Aspects of Human-Machine-Communication

Manfred Lang University of Technology, Munich, Germany

CRIM/FORWISS Workshop, Munich, Sept. 5–7, 1994

Abstract

Efforts to improve the user interface of technical information systems of growing performance but also ever-increasing complexity become more and more important. Promising investigations include adaptive user interfaces and learning strategies in combination with image, natural language, speech and gesture processing.

1 Introduction

The integration of computer and communication technologies offers a continuously increasing number of services since the invention of Morse Telegraphy. Microelectronics and optical technologies in conjunction with advanced software solutions and new approaches to system architecture allow the results of research in modern computer science to be increasingly applied cost-effectively. Networked information and communication systems for voice, data, text, and image make high-quality multimedia and multimodal dialog and database access possible over long distances. Growth of technical communication services means also increasing information exchange at the human-machine interface. Humans of course are the most important part of any information system. Not only engineers and computer experts, but also users with different backgrounds and work tasks are concerned with computer operated systems. The community of potential users is growing rapidly. In order to meet the demand for future systems, we have to keep their complexity within justifiable limits. Hence, user-friendly, cooperative interfaces become a first priority development goal. This requires progress in adapting principles of information representation to human modes of communication, and to human cognitive limitations and objectives [6].

2 Human-Machine-Interaction

Essential guidelines in developing information and communication systems are: Improving performance capability, extending functional effectiveness, reducing costs per transmitted, stored, and processed information unit, and, with increasing importance, developing user adequate interfaces.

We humans prefer to communicate via spoken sentences, written text, and pictures, We are accustomed to recognize and to percept, to process and to interpret pictures, scenes, written and spoken text, music and noise with high performance not yet technically achieved. For the dialog with computers we use formal languages, and we operate the system with the help of keyboard, mouse, joystick, trackball, light pen, touch screen, graphic tablet, and other manual techniques. With respect to the importance of image and speech processing for men, technical

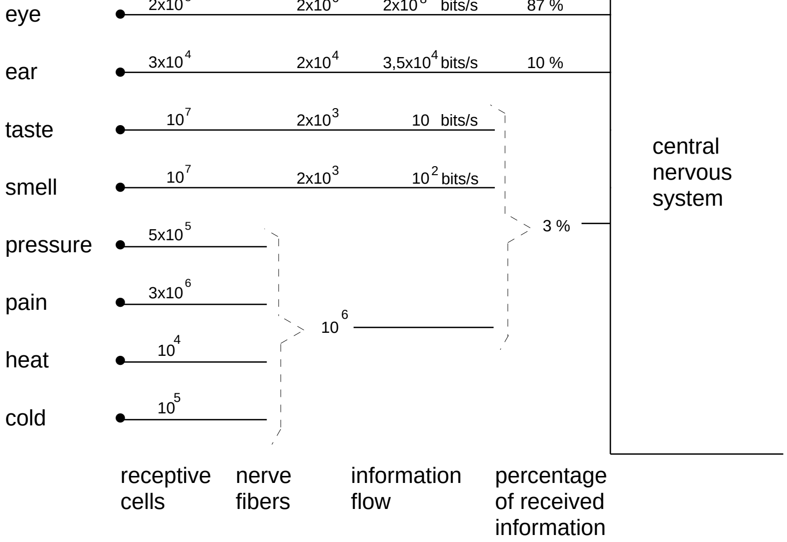


Figure 1: Receptiveness of sensory organs [2]

image and speech processing systems represent still a rather modest subgroup within the family of information processing systems. As we all know, there are a couple of reasons for that. Two important ones among them are: 1. Image, language and speech processing are hard to please, are costly with respect to required computing performance, and there are still open scientific problems to be solved. 2. Economic applications are strongly coupled to the progress in processor and computer technology.
In both areas there is promising progress! However, before transforming a task to the "computer world", a task which is formulated in natural language and with the help of pictures, we have not only to learn at least one computer language, but also to read voluminous user manuals. These barriers and obstacles could be reduced by:

- Replacing user manuals by computer implemented models and strategies successively adapting the technical systems to the user and his knowledge about the system and the task.
- Large scale application of natural information representations such as natural languages, speech, images, pointing devices for the human-machine communication.

