A new paradigm of interaction for human controlled technical systems

R. Möller
Automation and Process Control Engineering, University of Wuppertal,
Dept. of Electrical, Information, Media Engineering,
Rainer-Gruenter-Str. 21, Bld. FC, 42349 Wuppertal, Germany
e-mail: r.moeller@computer.org

Recibido el 27 de octubre de 2004; aceptado el 25 de mayo de 2005

With today's multimedia computing technology, we are able to return to simple and intuitive control of technical equipment by using all major human senses in combination, that is, auditory, visual, haptic and olfactory modalities. This leads to a new paradigm of human-computer interaction. This paper will present the state of development and use of multimodal interactive user interfaces. Further topics for discussion will be aspects and issues of human engineering concerned with the different modalities noted in this abstract. Actual applications and research will be referenced.

Keywords: Human-computer engineering; virtual reality; multimodal sensory.

Con la tecnología multimedia de hoy en día somos capaces de volver al control simple e intuitivo del equipo técnico, combinando los sentidos principales del ser humano, es decir, las modalidades auditivas, visuales, tactilares y olfativas. Esto da lugar a un nuevo paradigma interactivo entre el hombre y el ordenador. Este documento presenta el estado de desarrollo y uso de los interfaces interactivos del usuario. Otros temas de discusión son los aspectos y aplicaciones de la ingeniería humana en lo que respecta las diferentes modalidades reflejadas en este extracto. Se hace referencia a las aplicaciones e investigación reales.

Descriptores: Ingeniería humana; interfaces del usuario; realidad virtual; sentidos multimodales.

PACS: 89.20.Bb; 89.20.Kk; 43.72.+g; 87.19.Dd

1. Introduction

Since the days when humans started using computers to control processes and equipment, they had to learn how to use and interact with a system that only understands numerical input and produces numerical output. The interaction raised a high level of abstraction and a loss of necessary "feeling" of how the real process works. The activities of man changed from mainly handcrafting, which demands all human senses and skills, to information processing, which mainly demands mental skills. During the phase of industrialization, the paradigm was to teach human "operators" how to "serve" or use machines for manufacturing or production. While the degree of machine and process automation has continuously increased, computer technology, control software, telecommunication technology and visualization systems have also continuously improved. Computer control can be found in all every-day environments, from embedded controllers in our car or home automation to intelligent clothing. This leads to a new complexity problem: users need "easy to use" interfaces to the equipment, adapted to human-sensory based skills. Today's multimedia techniques for reading, communication and presentation of machine or process data are well developed and can help to implement such interfaces, although most of them use only one human modality, that is, the visual sensory channel. There is still no universal and overall integrated solution available, but in order to achieve simple and intuitive control of any technical equipment, all abilities of human communication should be efficiently combined. The most important human compatible perception media consist of auditory, visual, haptic and olfactory modalities. Suitable communication media offered for a human-machine dialogue are, for example, speech, sound, gesture and mimics, text, drawings, static and moving images and realtime interactive graphics [9]. Today's developments tend to integrate multimedia information streams that use several different human modalities at the same time, in order to achieve high-bandwidth human computer interaction. All this leads to a new paradigm of human-computer interaction.

2. State of the art and ongoing development

2.1. State of the art

Multimedia user interfaces are well known from hypermedia learning systems, interactive multimedia catalogs or training simulators. The user can interact within the real or a simulated environment using the same interaction devices and procedures. Intelligent hypermedia-based assistants train and guide the operator during his work shift. The systems used to create such hypermedia applications change from special multimedia authoring tools to web authoring tools based on meta-languages like XML, SMIL or VML, for example[5].

During all phases of operation, but especially during implementation of a plant or machine control system, an ordinary *video conference* application can be used to improve understanding between operators, support and project engineers, if they are all at different locations. Beyond this, efficiency of error diagnosis and fault recovery can be much improved by using a video conference system that is configured for remote diagnosis and remote maintenance. Related systems based on computer controlled video conferences allow *application sharing*, that is, synchronous cooperative work with visualization of the same electronic documents.

Multimedia-based control systems are also of growing importance in automated production and manufacturing industries, although they are not new in process control engineering. Larger process control stations or control rooms contain video monitors and telephone besides regular media, that is, text, pictures and computer-based interactive supervisory and control systems. But there is no sequential control or synchronization between the different media. Today's workstation technology has changed this paradigm. Regular personal computers have the power and capacity to handle multimedia data as well as process or machine control with the same platform and at a low cost. They are supported by the newest technical development, increasing abilities and a good price-performance ratio of computer components that process static or moving images, speech or sound. This raises the question, how these components can be made available to supervisory, control, simulation, documentation and training in process or production control. Multimedia control room technology, consisting of large-screen-presentations, real-time video monitoring inserts, video conference inserts and interactive workstations is state of the art. Audio integration, on the other hand, for example noise from the process or machine or spoken warnings, is still at its beginning. It is well known that spoken warnings can help an operator in difficult situations. The presentation of audio information, especially the communication of noise and sound from a remote process or machine, can be of significant advantage for a precise remote-diagnosis [3]. Research is going on in this field [4].

