Intelligent Assistants for Handicapped People's Independence: Case Study

*N.G.Bourbakis and Despina Kavraki**

Binghamton University, Dept. EE & CIS Center, Binghamton, NY 13902
University of Crete, Dept. ECE, Chania 73100, Crete, Greece
*** AIS, 3916 Fuller Hollow Rd, Vestal NY 13850**

Abstract

This paper presents the first stage of the development of two intelligent assistants for handicapped people's independence. The first intelligent assistant, called Tyflos, will help a blind user to be independent and able to walk and work alone in a 3-D dynamic environment. The system captures images from the surrounding 3-D environment, either by the user's command or in a continued mode, and converts the visual description of each image into a verbal (low or high level) descriptions. With other words the system plays the role of human assistant, who describes to the user the 3-D visual environment. The second assistant, called Koufos, will help a deaf person to be more independent in the living environment by making him/her to visually hear sounds and spoken natural language and assisting him/her to visually learn to speak as well.

KEYWORDS: AI Systems, 3-D Visual Navigation, Speech & Image and Sound Understanding, Natural Language Processing, Knowledge Representation and Conversion, DB Information Retrieval, Learning.

1. Introduction

Recently the Modern Computer Engineering, the Artificial Intelligence (AI) and Intelligent Systems technologies have advanced to the stage of successfully supporting large scale complex engineering and scientific projects, such as MYSIN an expert system for medical diagnosis, QBIC a multimedia based system [9], EUROTRA an automatic natural languages translation system [10], STARS for automatic software development at megaprogramming level [11], GALAXY a spoken natural language system [12], INFOMEDIA a system for accessing news electronically by voice [13], MEL an autonomous intelligent vehicle [14], intelligent walking robots [15,16], 3-D space maps generation [17,18], autonomous vision systems [19], neuromorphic vision chips [20], the robot with feelings [21], an autonomous helicopter using fuzzy decision making [22], etc.

These technological achievements are now ready to successfully serve the humanity in very sensitive fields, such as visually impaired and deaf people. Recent challenges came to the scientific surface with the development of prototype artificial eyes [7], available after 2010. Also, implants for blind people [23,24] will be available at that time.

These technological breakthroughs need some time before to be available to the public. In the meantime, the advances of the state of the art in intelligent technologies could help these evolutionary efforts above by developing

0-8186-7728-7/96 $5.00 © 1996 IEEE

intelligent assistants to help handicapped people. In particular, intelligent systems to make blind people to "verbally see" via a machine vision system equipped with speech and natural language understanding facilities, and deaf people to "visually hear" and learn to "speak" via the same technologies. Moreover these technological facilities are now available and more than 70% functional, and there is no need of waiting 15 or more years to use them.

Based on these challenges above, this paper presents the first stage of the development of two intelligent assistants for handicapped people's independence. The first intelligent assistant, called Tyflos, will help a blind user to be independent and able to walk and work alone in a 3-D dynamic environment. The system captures images from the surrounding 3-D environment, either by the user's command or in a continued mode, and converts the visual description of each image into a verbal (low or high level) descriptions. With other words the system plays the role of human assistant, who describes to the user the 3-D visual environment. The second assistant, called Koufos, will help a deaf person to be more independent in the living environment by making him/her to visually hear sounds and spoken natural language and assisting him/her to visually learn to speak as well.

This paper is organized into 5 main sections. Section 2 describes the Tyflos system. section 3 presents the Koufos system. Section 4 briefly discusses the systems status, and section 5 concludes the overall presentation.

2. The Tyflos System [1]

The global configuration of the Tyflos system shown in figure 1. It consists of seven main parts, such as the helmet, the vision camera, the laser range scanner, the microphone, the portable computer, the ear speaker, and the communication system.

