TOWARDS USER ADEQUATE HUMAN-COMPUTER-INTERACTION

Manfred Lang

1 INTRODUCTION

Human-Machine-Communication refers in general to an interdisciplinary matter. The following remarks restrict on engineering aspects as an overview which doesn't enter into too many details. Furthermore, recently started research efforts of the Munich University Institute for Human-Machine-Communication are outlined briefly. Computers and technical communication systems are the most complex machines in this connection. Humans, on the other hand, 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 increasingly 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 another 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.

We humans prefer to communicate via spoken sentences, written text, and pictures. We are accustomed to recognize and percept, to process and to interpret pictures, scenes, written and spoken text, music and noise with high performance not yet technically achieved. For the dialogue 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 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.

As we all know: 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:

1. 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.
2. Large scale application of natural information representations such as natural languages, speech, images, pointing devices for the human-machine communication.

This vision is a not yet realized goal, it is still a matter of research and development effort.

2 MAN AND COMMUNICATION TECHNOLOGIES

Overall guidelines in developing information and communication systems are:

- improving the <u>performance</u> capability
- extending the <u>functional</u> effectiveness

- reducing costs per transmitted, stored, or processed information unit
- developing user adequate interfaces

"Performance capability" with respect to computers means essentially processing speed and storage capacity. 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 dialogue and database access possible over long distances. The integration of computer and communication technologies is going on and it offers a continuously increasing number of services since the invention of Morse Telegraphy in 1837.

Fig. 1 shows the number of subscriber terminals of networked communication systems in logarithmic scale over the time axis. - 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 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.

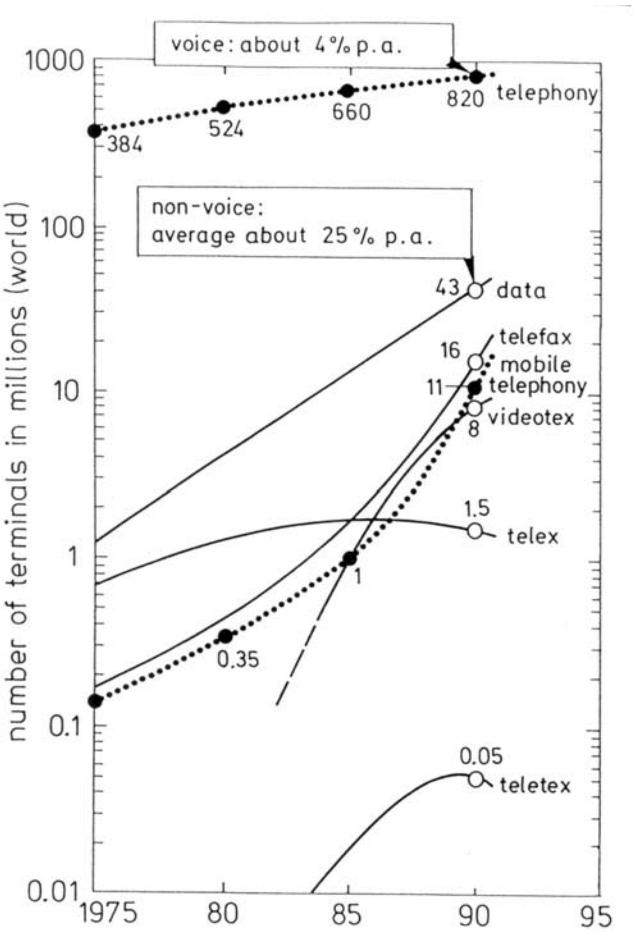


Figure 1: Growth of voice and non-voice communication (source Siemens AG [9])

Growth of technical communication services means also increasing information exchange at the human-machine interface. Human-machine interfaces should 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 50th. With respect to the simplified sketch of fig. 2 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, of course, don't say anything about the importance and relevance of the transmitted information.