2.2. New technologies

A very useful development based on multimedia technology is presented by *Virtual- and Augmented-Reality-Systems.* A VR-system merges a human user into a 3D projection of a locally constrained real scene, the virtual environment. An AR-System integrates virtual representations of real objects into an unconstrained real environment. AR systems can reveal, for example: presentations of measurement values, system-known information about hidden relevant objects, context-derived constraints and supplementary information. Both systems are based on the same display technology. An AR-user wears a head-mounted display which displays context-sensitive information blended into a presentation of his actual visual surrounding (Fig. 1). By optical or geo-positional tracking of his movements, an AR-system can guide a user through a specially marked complex surrounding. Interaction within the AR environment is possible with sensor equipped data gloves. Data gloves are well known as very precise and intuitive interaction devices for navigation and visualization in large data sets. They can be combined with force feedback elements in order to enhance visual feedback with tactile information [14] or just as input for gesture recognition. Although gesture recognition is seen as an important input paradigm for future applications in automation, it was found impractical to use data gloves. Simple one-hand gesture can be better obtained from camera based systems when the operator keeps his position within a small spatial region [1,2]. This is true for all cockpit-based control rooms.

Gesture and mimics recognition are, like many other modalities, part of biometric control systems. They are used for user identification, which is a key operation in access control for security-sensitive systems. Access control is also an issue in large distributed and mostly hierarchically organized automation systems. Operators and engineers with different responsibilities should also have different levels of access to a certain system. Input devices for eye (iris-) recognition and – much more accepted – fingerprint recognition are available for industrial environments. It can be expected that biometric input devices will soon become standard parts in complex industrial control applications.

3. Human factors

3.1. New paradigms for user interaction

During the seventies, the direct manipulation paradigm was introduced and led to an "increasing visual nature of computer interfaces" [12], sometimes called WIMP (Windows, Icons, Mouse, Pull-down menus) interfaces. Since then, the concepts of user interface design are driven by technical functionality requirements only. None of them clearly tackles the situation-dependent human abilities to perform a task or human mental constraints. It was found that this can lead to potentially dangerous system designs, especially if only visual interfaces are used [11]. Multimodal user interfaces therefore require special interaction strategies, because the information for all supported human senses must be coordinated in time, order and presentational form. On the other hand, they can help to reduce the complexity of interactive tasks because they can make human-computer interaction more natural and

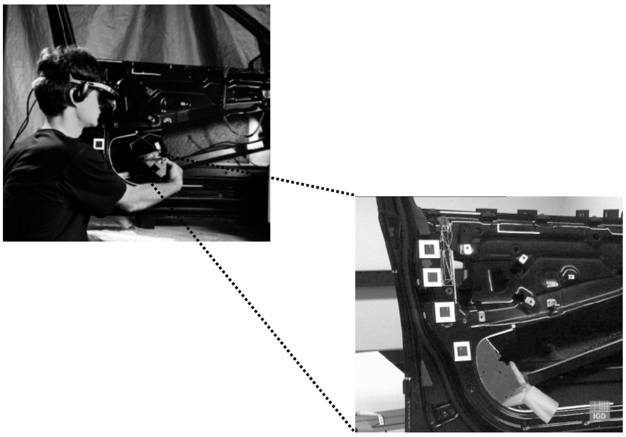

FIGURE 1. AR-guided system montage [8].

sometimes improve it to become invaluable, to elderly and handicapped people, for example, who often suffer from various losses in human performance such as visual accommodation, auditory sensitivity, weakened or distorted motor control, as well as the slowing down of information processing. Beyond this, it is proven that also ordinary users or operators of technical systems will benefit from multimodal interfaces if these are deliberately designed. The use of multiple sensual channels for human-computer interaction will increase the bandwidth of interaction significantly and thereby reduce the loss of information or delays.

With ongoing research, the difference between human-to-human and human-to-machine dialogues will soon vanish. Actual projects, for example CHIL (Computers in the Human Interaction Loop [16]), are focusing on a more sophisticated approach that will utilize intelligent communication agents, which can interpret, produce and understand time-correlated input and output modalities in real-time, according to individual human information processing abilities.

