed denotation, the following communication steps will fail even at level 2 tests, if they are constitutive.

Q2 Can ACUs have a high complexity? i.e. complex and compound messages can be processed by our perception apparatus as one unit.

5. EXPERIMENTS

Several observations of human-human communication were recorded and analysed. In the experiments two people had to complete a specific communication task (T). One person acted as primary sender (S) and the other as receiver (R) of a message.

T1 *consultation hour*: a student (S) tries to postpone his examination. A professor (R) decides on the matter.

T2 *route description*: a person living in a city (S) asks a foreigner (R) to fetch a book from a bookstore and explains how to reach the bookstore.

T3 *insurance agent*: An insurance agent (S) sells a pension insurance policy to a customer

T4 *work permit*: a refugee (S) explains to his friend, also a refugee (R), how to get a work permit.

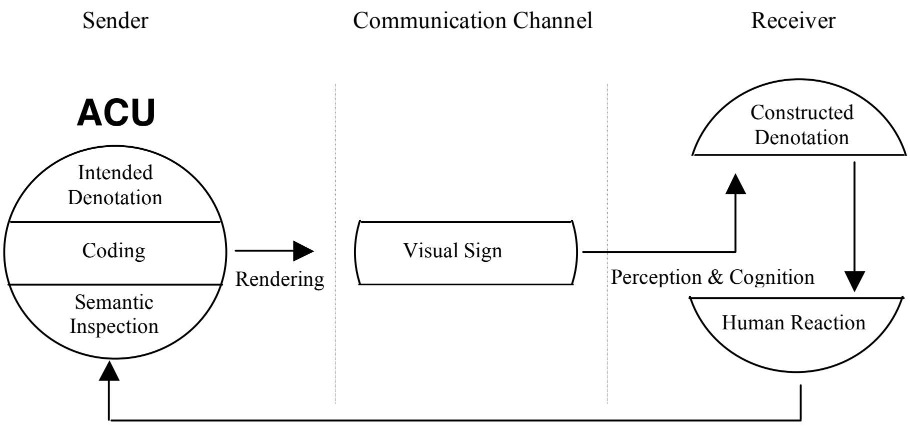

Figure 3. Atomic Communication Unit (ACU)

The following non-verbal signals were evaluated in the video analysis of the experiments:

A. The willingness to receive a message (attention level),

B. continuous message receiving signals,

C. take-over signals , switching from S to R or R to S,

D. assignment of a denotation,

E. concordant level.

The analysis showed that signals for A and B were predominantly eye contacts. Signals in category C were a mixture of eye contacts and verbal interruptions of the sender, and signals in D were mainly mimic and body signals (raising the eyebrows, nodding, affirmative noises). In category E, the communication took place on a higher level, by asking specific questions to test the concordance level. In the course of the communication tasks, the distinction between sender and receiver became blurred.

The direct conclusion for our research was that eye-gaze parameters embedded in a feedback loop should be prioritized as a sensor input for adaptive visual signs.

6. REFERENCES

[1] H. von Foerster, Cybernetics of Cybernetics (2nd edition), Future Systems, Minneapolis, (1996).

[2] P. Patrik, Intelligent User Interfaces – Introduction and survey – Research Report DKS03-01 / ICE 01, Delft University of Technology, (2003).

[3] E. Ross, Intelligent User Interfaces: Survey and Research Directions, Technical Report: CSTR-00-004, Bristol, (2000).

[4] P. Langley, User modeling in adaptive interfaces. Proceedings of the Seventh International Conference on User Modeling. Banff, Alberta: Springer, (1999), pp. 357-370.

[5] P. Brusilovsky, D. W. Cooper, Domain, task, and user models for an adaptive hypermedia performance support system, Proceedings of the 7th international conference on Intelligent user interfaces, (2002), pp. 23-30.

[6] R. M. Young, T. R. G. Green, T. Simon, Programmable User Models for Predictive Evaluation of Interface Designs, in K. Bice and C. Lewis (eds.) Proceedings of CHI'89 Human Factors in Computing Systems, ACM Press, New York, (1989).

[7] B. Myers, Creating User Interfaces by Demonstration, Academic Press, (1988).

[8] W. Kintsch, Comprehension: A Paradigm for Cognition, MA: Cambridge University Press, (1998).

[9] Y.W. Sohn, & S.M. Doane, Evaluating Comprehension-Based User Models: Predicting Individual User Planning and Action. User Modeling and User Adapted Interaction. 12(2-3), (2002), pp. 171-205.

[10] L.A. Suchman, Plans and Situated Actions, Cambride University Press (1987).

[11] G. Fischer, User Modeling in Human-Computer Interaction, in Human-Computer Interaction, User Modeling and User-Adapted Interaction Volume 11, Numbers 1-2 (2002), pp. 65-86.

[12] D. D. Salvucci, An integrated model of eye movements and visual encoding. Cognitive Systems Research, 1(4) 2001, pp. 201-220.

[13] D. D. Salvucci. J. R. Anderson, Intelligent Gaze-Added Interfaces, in Human Factors in Computing Systems: CHI 2000 Conference Proceedings. (2000), pp. 273-280.

[14] P. Qvarfordt, S. Zhai: Conversing with the user based on eye-gaze patterns, Proceedings of the SIGCHI conference on Human factors in computing systems CHI 2005, ACM Press (2005), pp. 221-230.

[15] B. Kiss, B. Benedek, G. Szijarto, B. Takacs: Closed Loop Dialog Model of Face-to-Face Communication with a Photo-Real Virtual Human, Visual Communications and Image Processing 2004 (EI25) SPIE Electronic Imaging, (2004).

16] M. Argyle, M. Cook, Gaze and Mutual Gaze, Cambridge University Press, London, (1977).

[17] P. Ekman, W. Friesen, Unmasking the Face. A guide to recognizing emotions from facial expressions, Palo Alto: Consulting Psychologists Press, (1984).

[18] S. Morishima, Face Analysis and Synthesis, IEEE Signal Processing Magazine, Vol.18, No.3 (2001), pp.26-34.

[19] M. Pantic, Facial gesture recognition from static dual-view face images, International Conference on Measuring Behaviour, (2002), pp. 195-197.