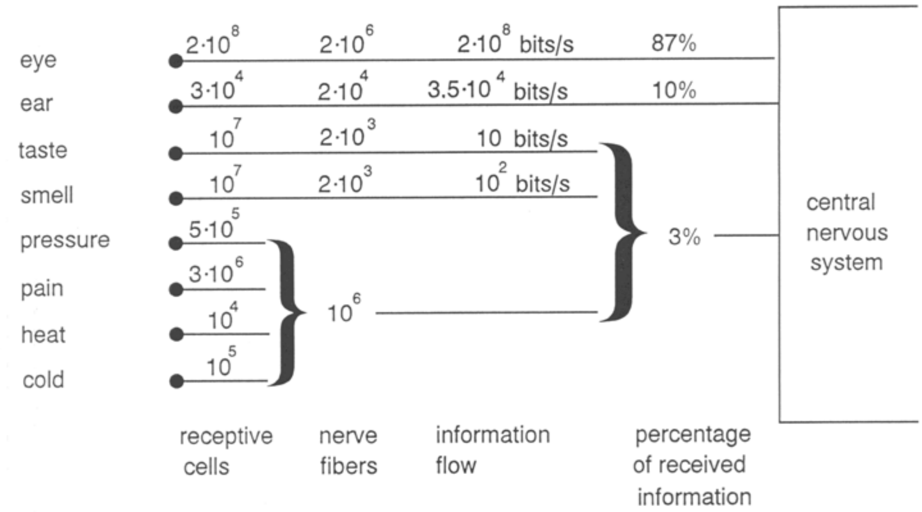


Figure 2: Receptiveness of sensory organs [2]

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.
Having these numbers in mind, let us come back to technical input/output devices (tab. 1) .

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

Table 1: Raw data rates for input/output devices

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. This table shows some rough numbers concerning information transfer between man and machine.

3 INSTITUTE FOR HUMAN-MACHINE-COMMUNICATION AT THE MUNICH UNIVERSITY OF TECHNOLOGY

The Institute for Human-Machine-Communication at the Munich University of Technology took its origin in 1992 from the well-known former Institute for Electroacoustics. Acoustics and psychoacoustics, of course, contribute essentially to human-machine communication, as the acoustic channel plays an important part of it. But in the meantime we use also optical and tactile modi in interchanging information between man and machine. Besides telephoning and broadcasting we retrieve and transmit text, graphics, images and scenes. We use television sets including coming HDTV-sets, we operate networked computer terminals with high resolution screens, plotters, and a couple of manual interaction techniques.

With reference to its history and to the above mentioned human-machine requirements the institutes main research topics are the following ones:

- Processing and interpreting of image, text, and speech.
- Adaptive dialogue models and user interfaces.
- Acoustic communication.
- Electroacoustics, psychoacoustics, technical acoustics including noise abatement.

The institute is provided with a computer network based on modern workstations, and it has noise protected rooms at its proposal for carrying out dedicated measurements.

4 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 following examples show already solved intermediate steps on the way to the long range goal mentioned.

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.

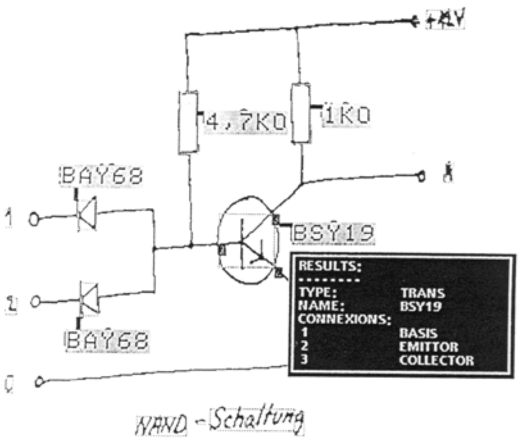


Figure 3: Automatically acquired hand-sketch [3].

Fig. 3 shows a simple hand-drawn sketch of a simple electric circuit diagram with text and graphic symbols. Text is automatically separated from graphics by computer analysis, line elements are traced, and defined points such as corners and branches are recorded. Typical components such as resistors, diodes, transistors, and their interconnections are recognized by matching them with models stored in the knowledge base. Finally we get a logical description of the circuit diagram. The outline can be stored as a standardized drawing, it can be output supplemented with specific explanations, retrieved from the knowledge base, as shown in the window of fig. 3, or it can be transferred to a CAD system for further processing.

More recent investigations in our laboratories are concerned with the recognition and interpretation of handwritten systemtheoretical expressions. A block diagram of the system is given in fig. 4.

