

Adaptive Visual Language Communication

Heimo Müller¹ and Herman Maurer²

¹ Medical University Graz, Austria

² Graz University of Technology and JOANNEUM Research, Graz, Austria

We present a model for adaptive visual language communication based on eye gaze pattern 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 analyzed feedback cycles in human-human communication tasks, and propose applications scenarios for ACUs.

Keywords: visual language, adaptive interfaces, eye gaze patterns

1. Introduction

Today usability engineers have quite a number of methods to evaluate the quality of an interface [1]. If such usability tests give good results the design process is finished and the users and the designer are happy, when this is not the case we have two choices:

- (A) The user interface will be redesigned according to the input from the usability tests, and we restart the evaluation, or
- (B) the users adapt their behaviour to the not perfect user interface.

Usually alternative (B) is chosen, not due to the fact, that the user interface designer and the usability engineer are lazy guys, 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 users requirements, see Figure 1.

The solution to this problem seems to be obvious: just 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 but the assessment of the user satisfaction.

Such an assessment - a usability test carried out by a machine - can be done either indirectly by analyzing the interaction behaviour (number of wrong clicks, search time, etc) or directly by observing the user with a number of sensors.

We believe that most real world word applications don't provide good adaptation, because they are based on an indirect assessment, which

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

We propose therefore control variables based on eye gaze pattern and facial expression analysis as an indicator for the understanding of a visual sign.

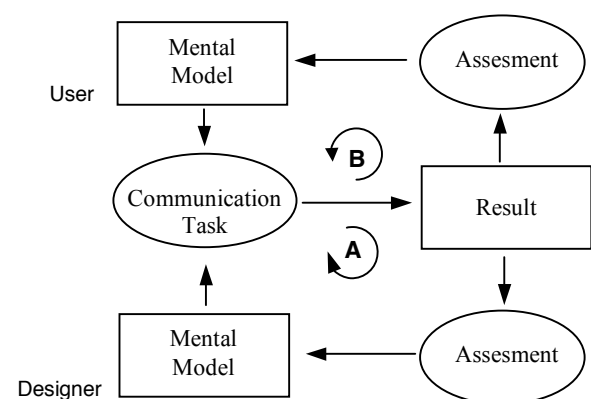


Figure 1. Feedback cycles in user interface design

With the help of such variables interface designers will be able to develop new types of visual communication objects, which are able to adapt their visual appearance and level of detail according to the understanding of the user, just in line with the statement of Heinz von Foerster "the hearer and not the speaker determines the meaning of an utterance" [2].

2. Related Work

2.1. Dynamic Visual Languages

Research in humanities and social sciences (psychology, linguistics, philosophy) indicates that people using visual information more than today would be more creative, exploit better the power of the human mind and probably could communicate across language-borders in a more intuitive way than is possible today with conventional text and natural language. Computer technology (information systems, telecommunication, visual tools) in turn promises to provide a wide range of highly effective tools to support visual, dynamic communication.

The underlying assumption of Dynamic Visual Languages is that by the beginning of the next decade computers allowing the presentation of high resolution moving images will be as ubiquitous as today's mobile phones. This will for the first time allow to implement a novel concept of communication systems that employs modern multimedia concepts (animations, movies, interactive pictures, dynamic maps, etc.) in combination with dynamic visual languages (i.e. visual languages in which symbols can change in shape, colour, size, etc. in time).

Putting it differently, there is no reason why information in the future should be recorded only using static text, static images and such, but rather it is conceivable to build on earlier attempts to construct new artificial languages that are not based on letters, but on icons. It was Otto Neurath [3] who showed with his isotypes that in many aspects symbols are superior to textual representations; a number of attempts to construct full communication systems based on symbols have been developed since. A brief survey of the main approaches is given in [4]. However, even in earlier papers such as [5], [6], [7], [8] and [9] the idea to use dynamic symbols, rather than static symbols has been discussed: the idea that symbols can change size, colour, shape, contours or can move and even change position to convey additional semantic meaning is rather appealing: why not e.g. use a symbol for 'eye' and to use the same symbol when slightly moving to indicate the verb associated with the noun, i.e. 'seeing'. This and other techniques as explained in above cited papers, techniques such as orthogonality, macros and using picture dictionaries that explain the same items in a variety of languages [10], changing the level of abstraction where desirable, etc. are the reasons why it does not seem far-fetched that written language, communication and interfaces as we now know them will be partially replaced by methods involving dynamic visual languages.