Human-machine interfaces have to be adapted to humans sensory, motory, and cognitive capabilities. These are very flexible, but not unlimited with respect to acquiring, maintaining, retrieving, manipulating, interpreting different kinds of information. Some reference data were already given by Karl Kuepfmueller in the 1950th. With reference to the simplified sketch of fig. 1 our sensory system is capable to receiving up to roughly 10^9 bit/s, most of them by the optical channel, followed by the acoustic one. Unnecessary to remark that statistical bit numbers, don't say anything about the importance and relevance of the transmitted information.

As our central nervous system is able to consciously process only between about (10 to 100) bit/s, a considerable data reduction takes place between information reception and central processing. A well-organized memory and the ability to learn are decisive elements to appropriately cope with such an enormous data reduction. We correlate actually received sensory information with earlier learned and stored information. That is, simply spoken, the way how speech understanding and image interpreting works. This may be outlined by a very simple example: Many people recognize in a certain scenery not only, "there is a tree", but also, "this is an oak, about 150 years old, it is spring time, the tree is healthy ... ", and so on. The visually

Input:		*Output:*	
Keyboard	80 - 130 bit/s	Reading Text	150 - 300 bit/s
Handwriting	10 - 20 bit/s	”Reading” Pictures	ca. 10^6 bit/s
Speaking	80 - 200 bit/s	Hearing	ca. 10^4 bit/s
Compare: Scanning	ca. 10^6 bit/s	Compare: Laserprinter	10^5 - 10^6 bit/s

Table 1: Raw data rates for input/output devices

perceived signals obviously activate additional knowledge which was learned earlier and stored in the memory.

Information input in technical systems is usually done by manual or sometimes foot-operated control; occasionally by gesture or mimicry-based control, or even by speech. For information output visual, auditive, and occasionally tactile modalities are available. Table 1 shows some rough numbers concerning information transfer between man and machine.

Terminals are the equipments where the human-machine interaction takes place. With respect to subscriber numbers, telephony is the most dominant service; its worldwide annual growthrate is smooth, is about 4 %. Non-voice data, telefax and mobile services undergo a considerable increase in which the worldwide annual growth rate of date services is about 25 %. Organizing new broadband services such as videophone, video conference, high definition digital television (HDTV) will go on beyond the year 2000.

3 Some Research Goals

User-friendly human-machine interaction results to a large extent in combining tactile modi with visual, natural language, and perhaps gesture ones. Transitions between different modi should be possible, and the system may adapt alternatively to professionals as well as to beginners. The present state-of-the-art is still far away from this envisioned goal. Research efforts, however, approach it step by step.

The block diagram of fig. 2 shows interacting human and technical system forming a compound entirety. **Communication** normaly takes its course while a special **task** is being done. Examples for tasks are computer aided applications such as CAD, data base access, tutorial systems, cockpit operations, robot control, etc. With respect to communication there are well established methods such as question-answer dialog, menu selection, WYSIWYG and so on. There are also new methods under investigation such as natural language and picture processing interfaces, or multimedia dialog.

Interdisciplinary research activities are studying thoroughly humans sensory, motory, and cognitive capabilities including learning processes. Promising investigations are concerned with the development of adaptive user interfaces. Normally the user of a technical system has a model of the technical system its functions and operations in mind. He elaborates also a model of the task and the adequate problem solving strategies. Simply spoken: the existence of these 2 models characterises the user as being an expert. New attempts are providing the technical system with software implemented task models and user models. With the help of these 2 models the system may be able to adapt to the user and his knowledge about the system and the task. In the long run, also learning strategies will be modelled which means that the user may start as a beginner and end up as an expert.