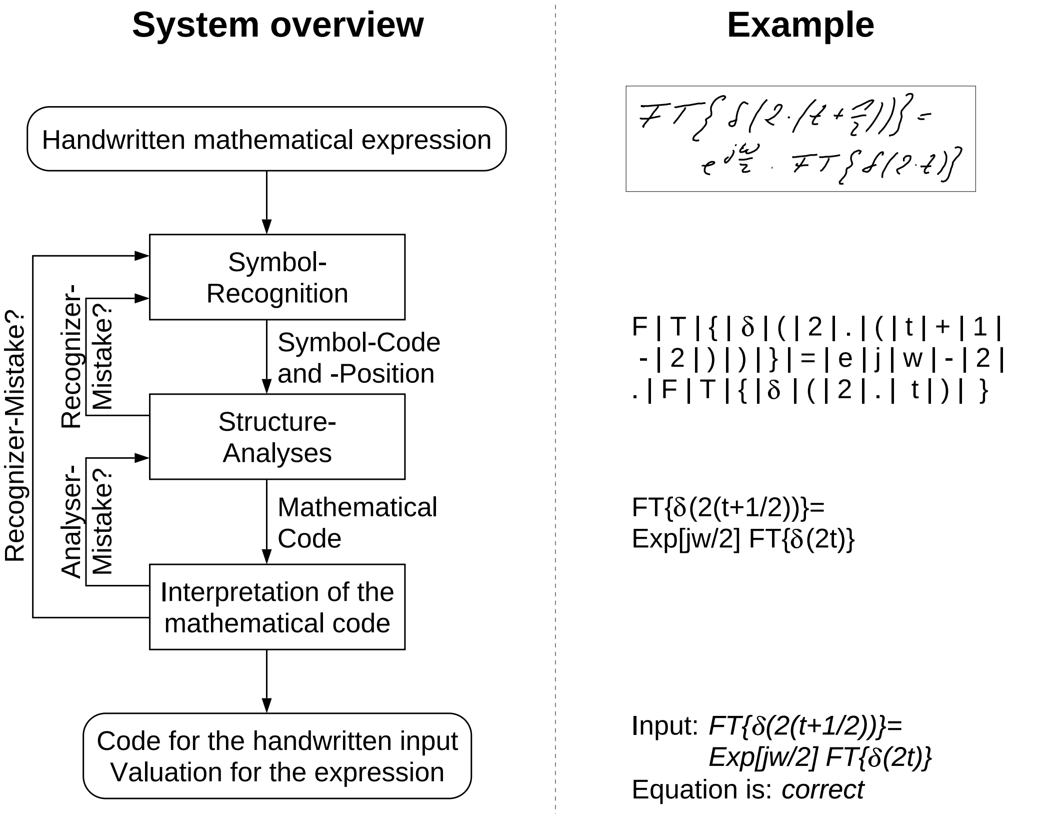


Figure 4: Recognition and Interpretation of handwritten mathematical expressions, system overview and illustration by an example

In the first stage the handwritten input is segmented into single symbols and classified by a symbol recognizer system based on Hidden-Markov-Models [8]. 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 [4]. A forecas for speech recognition products predicts a market growth from just under $200 million in 1992 to around $750 million by 1997 [11].

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 dialogue systems are very high. Their functionality may be roughly divided into four ranges:

- *Speech recognition* records spoken utterances and converts them into one or more word or sentence hypotheses.
- *Language understanding* interprets these word or sentence hypotheses and extracts the semantic content of the utterance usually by a rule based algorithm.

- The *dialogue manager* controls the incoming and outgoing information to and from a system or a data base.
- *Speech synthesis* converts an output text into natural speech.

In our present investigations, we try to integrate speech recognition and language understanding based on stochastic knowledge bases [7].