2.2. Intelligent Interfaces

A good overview about intelligent and adaptive interfaces can be found in [11] and [12]. Intelligent interfaces do not exist in isolation, but rather improve their ability to interact by constructing a user model based on the interaction. This brings the problem of intelligent interfaces close to the area of machine learning, where the user plays the role of the environment in which the learning occurs and the user model corresponds to the learning knowledge base. In such a scenario the interaction acts as performance task, on which learning should lead to improvements [13]. Many applications of adaptive interfaces focus on information filtering and recommendation task, e.g. in content-based filtering and collaborative filtering applications, see [13]. Intelligent interfaces can be divided into 3 classes [12]:

- (A) Adaptation within direct manipulation interfaces by adding extra interface objects for predicted future commands.
- (B) Intermediary interfaces : The nature of interaction is changed in order to act as an intermediary between the user and the direct manipulation interface.
- (C) Agent interfaces: In this case the user retains full control over the direct manipulation interface and is advised by an autonomous agent.

Our approach will focus 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 mental state of the user.

All intelligent interfaces have as central part a user model [13] [14]. Extensions to the simple model of stereotypical user models are programmable user models [15], user models for demonstrational user interfaces [16] and comprehension based user models [17].

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 to a runnable cognitive model. [18]

With the help of demonstrational user interfaces the user gives examples in the direct manipulation interface and the application generalizes from the examples 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 guesses the operation that the user is going to do next based on previous actions, so that the user might not have to perform it. [2.2.f]

Comprehension based user models are based on the construction-integration (C-I) theory of comprehension [2.2.g]. 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 a knowledge representations about the current task (world knowledge), about context independent declarative facts (general knowledge) and possible plans of actions (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 [2.2.h].

2.3. 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” [19]. 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 [20]

- (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 construct feedback cycles.

Several interdisciplinary projects are currently 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 taskforce 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 due to sensor inputs, primary eye gaze patterns. In [21] and [22] gaze added interfaces using a probabilistic algorithm und user model to interpret gaze focus are described. An interactive system (iTourist) sensing

user's interest based on eye gaze patterns is presented in [23].

In [24] the principle of a closed loop dialog model in face to face communication is introduced. Closed loop systems treat the recognition of users internal state as an active optimization problem in contrast to a passive observer model.

3. Reference Model

In human-human communication the information flow, consisting of primary information and nonverbal communication elements, 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 is typically made up of simple mouse and textual interactions.

In Figure 2 a reference model for the information flow between a human person and a artificial system is shown. On the interaction surface, which can be seen as 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.

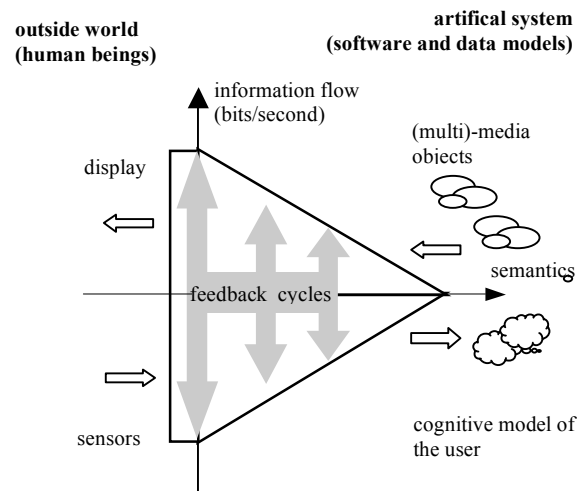


Figure 2. Reference Model

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 input, e.g. cameras, microphones, movement sensors, is used, the information flow can reach the magnitude of the rendering process. However, the fusion and interpretation of the sensor data is a very challenging task, because of the high degree of

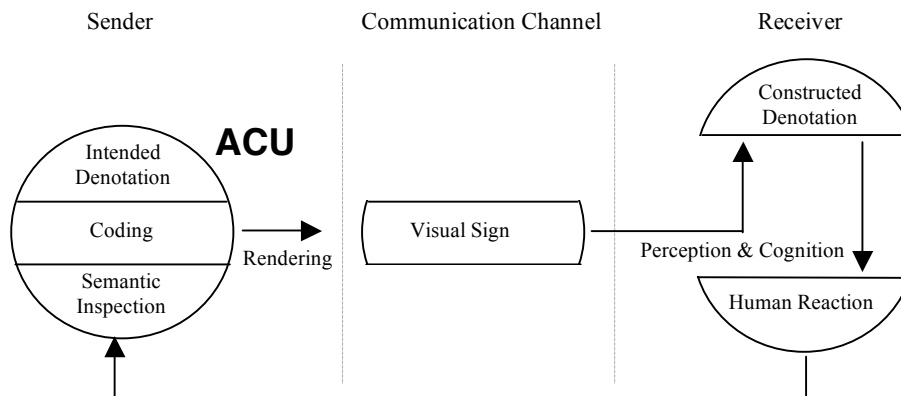


Figure 3. Atomic Communication Unit (ACU)

ambiguity and – especially in high level analysis – context dependent semantics.