In 1983 the seven-layered Open Systems Interconnection (OSI) reference model was approved as an international standard. It successfully opened worldwide communication between

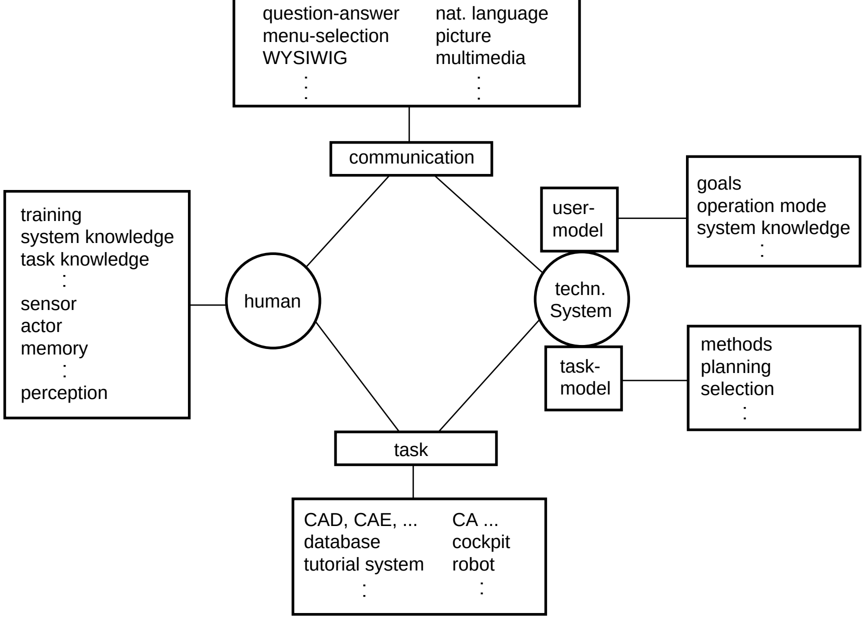


Figure 2: Human-machine Communication

systems of different manufacturers. In accordance with this reference model, several remarkable proposals describe human‘s decision behaviour, as well as the process controlling an adaptive human-machine interaction with the help of layered models too. A well-known example is J. Rasmussen's [10] three-tiered hierarchy which discriminates between skill-based, rulebased and knowledge-based behaviour.

Proposed 3-level communication models, for instance [8], [13], planning and evaluating of actions are embedded in a conceptual level, a dialog level, and an input/output level. Actions of the user occur top-down, and his reactions to the system‘s output bottom-up. The user may be able to propose, within certain limits, the system‘s response to his actions, and therefore to plan further actions in advance. On the other hand, as far as the system is equipped with a software implemented user model, future actions of the user will be proposed, appropriate objects activated, and required operations allocated. The purpose of this interaction model will be adapting system and user, and reducing transaction time. Similar models described in the literature use different numbers of levels.

In the laboratories of the Munich Institute for Human-Machine Communication we recently started first content steps in building adaptive dialog models for tutoring systems and we carry out acceptance tests using a closed control loop, see fig. 3. The system installed by my co-worker A. Obermaier [7] and several students consists of an user model and an adaptive dialog component [1]. Information about the user concerning the interaction is collected and used again in suitable situations. For instance, if the system recognizes some parameters often used, and so obviously preferred by an individual user, it may apply this knowledge to the dialog component and offer these settings when appropriate. In addition to this structure the Intelligent Tutoring Systems (ITS) needs a task model to give instructions and hints to the user, and a solution for the given task. This solution can be demonstrated step by step in visually animated form if needed. Another main feature of the ITS is a tool to integrate user's acceptance in the dialog. If some action is initiated by the system, it can never be proof against doing something unwelcome, surprising or confusing to the user. The acceptance component tries to get the feedback from the user to regulate the dialog, to correct or confirm the presumption taken by the user model. In this way we get a self-regulating, user-centered approach to fit user's requirements.