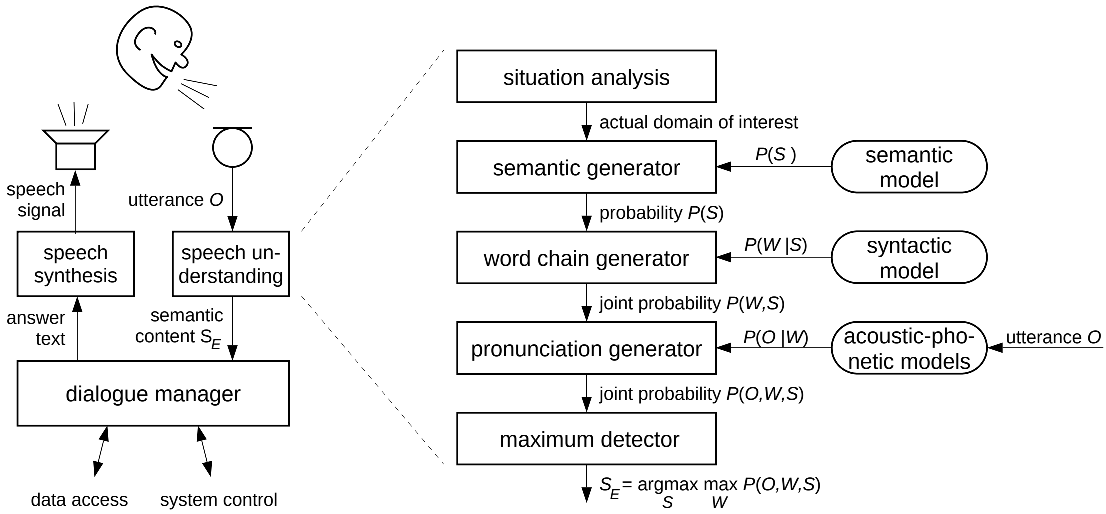
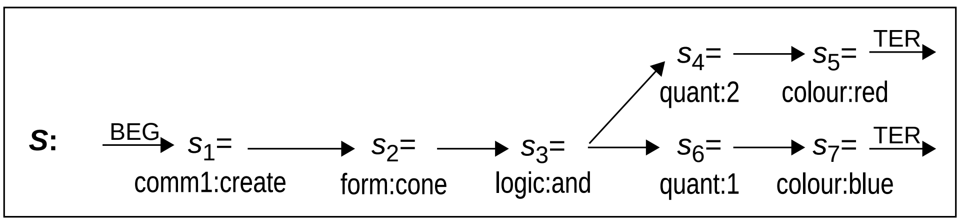


Figure 5: Block diagram for a spoken human-machine-dialogue with the hierarchy of a system for extracting the semantic content S_E of an utterance O

Within a given 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 [6], and calculates the conditional probability $P(W|S)$.



W_1 = 'zeichne zwei rote und eine blaue kugel'
W_2 = 'erzeuge eine blaue und äh zwei rote kugeln'
W_3 = 'erzeuge doch bitte eine blaue und auch noch zwei rote kugeln'

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

In the next step, using phonetic and acoustic models, the probability $P(O|W)$ 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:

$$S_E = \underset{S}{\operatorname{argmax}} \ \max_{W} \ [P(O|W) \cdot P(W|S) \cdot P(S)] = \underset{S}{\operatorname{argmax}} \ \max_{W} P(O, W, S) \qquad (1)$$

With eq. (1), the extraction of semantic information will be carried out by pure stochastic methods.

Promising investigations are concerned with the development of adaptive user interfaces (fig.7). Normally the user of a technical system has a model $U(S)$ of the technical system its functions and operations in mind. He elaborates also a model $U(T)$ 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 $S(T)$ and user models $S(U)$. 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 our laboratories we recently started first content steps in building adaptive dialogue models for tutoring systems and carry out acceptance tests using a closed control loop.

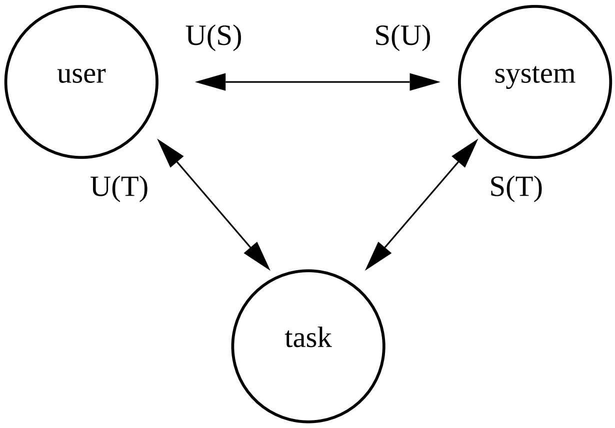


Figure 7: Models in user's and systems's mind

Following T. Kühme[1], an adaptation model in an interactive application consists of an user model and an adaptive dialog component. 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 to the user and, of course, hints and a solution for the given task. This solution could be demonstrated step by step in visually animated form if it is needed.

Another main feature of the ITS is a tool to integrate user's acceptance in the dialog [5]. 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.

In our case, the tasks are part of our exercises and practical demonstrations, and the users are our students. The 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.

Organizing useradequately a distributed multiuser environment for multimedia applications is an essential part in improving the human-machine-interaction. The block diagram of fig. 9 shows an approach which is under investigation in our laboratories.