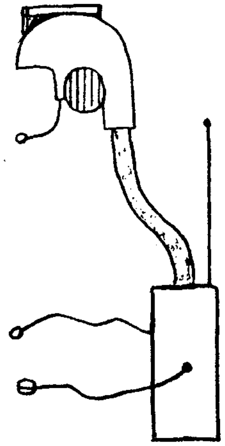

Figure 1.

2.1. The Helmet

The helmet plays a very significant role the user of Tyflos system. Except that it protects the user's head, most of the system's parts are mounted on it, see figure 2.

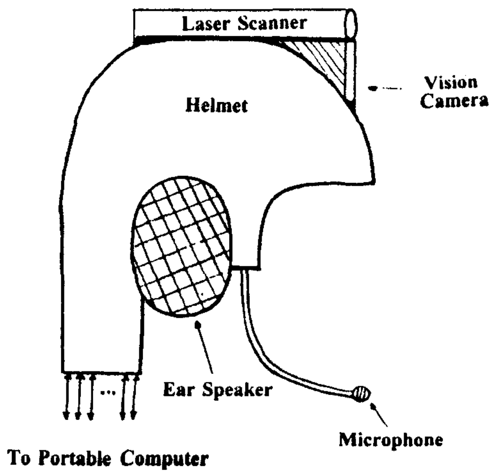

Figure 2.

2.2. The Vision Camera

The vision camera is mounted on the top of the helmet in order to provide a better view of the targeted scene. The vision camera is associated with a system of adjustable lenses that provide to the system the ability to capture images from long and short distance according to the user wishes. The camera is high resolution, color, small in size and weight.

2.3. The Laser Range Scanner

The laser range scanner is attached to the vision camera, with the ability to "scan" the same view with the camera. It is used for the determination of the distances of the existing objects and for the 3-D view of a captured image. The size of the laser scanner has also to be small and with resolution at least 1cm radius in a distance of 15 meters [28].

2.4. The Ear Speaker

This part is very important for the user. In particular, the Tyflos system has only one ear speaker through it the user receives the detailed description of the surrounding environment, and gets explanation about the sounds that the user's free ear receives. The ear speaker has to be a high fidelity component.

2.5. The Microphone

The microphone is the device through it the user communicates with computer and the appropriate Databases. It has a filter to reduce or if it possible to eliminate the noise from the user's environment.

2.6. The Portable Computer

The portable powerful computer is the most valuable component of the Tyflos system. It hosts all the important software tools that make the Tyflos system functional. In particular, it hosts all the DBs (Image, Speech, Natural Language DataBases), Software Interfaces, Knowledge Conversion Tools, Communication Tools, etc.

2.7. The Communication System

This component is important when the user needs to establish communication (email, software transfer) with other resources, such as portable or mobile information DBs, emergency centers and etc.

2.8. The Operation of the Tyflos System

The Tyflos system has a role to assist a blind person by describing and explaining with natural language words the scene and the sounds which surrounds the user. It has to able to assist the user in several cases such as,

- where is a particular object;
- describe the surrounding environment;
- in what distance a certain object is;
- what is the object which produces a certain sound;
- read me a book or a text;
- etc.

With other words it will play the role of the human assistant with the difference that it does "hold" the blind person, but it explain and describes the 3-D visual environment by making the user more independent. In particular, when the user wishes to find an "object" in the 3-D environment that he/she is working, he/she speaks though the microphone asking for that particular "object". The voice is converted into a natural language text (command), which accesses that portable DB. The system then requires from the user to slowly rotate his/her head scanning the surrounding space. Each image from a sequence of images captured by the vision camera is analyzed and recognized by the portable computer. If the "object" asked by the user is detected and recognized, then the system informs the user about it though the ear speaker. With a similar way the system explains and describes to the user via the ear speaker what object made a certain sound and where that object is from the current position of the user.