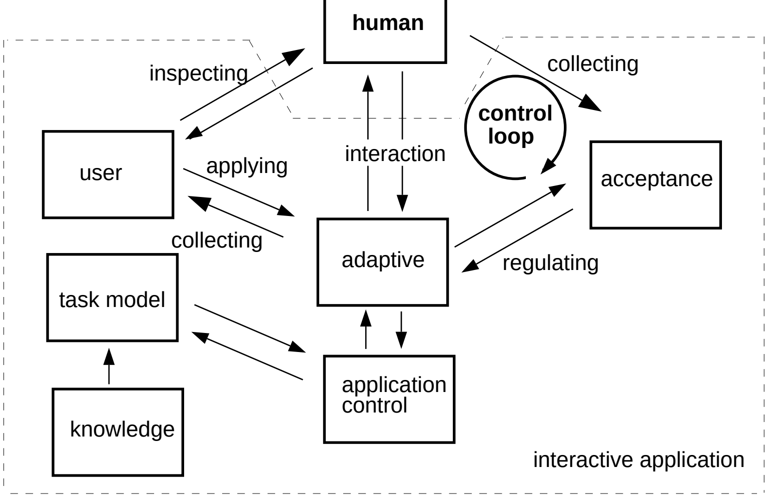


Figure 3: A closed control loop in an interactive application for acceptance tests

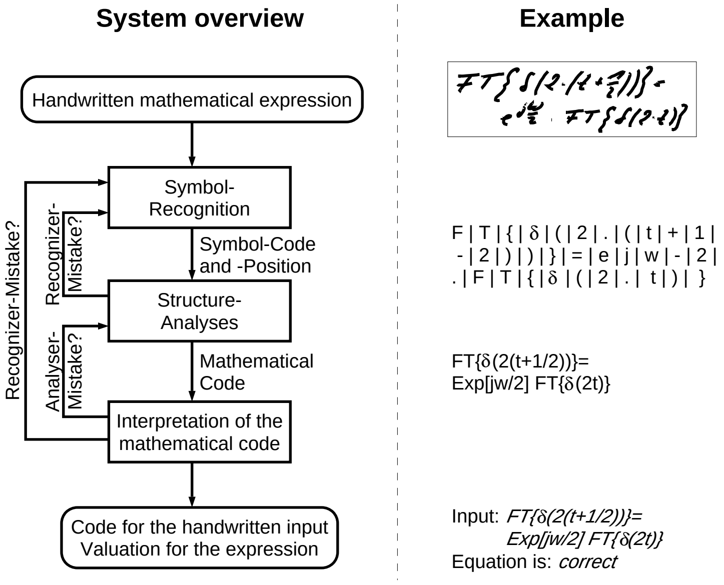

System overview
Example
Handwritten mathematical expression
Symbol-Recognition
Symbol-Code and -Position
Structure-Analyses
Mathematical Code
Interpretation of the mathematical code
Code for the handwritten input
Valuation for the expression
Recognizer-Mistake?
Recognizer-Mistake?
Analyser-Mistake?
F | T | { | δ | (| 2 | . | (| t | + | 1 |
- | 2 |) |) | } | = | e | j | w | - | 2 |
. | F | T | { | δ | (| 2 | . | t |) | }
FT{δ(2(t+1/2))}=
Exp[jw/2] FT{δ(2t)}
Input: FT{δ(2(t+1/2))}=
Exp[jw/2] FT{δ(2t)}
Equation is: correct

Figure 4: Recognition and interpretion of handwritten mathematical expressions, system overview and illustration by an example

The tasks in the running tests are part of our exercises and practical demonstrations, and the users are our students. A task, for instance, may be “analyzing a speech signal”. In this case the task model offers help in signal analysis, as far as it is necessary, and the user model follows the students actions recognizing when they need help. Acceptance is evaluated by observing the students, by measurements, with the help of questionnaires, and by taking into account already existing knowledge about the persons, and knowledge from preceding tests. The user model, which is under test for tutoring systems now, will be implemented for other applications too.