Todays multimedia systems are limited in their network capabilities. Only some approaches [10] use a client server concept. The system we are designing is based on independent services, which

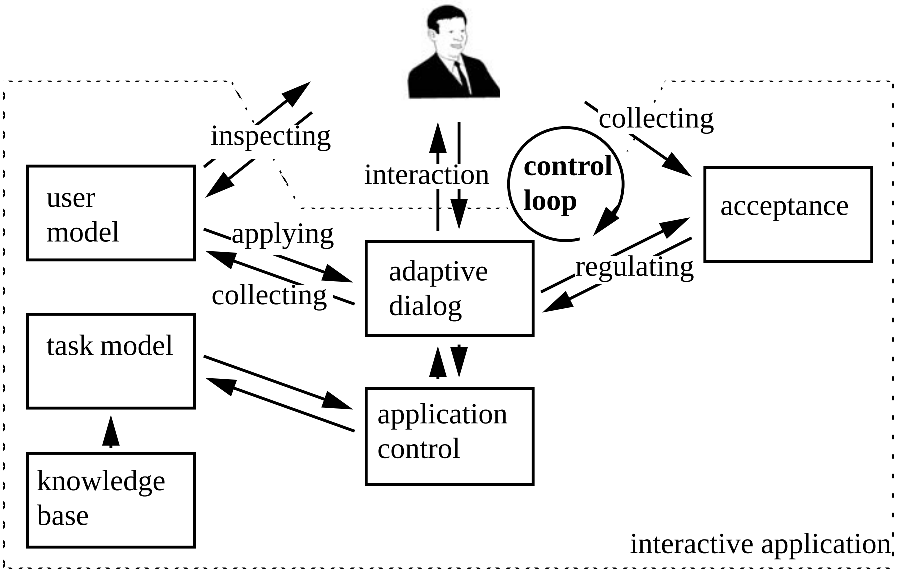


Figure 8: A closed control loop in an interactive application for acceptance test

are spread over networked workstations. They can communicate between eachother via unidirectional channels, that transport continous data streams as well as command streams. One Server per workstation is responsible for the management of all local services. It receives calls like service start commands and connection requestes from the clients.

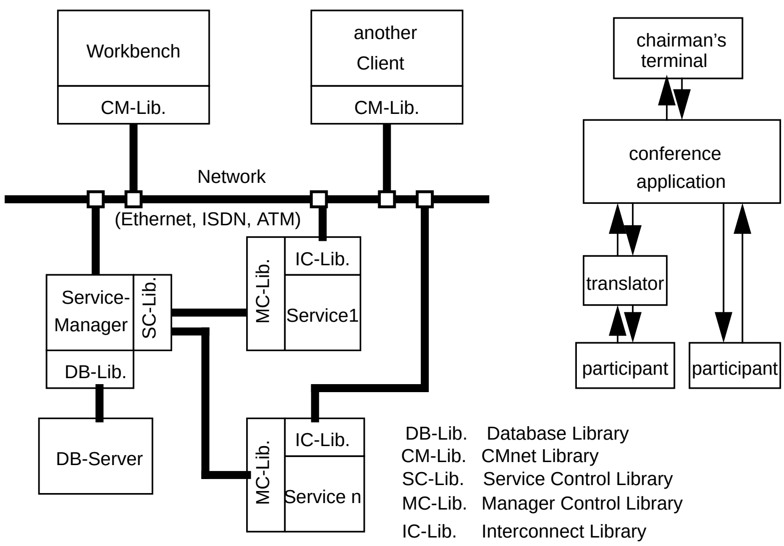


Figure 9: Distributed Multi-user environment for continous media applications

Therefore the task of a client is reduced on control actions. All data handling is done inside of the services. We use 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 used for user interfaces. They are based on public domain interpreters.

3. Compound Services are used for more complex systems like conferences. They can also contain other compound services to increase 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. 9 demonstrates a typical conference application.

5 OUTLOOK

Coming back to the development goals mentioned in the beginning, the scenario appearing on the horizon may be:

- Performance, functionality, and cost/performance ration 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.

6 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

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

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

[5] 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

[6] 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

[7] 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

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

[9] *The intelligent integrated broadband network - Telecommunications in the 1990s*, Siemens AG, ordering No. A30930-N2050-126-1-7618

[10] *AudioFile System,* Version 2, Release 2, Release Notes and Installation Instructions, DEC Cambridge Research Laboratory, 1993

[11] *Industry Watch*, Voice International, vol. 2 (1993), no. 6, pp. 20-21