2.8.1. Visual to Audio Conversion

The conversion of information plays one of the most critical roles in the Tyflos system operation. In particular, visual information extracted from the surrounding environment is converted into speech. There are two different cases here:

. Recognition of a single object

In this case the particular object is converted into a hierarchical SPN graph [4,25]. The SPN graph is compared to the system's KB in order the object to be recognized or appropriately classified. Thus, the recognizable object activates the appropriate speech patterns, from the Speech DB, to be played by the ear speakers

. Description of a scene

In the second case a part of the surrounding scene is described. In particular, the description of each object and its relationships with the other objects in the same scene are extracted and described by SPN graphs. These SPN graphs are transformed into NL sentences. Then the sentences are appropriately converted into speech patterns, which are played by the ear speakers.

2.9. Communication and Mobile DBs Issues

Today's technological limitations don't allow a portable computer to host many different databases, such as sports, travel, market, medical, etc. Thus, for the Tyflos system there are two possible alternatives, either the user to carry CDs with different DBs, or its computer to be connected with a multimedia network in order to download the appropriate DB at each particular time. The use of different databases is a necessity for the Tyflos system, since the user needs to access different forms of information according to the circumstances.

3. The Koufos System [2]

The global configuration of the Koufos system shown in figure 3. It consists of seven main parts, such as the light helmet, the TV screen, the microphone, the portable computer, and the sounds sensor.

3.1. The Light Helmet

The light helmet is a basic part for the user of the Koufos system. On this helmet two

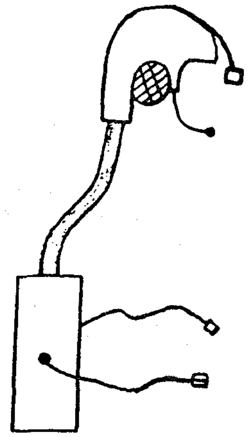

Figure 3.

important components are mounted, the microphone and the TV screen, see figure 4.

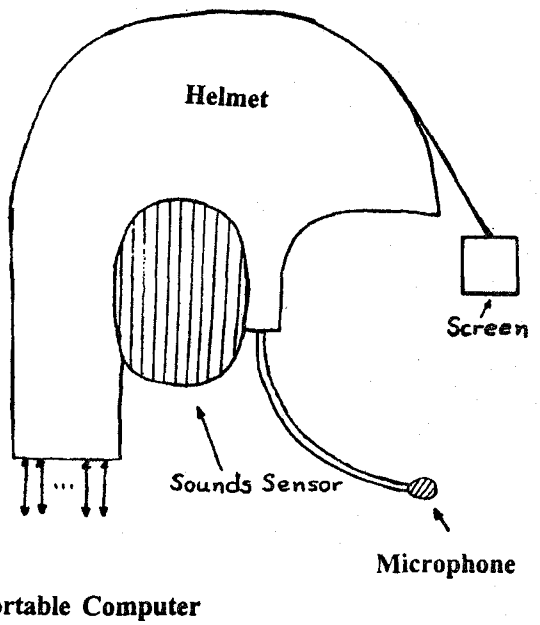

Figure 4.

3.2. The TV Screen

The TV screen is the state of the art in imaging & display. It is small in size (2.5x2.5 inch.) and weight, and attached to the light helmet. The TV screen is going to visually display in written natural language all the important sounds and voices from the surrounding environment. Thus, the user is able to continuously see the changes on the screen and understand - to some degree - what it is going on around.

3.3. The Sounds Sensor-Receiver

The sounds sensor-receiver is a very important component of the Koufos system as well. It receives sounds from the surrounding environment and displays them on the TV screen after filtering. Thus, the sounds sensor-receiver will play the role of a visual ear.

3.4. The Microphone

The microphone is the device through it the user communicates with computer and the appropriate speech and natural language DBs. It has a filter to reduce or if it possible to eliminate the noise from the user's environment. The microphone is used to help the user to learn to speak.

In particular, when spoken patterns are displayed on the TV screen, the user will try to produce sounds via the microphone and these sounds will be displayed bellow the correct speech patterns.