Today's research is concentrating on time-correlated multimedia modalities such as the speech, gesture and mimics, the last to a detail like movements of an eyebrow. Research with *talking heads* and *talking faces* **[**7**]** that simulate mimics has shown that the requirements for multimodal input and output are distinct from each other. Humans are extremely sensitive to slight misalignments between audible and visible speech, that is, the closing and forming of lips that "mirrors" the transferred information. Besides that, the effect of co-articulation needs to be addressed [6]: invisible geometric parts must be taken into simulation, *i.e.* palate and tongue. Computers, on the other hand, will only interpret the spoken sound. Gesture is a separate modality that can be used for control or additional information.

Multimodal interaction with cameras, headsets and data gloves is the first step toward freeing a user from a fixed desktop workplace. For some AR-applications, it is necessary to use *wearable computers.* Others use pen-computers or PDAs. The question arises how to utilize our computerized everyday-equipment such as digicam, mobile phone and PDA for *interaction* with our laptop, for example. The defining characteristic of ubiquitous computing is the change from the traditional desktop paradigm to a networked computing environment that unnoticeably surrounds the user, wherever he goes. Depending on his location or behavior, the user will have different dialogues with his computing environment. Admittedly there are some open issues to solve, namely granularity of integration, context-awareness and scalability of interfaces.

Technologies of biometric input, generally developed for security applications, are also interesting for multimodal human-computer interaction. Especially face, mimics and gesture recognition together with speech input and output are of growing importance for advanced user interfaces. Gesture and mimics are natural and individual for every human being. Today it is necessary to define unique and unambiguous gesture features, in order to use them for control applications. If a human operator learns these features, he can enhance his interaction bandwidth significantly. But natural gestures as well as facial expression are due to dynamic and unconscious movements that could produce unpredictable results, during process control for example. Research is therefore focused on prediction and online-analysis of human gesture as well as mimics. Simple applications for typewriting without keyboards or multimodal information kiosk interface are already available [13,15].

Beyond 2D interaction paradigms (WIMP), VR and AR applications require 3D-interaction elements. Menus and buttons are presented in 3D. Selection and manipulation happen to be done in space. There is no keyboard or mouse available; instead data-gloves are used for pointing and positioning of 3D objects and user navigation. Interaction metaphors consist of rotary tool choosers, ring menus and palettes, but all of them are of a visual nature.

3.2. Adaptive user interfaces

An important term in this context is *adaptivity*. An adaptive user interface deals with one or more of three different subjects, *i.e.* situation, user and task. Adaptation to a situation is marked by automatic adjustment of modality parameters. User adaptivity is concerned with the user's limitations. An intelligent agent is adaptive, able to learn and recognize user interaction strategies, and can thereby modify the basic user model. Task adapting agents will just be activated whenever a certain task starts, without any influence from or to the user. The most critical problems in user interfaces for automation are mental constraints and parallelism of interaction tasks. If an operator's instant visual and manual attention is interrupted by an unpredicted event, this can raise temporal shortcuts in mental or physiological resources. In such cases, mimics or gesture recognition or speech can enhance efficiency. Situation and task adaptivity makes it possible to substitute one modality with another, if losses of information due to user strain could be expected or the user just "feels more comfortable" with the substitute.

3.3. Speech and sound

From the technical point of view, spoken input and synthetic audio output, *i.e.* production of spoken output from phonetic representations, will be beneficial if distortions by background noise or affections to neighboring production processes can be excluded. Besides *speech understanding*, which has its focus on intelligent agents, search engines or automatic translation and can therefore improve human-computer dialogue at a high level of abstraction, *speech recognition* is generally used for human-computer interaction. Speech recognition is the projection of acoustic signals on written words or phrases as in dictation systems. Continuous speech recognition has an optimal recognition rate of 92 – 98%. This is unacceptable for surgery control or for tool control in a production workshop. A lot of parameters must be optimized before speech recognition can be applied to production control systems, for example acoustic environment, special vocabulary or operator training [10]. The simplest and most intuitive way, of course, is *interactive speech recognition*. As in real communication, the system provides feedback, and the user can enter a dialogue to correct interpretation errors. Advanced systems like these can supplement existing or replace other modalities. Simple *speech control* based on word recognition is often implemented in office systems and also in machine control. As there is only a limited vocabulary, a speaker-independent recognition rate of 100% is easily achievable, which is perfect for simple command driven systems. *Voice recognition* is not used to recognize what is said but *who* talks with the system. This can be used to implement hierarchical access structures in control. Issues are false acceptance or false rejection, which are both unacceptable for security applications. *Speech synthesis* could be useful for dialogue applications, although it has a major drawback compared with the replay of prerecorded human speech: there is as yet no automatic method to determine which word or part of a sentence must be emphasized in a certain context. Non-speech *sound* is well known in user interface design. It is implemented as auditory icons or earcons. Auditory icons use natural sounds to represent different types of objects and actions in the interface, for example files arriving in a mailbox producing a sound like a real letter would do. Natural sounds are intuitive but must be learned, anyway. Earcons are alike but synthetically generated sounds. An experienced operator, for example, will intuitively recognize the failure or maintenance state of a transmission system, if he can hear the noise or sound directly from the process. Acoustic data obtained from microphones applied to a certain automated system can be simply used and understood as a new brand of sensors or actors that are part of a process control unit. Visual data can be integrated in the same way into a multimedia control system. Supervising and monitoring of certain process states is simplified, if the operator can "see" and "hear" the real process.