With the aid of modern pattern recognition methods and by using problem-oriented knowledge bases, computers will be able to analyse and interpret drawings, graphics and images. Automatic reading, interpreting and computer-controlled preparing of manually drawn diagrams or interpreting of forms and tables are illustrations of the research stage [3]. More recent investigations in our Munich laboratories are concerned with the recognition and interpretation of handwritten systemtheoretical expressions. A block diagram of the system implemented by my co-worker H.J. Winkler [12] and our students is given in fig. 4.

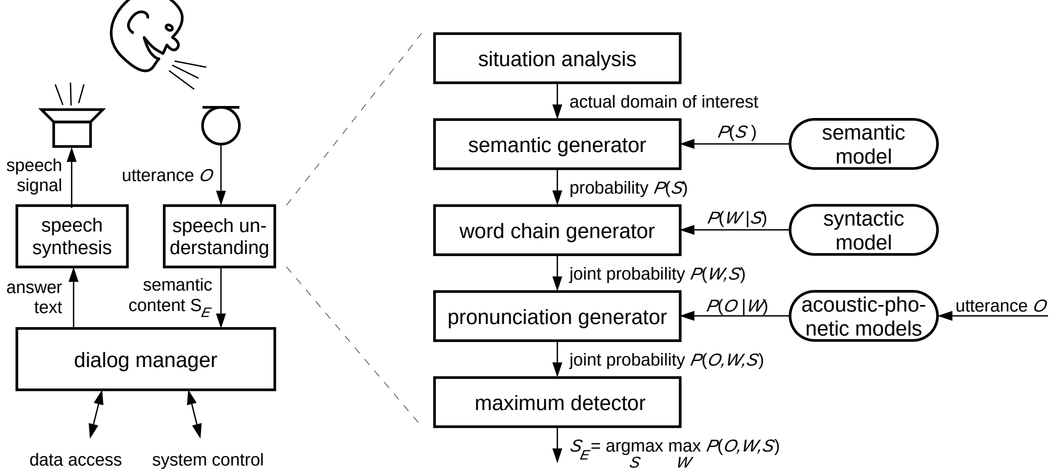


Figure 5: Block diagramm for spoken human-machine-dialog with the hierarchie of a system for extracting the semantic content S_E of an utterance O

In the first stage the handwritten input is segmented into single symbols and classified by a symbol recognizer system based on Hidden-Markov-Models. As recognition of mathematical formulas implies symbol recognition and structure interpretation, the relations between the expression's symbols are detected by analysing their relative positions. The result is a mathematical code (a one-dimensional line of text) for the handwritten input, which contains the complete two-dimensional (mathematical) information. If this transposition fails, a verification of the symbol recognizer results is carried out. In the last stage the mathematical code is interpreted relative to its syntactical and its mathematical-systemtheoretical correctness.

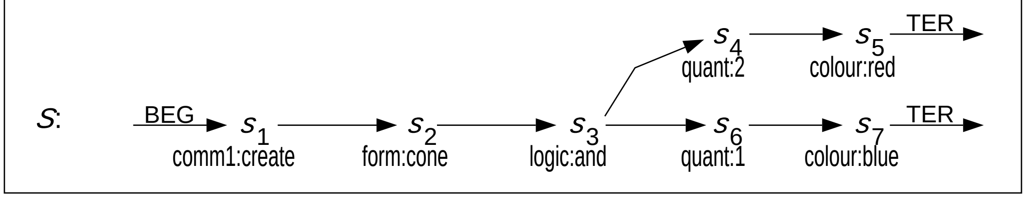
Natural language interfaces are matter of thoroughly but also controversly discussed investigations. Nevertheless, speech recognition and speech synthesis offer an especially user-friendly way for human-machine-communication. Relatively simple question/answering systems are already on the market. More sophisticated experimental systems which interpret fluently spoken sentences referring to a limited application scope, or even translation systems are still a matter of research and development efforts. The requirements to be met by future spoken dialog systems are very high. Speech recognition and speech synthesis have to be well integrated into the application, and the dialog management has to be adapted to user‘s communication habits. There is, however, remarkable progress in speech technology, partly depending on available and economically acceptable computing performance. There is also increasing commercial interest in this technology.