3.5. The Portable Computer

The portable powerful computer is the most valuable component of the Tyflos system. It hosts all the important software tools that make the Tyflos system functional. In particular, it hosts all the DBs (Image, Speech, Natural Language DataBases), Software Interfaces, Knowledge Conversion Tools, Communication Tools, etc.

3.6. The Operation of the Koufos System

The Koufos system also has a significant role to assist a deaf person by visually describing and explaining with spoken natural language words the scene and the sounds which surrounds the user. With other words it will replace the role of the human assistant, who describes with signs audio information to the deaf person, by making the user more independent.

3.6.1. Audio to Visual Conversion

When audio information is on the air, the sounds sensor receives that piece of information and the software systems convert it into visual pieces of information understandable by the user.

A. Speech and Sounds to NL Conversion

The first part is the conversion of the spoken natural language into natural language text displayed on the screen. The second part is the conversion of sounds from the surrounding environment into visual patterns recognizable the user.

. Visual Representation of Spoken NL (speech)

The visual representation of the spoken natural language is displayed on the user's screen in forms of NL text, see figure 5. This task is very difficult, since noise from the environments makes the speech patterns non-understandable. In addition, different speech patterns spoken by different people make the entire problem more difficult.

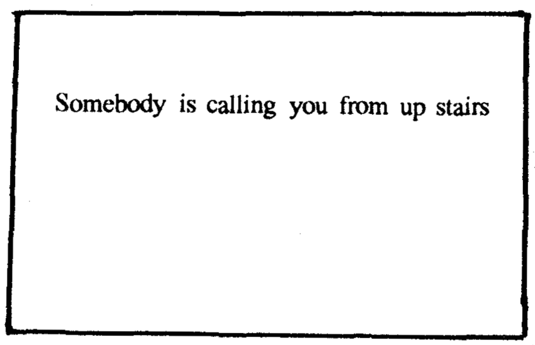

Figure 5: An illustrative example

. Visual Representation of Sounds

The visual representation of the sounds (non NL spoken sounds) is displayed on the user's screen in form of visual icons, see figure 6, thus the user will make a quick distinction between spoken and non-spoken sounds.

One very interesting feature of the sounds sensor is that it can define the direction (orientation) of the source, which created certain sounds. The direction is also displayed on the user's screen to make him/her the option to "see" the scene to that direction.

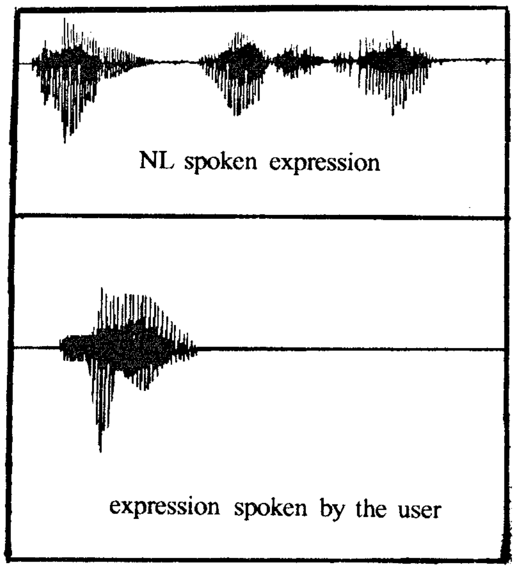

Figure 6: An illustrative example

3.6.2. Learning to Speak

The Koufos system provides the ability to the user to teach him/her to speak. In particular, a speech patterns, which represent certain NL expressions, are displayed on the upper part of the user screen. The user creates sounds whose the patterns are displayed on the lower part of the user's screen, see figure 7. These two different patterns are compared and the differences are displayed to the user. This feature gives the ability to the user to create "sounds" whose patterns will gradually "match" the "correct" speech patterns available in the speech DB.

