

COGNITO - AN EXPERIMENTAL VOICE-CONTROLLED TELECOMMUNICATION SYSTEM

U. Schultheiß, B. Lochschmidt

Research Institute of the Deutsche Bundespost TELEKOM
Am Kavalleriesand 3, 6100 Darmstadt, Germany

Abstract

The available technology may enable many handicapped people to make use of telecommunication equipment and services. In this field, modern speech processing technology plays an important part. As a first step, it is necessary, in this context, to develop appropriate concepts for equipment and services which fulfill the needs of a maximum number of handicapped persons and, at the same time, promise economic success. For these purposes, a PC-based experimental system named COGNITO has been developed. Seen from a functional viewpoint, the present COGNITO version constitutes a telephone with advanced functions all of which are controllable by the human voice.

1. Introduction

In Germany alone, the number of persons subject to motor or sensory handicaps is estimated at several hundred thousand. Moreover, there is an unknown number of people subject to mental and multiple handicaps. These include a large group of persons who, at present, frequently have trouble in using telecommunication equipment or need the support of attendants although it is, above all, the telephone which serves as a very important means of communication and organization for the disabled. In addition, we have to bear in mind that the envisaged integration of video, speech and data communication offers the chance of rehabilitating and integrating handicapped persons provided that they are enabled to operate the systems transmitting this service. The development of user-friendly types of operation is, therefore, a real milestone to be set for handicapped and elderly

people on the road towards a more self-sufficient and less isolated way of life. Other important and economically justifiable reasons for intensified activities in this field are the improved possibilities of rehabilitation. In this context, we should also mention related European projects such as COST [1], ESPRIT and RACE.

2. Use of Speech Processing Systems

Various attempts have been made to design telecommunication terminals that can be operated by disabled persons. Examples are external devices and/or modifications introduced in the telephone services such as larger keys with adjustable activation force, goose-necks for microphones and special receiver amplifiers. For many disabled persons, however, these adaptations are inadequate because they do not allow terminals or services to be used at all or, at least, in fairly convenient ways. A large group among these people may nevertheless be assisted by modern speech processing technology serving as a link between speech and data communication.

3. Useful Concepts for Terminals Tailored to the Needs of Handicapped Persons

For the design of a terminal tailored to the needs of the handicapped, it is necessary to elaborate a structural concept. In doing so, we have to bear in mind that it does not suffice to develop some kind of functional laboratory model. Our concepts must indeed promise economic success. In addition to technical features, such concepts have to include aspects of marketing and customer information.

As far as the technical concepts are concerned, it is impossible, as a matter of course, to develop a terminal

uitable for the disabled whose basic design meets the requirements of all handicapped individuals. It is simply a fact that there are too many different groups of disabled persons. It is, however, neither reasonable nor economically promising to simultaneously develop and take to the market a large number of different special solutions. The solutions to the problem just referred to may consist in two approaches we will call "interface solution" and "PC solution" in the following.

a) "Interface solution"

For this purpose, it is necessary to equip conventional telephone terminals with a defined and, if possible, standardized interface allowing special external devices to be connected to the terminal (see Fig. 1a). Such interfaces have to make possible the connection of, for instance, voice-controlled dial devices and speech output systems as well as Braille displays, suck-blow switches, and, possibly, special loudspeakers and microphones.

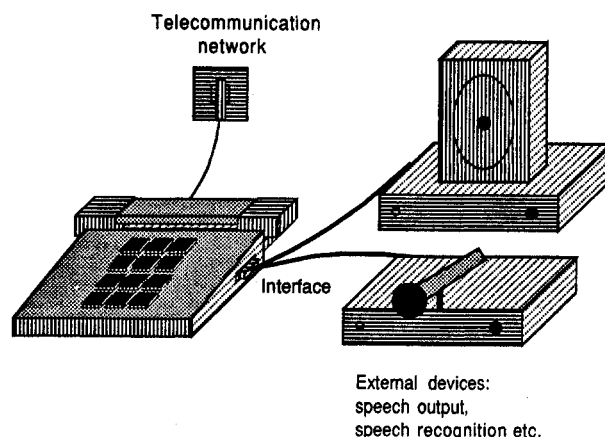
b) "PC solution"