Speech research and technology are a central topic of this workshop with outstanding experts speaking. Therefore let me restrict myself on mentioning my co-workers H. Stahl and J. Müller's [11] investigations integrating speech recognition and language understanding in the framework of a pure stochastic model (fig. 5).

Within a given limited domain of interest, a semantic model generates possible semantic structures S, which are semantic representations close to the word level, and estimates its a-priori-probability $P(S)$. Referring to a given semantic structure S (exemplary shown in fig. 6), the syntactic model generates word chains W using hierarchical Hidden-Markov-Models [9], and calculates the conditional probability $P(W|S)$.

In the next step, using phonetic and acoustic models, the probability $P(W|S)$ of a recorded observation sequence O given a word chain W is calculated. The decoded semantic meaning S_E contained in the most probable combination of S, W and O is given by:

$$S_E = \underset{S}{\text{argmax}} \; \underset{W}{\max} \; [P(O \mid W) \cdot P(S)] = \underset{S}{\text{argmax}} \; \underset{W}{\max} \; P(O, W, S)$$



W_1 = 'zeichne zwei rote und einen blauen kegel'
W_2 = 'erzeuge einen blauen und äh zwei rote kegel'
W_3 = 'erzeuge doch bitte einen blauen und auch noch zwei rote kegel'

Figure 6: Semantic structure S consisting of seven semantic units (semuns) $s_1 \ldots s_7$. Each semun is represented by a pair of type and value, separated by a colon, and has a certain number of successorsemuns. Some examples of possible word chains W_i (in German language) emitted by the associated syntactic model are shown below.

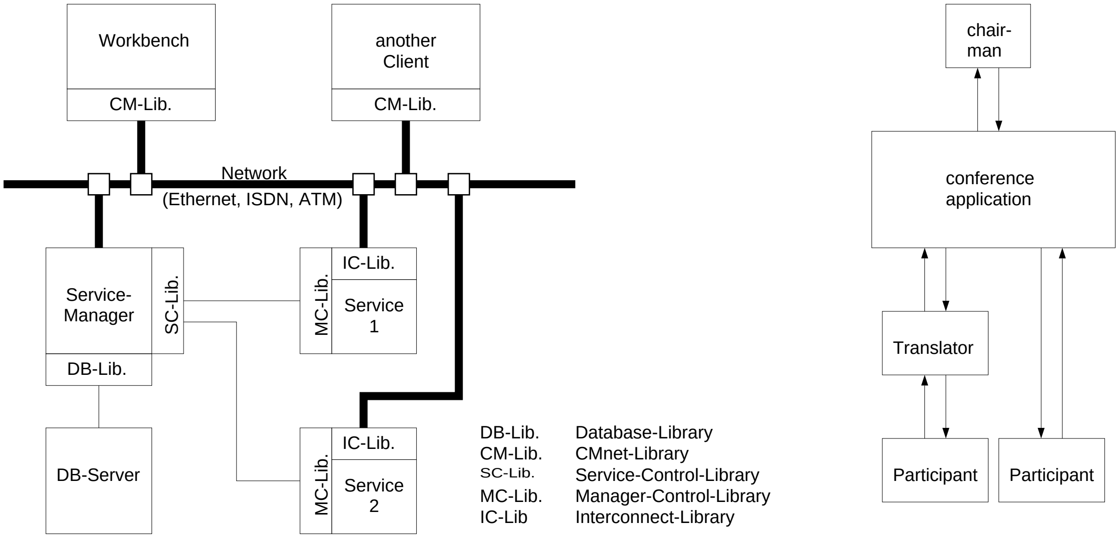


Figure 7: Distributed multiuser environment for continuous media applications

With the help of this equation, the extraction of semantic information will be carried out by pure stochastic methods.

Organizing useradequately a distributed multiuser environment for multimedia applications is an essential part in improving the human-machine-interaction. The block diagram of fig. 7 shows a client server approach which is under investigation in our laboratories.