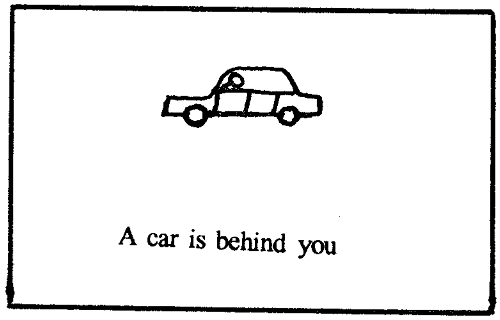

Figure 7: An illustrative example

4. Implementation of the Tyflos and Koufos Prototypes

4.1 . Integration of AI Methodologies[29,30]

The actual implementation of both Tyflos and Koufos systems combines the state-of-the-art in AI methodologies, such as image understanding, speech understanding, natural language understanding, knowledge representation and conversion, etc., with an efficient interfacing of software tools (DBs, OS and Communication interfaces). Figure 8 shows the architectural configuration of the systems computing part.

4.2. Prototypes

The hardware prototypes of the Tyflos and Koufos systems have been designed and their implementation is in progress at UC-AIS. At this stage the Tyflos system consists of the helmet, a vision camera (Hitachi HV C10 color, high resolution) , a portable computer NG-486, an ear speaker and a microphone. The Koufos system, at this stage, consists of a portable computer Mc-540, a speech synthesizer and a microphone.

The software tools available in the Tyflos system are: an image processing library with 150 algorithms, a pattern recognition tool for simple objects & OCR, an image understanding tool (75% functional), a speech to natural language conversion tool (50% functional), a

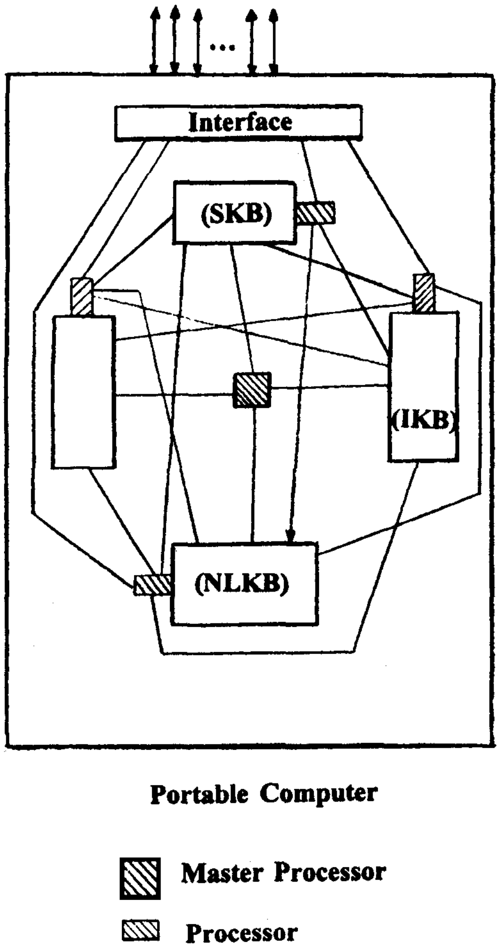

Figure 8:

learning tool, an expert system, and a small scale KB.

The software tools available for the Koufos system are: the Speaking Dynamically software, the visual representation of sounds package (50% functional), and a NL understanding tool.

4.3. Synchronization of Multiple Modalities

A very important issue for both systems is the synchronization and differentiation of different modalities (images, voices, sounds, etc.) occurred in the same "working" environment.

. **Differentiation** [4]

This is a very difficult task in particularly in the speech and sounds domains. Thus, the distinction of single modalities at the speech and sounds domains can be done by using Visual patterns, so that the actual problem is transferred from the S&S domain into the image domain.

. **Synchronization** [4,25]

The synchronization issue can be handled efficiently by using the stochastic Petri-net (SPN) model. This means that when an image or images are associated with certain voices and sounds, then these modalities are connected via an SPN model, where their activation and acquisition can be easily performed.