The basic structure used is a universal type of hardware such as provided by a PC. The PC is complemented by a telephone interface and, possibly, a modem or ISDN board and, on the other hand, by devices like a speech recognition board and a speech output system as well as a loudspeaker and a microphone (see Fig. 1b). The interaction among these components and with the PC will be determined by the software only.

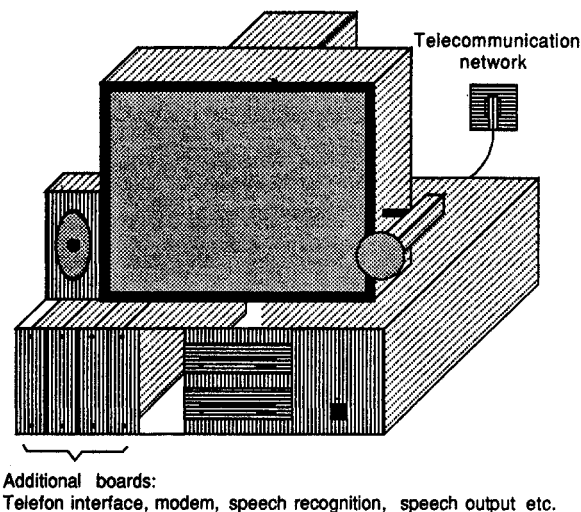
In view of the high degree of flexibility and modularity, we have decided to pursue, for our experimental system COGNITO, the approach to the "PC solution" (see [2, 3]). We are aware, however, of the strong reasons in favor of the "interface solution". The basic unit can be manufactured at substantially lower costs and many customers feel more inclined to use as terminal a telephone set which looks like a telephone. This suggests the idea of using the "interface solution" mainly in the private sphere while preferring the "PC solution" at the office.

4. COGNITO - an Experimental System

Seen from a conceptual viewpoint, the following main objectives have been pursued in the design and development of the COGNITO system:



a) "Interface solution"



b) "PC solution"

Figure 1: Two concepts for telecommunication terminals tailored to the needs of handicapped individuals

- development of an expandable modular system concept
- implementation of adequate user interfaces
- organization of the dialog

As far as the hardware implementation is concerned, it has been deemed essential to use a maximum of components available on the market. Seen from a functional viewpoint, the first development stage of COGNITO represents a telephone with advanced functions which are controllable by the human voice. Other options like, for instance, a speech output system may be added in the future.

4.1 Hardware

For the updated version it is sufficient to use a hardware base consisting of an IBM - compatible PC (XT) equipped with a RAM of at least 640kB, harddisk, a free serial port and enough room for a long plug-in board. The following components are added:

- speaker-independent speech-recognition board for a medium-size vocabulary consisting of approximately 100 words ¹
- telephone interface ²
- directional microphone and audio box (containing loudspeaker and amplifiers)

4.2 Software

The operation of the speech recognition system is based on special programs like driver software and the normally supplied training program which has been modified for the purpose of teaching the speech samples. For controlling the telephone functions, use is made of a subroutine library which belongs to the telephone interface and is activated by a program written in Turbo-Pascal. This program fulfills all other central functions and, in particular, the control of the dialog. Its design is based on a tree-shaped menu structure. The information required is represented by a window system. The dialog process includes possibilities of correction if digits and names are entered. Depending on the current state of the dialog, there is also the possibility of certain inputs only. This is particularly important whenever telephone

¹at present: VoiceScribe 1000 (Dragon systems)

²at present: Intelliphone (IBD)

calls are made because, during such periods, the speech recognition system is still in operation. This may activate unwanted functions arising from accidental incorrect recognitions.