The system installed by my co-worker R. Zwickenpflug and our students is based on independent services, which are spread over networked workstations. They communicate between each other via unidirectional channels, transporting continous data streams as well as command streams. One Server per workstation is responsible for managing all local services. It receives calls such as service start commands and connection requests from the clients.

A client's task is reduced on control actions. We distinguish between 3 different types of services:
1. Primitive services are typically used for audio and video signal handling tasks like microphones, loudspeakers, mixers and so on.
2. Interpreted services are responsible for user interfaces. They are based on public domain interpreters.
3. Compound Services are responsible for more complex systems such as conferences. They may also contain additional compound services increasing the complexity.

A service manager starts and controls different services for networked clients; the block diagram shows two out of a series of clients. The continuous-media-data are processed by decentralized service modules 1 to n. The right-hand block of fig. 7 demonstrates a typical conference application.

4 Outlook

Coming back to the development goals mentioned in the beginning, the future scenario may be:
– Performance, functionality, and cost/performance ratio of information and communication systems will continue to be improved according to technological progress.
– Actual challenges to scientists and engineers are: Adapting information systems of ever-increasing complexity to human's information processing and cognition capabilities.
– The question to be answered is: "How can we further support human's intelligence and creativeness by machine's performance?"
– The user of a technical system should be free to concentrate on the task he is going to master, and less on operating the technical system.

References

[1] T. Kühme: *A user-centered approach to adaptive interfaces*, Proc. 1993 International Workshop on Intelligent User Interfaces, Orlando, FL New York, ACM Press

[2] K. Küpfmüller: *Informationsverarbeitung durch den Menschen*, Nachrichtentechn. Zeitschrift 12 (1959), 68-74

[3] M. Lang: *Bild-, Sprach- und Wissensverarbeitung für den Dialog zwischen Mensch und Maschine*, ITG-Fachberichte 104, 1988

[4] M. Lang: *Kommunikation zwischen Mensch und Maschine*, mikroelektronik, vol. 5 (1991), no. 6, pp. 212-219

[5] M. Lang, H. Stahl: *Spracherkennung für einen ergonomischen Mensch-Maschine-Dialog*, mikroelektronik, vol. 8 (1994), no. 2, pp. 79-82

[6] M. Lang: *Towards User Adequate Human-Computer Interaction*, in B. Horvat, Z. Kacic (eds.) Proc. Workshop 'Modern Modes of Man-Machine Communication', University of Maribor, June 1994

[7] U. Malinowski, A. Obermaier: *Adaptivität und Benutzermodellierung - Was ist ein adäquates Modell?*, Workshop 'Adaptivität und Benutzermodellierung in interaktiven Softwaresystemen' im Rahmen der Tagung KI93, Bericht Nr. 30/93, Universität Konstanz

[8] D.A. Norman: *Cognitive Engineering*, in D.A. Norman, S.W. Draper (eds.) User Centered System Design, Lawrence Erlbaum Associates, Hillsdale, New Jersey 1986, pp. 31-61

[9] L.R. Rabiner: *A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition*, Proc. IEEE, vol. 77 (1989), no. 2, pp. 257-286

[10] J. Rasmussen: *Information Processing and Human-Machine Interaction*, North-Holland, New York, Amsterdam, London (1986)

[11] H. Stahl, J. Müller: *An Approach to Natural Speech Understanding Based on Stochastic Models in a Hierarchical Structure*, Proc. Workshop 'Modern Modes of Man-Machine-Communication', Maribor, 1994

[12] H.-J. Winkler: *Symbol Recognition in Handwritten Mathematical Formulas*, Proc. Workshop 'Modern Modes of Man-Machine-Communication', Maribor, 1994

[13] J.E. Ziegler, K.P. Faehnrich: *Direct Manipulation* in H. Helander (ed.), Handbook of Human-Computer Interaction, North-Holland, Amsterdam, New York, Oxford, Tokyo 1990, pp. 123-133