4. Conclusion

It has been shown that (and why) multimodal interfaces are of growing importance not only in everyday life but also in automated production and manufacturing industries. Although attainable failure rates are not acceptably low enough today, the technologies of biometric input devices and speech control are promising from the technical point of view. Recorded and synthetic audio output, *i.e.* production of spoken output from phonetic representations, will be beneficial if distortions by background noise or affections to neighboring production processes can be excluded and a method ensures that user attention is not lost for other important modalities. For technical supervisory and control tasks audio- and video-"sensors" will be highly beneficial, as users will get back the "feeling" of controlling a real process. Concerning cost, one can say that most of the regular control and supervisory tasks can be done with rather simple workstation equipment. But the extra cost for realtime-multimedia and virtual reality extensions is worthwhile if balanced against benefit and customer satisfaction.

1. S. Akyol, L. Libuda, and K.-F. Kraiss, *Multimodale Benut-zung adaptiver Kfz-Bordsysteme*, in: K.P. Timpe, T. Jürgen-sohn, Kraftfahrzeugführung (Springer-Verlag, Berlin, 2001).
2. A. Wu, M. Shah, and N. da Vitoria Lobo, *A Virtual 3D Blackboard: 3D Finger Tracking using a single camera*, Fourth IEEE International Conference on Automatic Face and Gesture Recognition, Proceedings, Grenoble (France, 2000).
3. M. Rauterberg, *Different effects of auditory feedback in man-machine interfaces*, in: N.A. Stanton and J. Edworthy (eds.), Human Factors in Auditory Warnings, Ashgate Publishing (1999) pp. 225-242.
4. *Multimedia Process Control Room*, German Research Foundation (DFG) project: http://www.imat.maschinenbau.uni-kassel.de/forschu/deutsche.html#multi (1997-2010).
5. W. Mueller, R. Schaefer, and S. Bleul, *Interactive Multimodal User Interfaces for Mobile Devices*, 37^{th} Int. Conf. on System Sciences, IEEE Proceedings, 2004.
6. P. Cosi *et al.*, *Labial Coarticulation Modeling for Realistic Facial Animation*, in: Fourth IEEE International Conference on Multimodal Interfaces (ICMI'02), IEEE Proceedings (2002) p. 505.
7. CSLU Toolkit V. 2.0 (http://cslu.ece.ogi.edu/toolkit/), Center for Spoken Language Understanding, Oregon Graduate Institute of Science and Technology, 2004.
8. http://www.arvika.de, *Augmented Reality for Develop-ment, Production and Service*, bmb+f lead project (German Ministry of Education and Research).
9. R. Möller, *Relevanz computergraphischer Methoden für die Automatisierungstechnik*. VDI Fortschrittberichte, Reihe 20, Nr. 171, VDI-Verlag, Düsseldorf, 1995.
10. R. Möller and R. Karger, *Multimediatechnologie – optischer und akustischer Kanal*, Proceedings of VDE/GMA Kongress, Stuttgart, 1998.
11. *Mensch-Maschine-Sytemtechnik*, K.-P. Timpe, H. Kolrep, T. Jürgensohn (ed.) (Symposion Publishing, 2002).
12. B. Shneiderman and C. Plaisant, *Designing the User Interface: Strategies for Effective Human-Computer Interaction*, Addison-Wesley Longman, 4^{th}. ed., 2004
13. C. Maggioni and H. Röttger, *Virtual Touchscreen - a novel User Interface made of Light - Principles, metaphors and experiences*, Proceedings of HCI International, Volume I: Ergonomics and User Interfaces (1999) p.301.
14. M. Lin and K. Salisbury, *Comp. Graph. App.* **24/2** (2004) 22.

15. E. Mäkinen and R. Raisamo, *Real-Time Face Detection for Kiosk Interfaces*. Proceedings of APCHI 2002, Vol. 2, Science Press, Beijing, China, p. 528.

16. A. Waibel *et al.*, *CHIL: Computers in the Human Interaction Loop*, 5th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), Proceedings (2004), Lisbon.