4.3 Facilities

The voice-controlled telephones available on the market are no real help for people subject to motor handicaps since, as a general rule, manual inputs are required, for instance, at the beginning and end of a telephone call and in order to program the keys used for abbreviated dialing. It is for this reason that we have taken care for all functions to be controllable by voice. This means, in particular, that - in contrast to conventional solutions - the following central functions can be fulfilled without manual activation:

- initialization of dialing
- acceptance or termination of a telephone call

Moreover, it is possible - by means of voice commands - to compile and edit a file serving as an abbreviated dialing directory. Names are composed with the aid of a given spelling alphabet. The file may assume any size within the storage capacity of the PC. The present program version offers the possibility of switching among 10 users with a file being provided for each one. Other facilities are the possibilities of making emergency calls and repeated call attempts.

4.4 Practical Experience Gained so Far

Experience has shown that, at least, users familiar with the operation of a PC are indeed capable of managing the voice input and the dialog control, too, if the ambient noise is relatively low. It will be necessary to examine, in a future not too distant, whether the problem of disturbing noise can be handled by more sophisticated speech recognition boards or appropriate noise reduction procedures [4]. Moreover, it has become apparent that incorrect recognition may, once in a while, give rise to call interruptions. But there are still some untapped possibilities of increasing the resistance to this type of malfunction by modifications of the software and/or the dialog structure. Another problem is attributable to differences in the quality of telephone links which result in the need for a correct adjustment of the loudspeaker amplifier. Given the overall concept described above, a manual control is unacceptable although a voice control of the loudspeaker amplifier after the call setup is not

convenient, either. This problem is, therefore, likely to admit only the solution of an automatic amplifier control. Moreover, it will be necessary, in the future, to achieve a particularly effective decoupling of the loudspeaker and microphone signals in order to avoid interfering echos and whistles caused by feedback and to prevent, on the other hand, the speech recognition from being influenced by the loudspeaker signal. This suggests the idea of using echo canceling procedures [5], possibly in combination with conventional approaches.

When COGNITO was used by people lacking experience with PC systems, it became evident that, with respect to the organization of the dialog, the ideal of a user interface to be largely used by intuition has not yet been achieved. In this field, it will certainly be necessary to do much more basic research and development in order to implement an appropriate voice interface with a reasonable integration of a speech output system.

5. Future Objectives

In addition to the necessary activities outlined in the foregoing section, we are concentrating our efforts on the following objectives:

- use of further telecommunication services like, for instance, VTX
- integration of an environment-control system
- integration of an ISDN interface and the possibilities resulting therefrom

In addition, it is possible to extend the system to obtain a voice-controlled workstation equipped with alternative input and output devices (such as a suck-blow switch or Braille line).

6. Conclusion

COGNITO represents an experimental system allowing concepts for voice-controlled telecommunication systems to be investigated in a flexible way. This system can be operated without difficulty by persons subject to motor impairments. First experience in the application of COGNITO holds out the promise that more advanced and sophisticated systems of this design will be capable of using entirely new possibilities in the case of certain impairments.

7. References

- [1] Report COST 219: Future Telecommunication and Teleinformatics Facilities for Disabled People. To be published by the EC.
- [2] H. Klaus, K. Fellbaum: Einsatz der Spracherkennung und Sprachausgabe bei Personal Computern für motorisch behinderte Benutzer. Proceedings of the DAGA'91, Bochum.
- [3] M. Bazzani, E.M. Mumolo: PC-Based Telephone Communication System for Deaf-Blind People. Proceedings of IEEE Conf. Globecom'88.
- [4] R. Zelinski: Noise Reduction for Hands-Free Telephony Based on a Microphone Array with Adaptive Post-Filtering. International Workshop on Acoustic Echo Control, 1989, Berlin.
- [5] T. Becker, E. Hänslér, U. Schultheiß: Probleme bei der Kompensation akustischer Echos, Frequenz 1984(6), 142-148.