5. Conclusions, Benefits and Impact

The benefits of the systems presented here are significant for the handicapped people. Each of them will provide the ability to the user to understand better the surrounding environment, and give them more independence in their motion. They will not be as good as the ones proposed in IEEE Spectrum [7,23,24], since the user will not gain back their lost senses (visual, or audio), but on the other hand these are some possible options for an independent life. Of course, liability issues will be raised, but nothing is perfect.

The status of the systems are not functional yet. Several important problems (vision, speech) have to be solved, before such systems to be available to the users. However, a significant progress has been done, and new methods in both vision and speech developed by other researchers will be appropriately used.

Acknowledgements

This work is supported by grants AIS 1995-96, ECC 1996-97 and patents have been filed.

References

[1] N.G.Bourbakis and D.Kavraki, Tyflos - an intelligent assistant for visually impaired people's independence,AIS-TR-1993,10pages

[2] D.Kavraki and N.G.Bourbakis, Koufos - an intelligent assistant for deaf people's independence, AIS-TR-1993, 12 pages.

[3] N.Bourbakis, 2-D space maps generation in unknown environments,IJPRAI,vol.11,1997

[4] N.Bourbakis and D.Kavraki, Visual representation and recognition of speech patterns, TR-1995, sub. to Int. Journal

[5] N.Bourbakis and R.Andel, Fusing laser and image data for 3-D representations, IEEE Symp. on ATP, 1997

[6] D.Kavraki and N.Bourbakis, Improving sounds by matching them with spoken NL patterns, TR-1993, 11 pages, sub. to Int. Symposium on NL.

[7] G.Dagnelie and R.Massof, Toward an artificial eye,IEEE Spectrum,May 1996, pp.20-29

[8] ELMO microcameras, Image processing past and present, 1995

[9] M.Flickner et.al., Query by images and video content: the IBM's QBIC system, IEEE Computer, 28,9,23-32,1995

[10] The EUROTRA project: Machine languages translation, ESPRIT, 1986-96

[11] STARS , Newsletter, 12, March 1995

[12] V.Zue,GALAXY: accessing the information Highway using spoken language, MIT,1995

[13] R.Reddy, INORMEDIA news on demand, CMU research project, 1995

[14] Integration testbeds for mobility: the unmanned autonomous vehicle, NIST, Intelligent Systems Div., 1996

[15] H.Mira and Shimoyama, Dynamic walk of a biped IJRR, 3,2,60-74,1984

[16] M.Raibert, Legged robots that balance, MIT Press Cambridge MA 1986

[17] N.Bourbakis Generation of 2-D space maps from unknown environments, IJPRAI, 1997

[18] N.Bourbakis, A survey on 3-D navigation, TR-1996

[19] N.Bourbakis and S.Mertoguno, Kydon - an autonomous, multi-layers vision system, IJEAAI,vol.9,1,43-52,1996

[20] C.Koch & B.Mathur, Neuromorphic Vision Chips,IEEE Spectrum,May 1996,pp.38-46

[21] The Nanorobots Research Project in Japan, 1990

[22] The Autonomous, unmanned Helicopter, Japan 1995

[23] J.Wyatt and J.Rizzo, Ocular implants for the blind,IEEE Spectrum,May 1996,pp.47-53

[24] R.Normann, E.Maynard, K.Guillory and D.Warren, Cortical implants for the blind, IEEE Spectrum, May 1996, pp. 54-59

[25] N.Bourbakis, SPN representation on spoken natural language, IJISNL, 1997

[26] J.Albus, 3-D Images via stereo audio signals, communication, NIST,July 25,1996

[27] J.Albus, Outline for the theory of intelligence, IEEE T-SMC, 21,3,1991

[28] Riegel Laser Measurement Systems,Orlando FL, 1994

[29] L.Tsoukalas and R.Uhring, Fuzzy and Neural Approaches in Engineering, J.Wiley & sons, NY 1997 in press.

[30] N.Bourbakis, AI Methods and Applications, WSP, Singapore, 1992