Following the structure of the human perception apparatus feedback cycles are 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.

4. Atomic Communication Units

Building on a number of previous studies in the field of static and dynamic visual languages [2.1.a- 2.1.e], we 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

<i>Intended Denotation</i>	Formal or natural language description of its mission.
<i>Coding</i>	Representation of the intended denotation.
<i>Semantic Inspection</i>	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 a camera.

Visual Sign:

“hello”

Intended Denotation:

To greet somebody.

Semantic Inspection:

Eye contact, to raise a smile, to reply the greeting.

Visual Sign:

? V $\hat{\phi}$ 2 | .

Intended Denotation:

How are you? (BLISS symbols).

Semantic Inspection:

Eye contact, to raise a smile, answering the question.

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 does 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 at this level a simple eye tracking systems. [25]

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 [26], [27] and [28]. 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.

The following hypotheses are currently under investigation:

H1 It is enough to provide feedback mechanisms at level 1 and 2 in order to implement an adaptive visual communication process. If the receiver has a completely wrong constructed denotation, the following communication steps will fail even at level 2 tests, if they are constitutive.

H2 ACUs can have high complexity, i.e. complex and compound messages can be processed by our perception apparatus as one unit.

Within a feedback loop facial expressions and body language signals can be answered by an interface using “humanization objects” in its presentation system simulating e.g. eye contacts, staring, and advancing backwards.

5. Experiments

Several observations of human-human communication where recorded and analysed. In the experiments two persons had to solve a specific communications task (T). One person acts as primary sender (S) and the other as receiver (R) of a message. The sender got the instruction to explain the message either in a very

complex (C) or in a simple way (\neg C). The receiver of the message got the instruction either to commit himself to the communication process (F) or to give as little feedback signals as possible (\neg F). The following communication tasks were recorded:

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

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

T3 *insurance agent*: An insurance agent (S) sells an old age insurance to a customer.

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

In 8 experiments, two persons where recorded for all 4 communication tasks, each 5 to 10 minutes, on 5 video streams (face1, face 2, body1, body2, overall situation). For 2 specific persons the role (sender/receiver) was the same for all tasks, the instructions how to act as sender (C, \neg C) and as receiver (F, \neg F) was varied. The following non-verbal signals where evaluated in the video recordings:

- A. The willingness to receive a message (attention level),
- B. continuous message receiving signals,
- C. take-over signals, switching from S-R or R-S,
- D. assignment of a denotation,
- E. concordant level.

The analysis showed, that signals for A and B were predominantly eye contacts. Signals in categories C were a mixture of eye contacts and verbal interruption of the sender, and signals in D were mainly mimic and body signals (raising the eyebrows, nodding, affirmative grumbling). In category E, the communication was done 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, especially in the (\neg C,F) settings.

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

A simple simulation was done to show that a feedback loop in the early stage of the sensor data analysis can already contribute effectively to human computer interaction. In this experiment the orientation and distance of a display was adapted to a person in front of the display, simulating eye contacts, looking after somebody and adjusting the personal distance in conversations, see Figure 4.

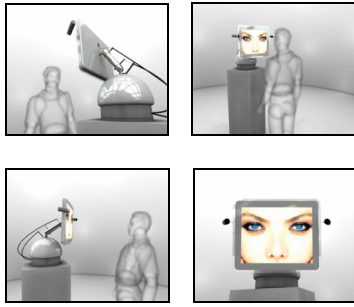


Figure 3. "A display that can see"

The orientation of the display is a simulation of looking at somebody, and can be controlled by the direct analysis of sensor data (cameras, distance sensors). This process is analogous to the control of the eye movements by the corpus geniculatum laterale (very early stage of the visual perception) in our brain.

6. Conclusion and Future Work

We developed a framework for adaptive visual signs and propose close feedback cycles already early in the processing of sensor data. The evaluation of human-human communication tasks gave valuable input for the design of these feedback cycles.

In the next step the hypothesis described in chapter 4 will be verified in real world applications. Application areas with good prospects for adaptive visual signs are e-learning applications, personal assistants and visual data mining.

In order to develop real world content using ACUs, it will be necessary to provide a simple API and/or a set of parameters as service for web based content development or popular (proprietary) applications as Flash or Director.

In a longer term we plan to develop a simple device for children of age 2-5 years, working title "grammar robot". Visual tokens and audiovisual sensors should control the device, which will perform a set of actions are triggered by simple communication goals known from toy robots. The goal of this experiment is to watch children in the development of a visual language grammar.

Acknowledgment

Our thanks are due to all partners of the EC Project SCALEX, especially Alexandra Preis and Martin Umgeher for their contributions and discussions.

References

- [1] A. Holzinger, Usability Engineering Methods for Software Developers, Communication of the ACM, Vol. 48, 2005, pp. 71-74.
- [2] H. von Foerster, Cybernetics of Cybernetics (2nd edition), Future Systems, Minneapolis, 1996.
- [3] O. Neurath, International picture language, University of Reading, 1978.
- [4] H. Maurer, R. Stubenrauch, D. Camhy, Foundations of MIRACLE - Multimedia Information Repository, A Computer-supported Language Effort, J.UCS Vol 9, No 4., 2003, pp.309-348.
- [5] J. A. Lennon, H. Maurer, MUSLI: A hypermedia interface for dynamic, interactive, and symbolic communication; J.NCA vol 24, (2001), 273-291.
- [6] J. A. Lennon, H. Maurer, Augmenting text and voice conversations with dynamic, interactive abstractions using P2P networking; J.NCA vol. 24 (2001), 293-306.
- [7] H. Maurer, P. Carlson, Computer Visualization, a Missing Organ and a Cyber-Equivalency; Collegiate Microcomputer, vol. 10, no. 2 (1992), 110-116.
- [8] J. A. Lennon, H. Maurer, MUSLI- A Multisensory Interface; Proc. ED-MEDIA'94, AACE (1994), 341-348.
- [9] D. H. Jonassen, R. Goldman-Segal, H. Maurer, Dynamicons as Dynamic Graphic Interfaces; Intelligent Tutoring Meida vol. 6, No.3-4 (1996), 149-158.
- [10] Duden- Oxford: Bildwörterbücher, Dudenverlag Mannheim (1994)
- [11] P. Patrik, Intelligent User Interfaces – Introduction and survey – Research Report DKS03-01 / ICE 01, Delft University of Technology, 2003.
- [12] E. Ross, Intelligent User Interfaces: Survey and Research Directions, Technical Report: CSTR-00-004, Bristol, 2000
- [13] P. Langley, User modeling in adaptive interfaces. Proceedings of the Seventh International Conference on User Modeling. Banff, Alberta: Springer, 1999, pp. 357-370.
- [14] 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.
- [15] 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
- [16] B. Myers, Creating User Interfaces by Demonstration, Academic Press, (1988)
- [17] W. Kintsch, Comprehension: A Paradigm for Cognition, MA: Cambridge University Press, (1998)
- [18] 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), 171-205.
- [19] L.A. Suchman, Plans and Situated Actions, Cambridge University Press (1987)
- [20] 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
- [21] D. D. Salvucci, An integrated model of eye movements and visual encoding. Cognitive Systems Research, 1(4), 2001, pp. 201-220.
- [22] D. D. Salvucci, J. R. Anderson, Intelligent Gaze-Added Interfaces, in Human Factors in Computing Systems: CHI 2000 Conference Proceedings. 2000, pp. 273-280
- [23] 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
- [24] 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.
- [25] M. Argyle, M. Cook, Gaze and Mutual Gaze, Cambridge University Press, London, 1977.
- [26] P. Ekman, W. Friesen, Unmasking the Face. A guide to recognizing emotions from facial expressions, Palo Alto: Consulting Psychologists Press, 1984.
- [27] S. Morishima, Face Analysis and Synthesis, IEEE Signal Processing Magazine, Vol.18, No.3, 2001, pp.26-34.
- [28] M. Pantic, Facial gesture recognition from static dual-view face images, International Conference on Measuring Behaviour, 2002, pp. 195-197.

Contact address:

Heimo Müller
Medical University of Graz
Stiftingtalstrasse 24
A-8010 Graz / Austria
heimo.mueller@meduni-graz.at

Hermann Maurer
Graz University of Technology and JOANNEUM
RESEARCH Graz
Inffeldgasse 16c
A-8010 Graz / Austria
hmaurer@iicm.edu
www.iicm.edu
