OULUN YLIOPISTO
UNIVERSITY of OULU

Interactive voice response system and eye-tracking interface in assistive technology for disabled

University of Oulu
Department of Information
Processing Science
Bachelor’s Thesis
Kalle Kenttälä
6.5.2019

Abstract

The development of ICT has been very fast in the last few decades and it is important that everyone can benefit from this progress. It is essential for designing user interfaces to keep up on this progress and ensure the usability and accessibility of new innovations. The purpose of this academic literature review has been to study the basics of multimodal interaction, emphasizing on context with multimodal assistive technology for disabled people. From various modalities, interactive voice response and eye-tracking were chosen for analysis. The motivation for this work is to study how technology can be harnessed for assisting disabled people in daily life.

KEYWORDS

Multimodality, eye-tracking, disabled, interactive voice response, eye gaze

SUPERVISOR

Aryan Firouzian

Contents

1. Introduction

Using full potential of computers has always been restricted by the interaction between human and computer. This restriction is reduced by focusing on usability in the design of user interfaces. The focus of multimodal interfaces is to make the interaction more natural to human behavior and allowing access to computers for people that are unable to use conventional tools like keyboard and mouse. This literature review focuses on interactive voice response systems and eye-tracking technology in multimodal interfaces that can be used to improve the usability when used by disabled people. Google Scholar, IEEE Xplore, and ACM Digital Library were the main databases used in this research.

In Chapter 2, the interactive voice response system will be presented along with some problems that have been identified. The interactive voice response system includes Text-to-Speech system, that decreases the need for visual feedback. In Chapter 3, eye-tracking technology and some of its basic principles and applications are presented. Eyes can be powerful and intuitive tool when interacting with interfaces. Chapter 4 focuses on special aspects that must be considered when designing user interfaces for disabled people. In many cases, people with restricted or absent movement capabilities need customized and individually adapted user interfaces. Chapter 5 presents the basic concept of multimodality and some examples of multimodal interfaces. Multimodality is used on challenging the traditional human-computer-interaction with mouse and keyboard. Finally, conclusions are presented in Chapter 6.

2. Interactive voice response (IVR) system

2.1 Basics of IVR

As presented by Kim,Liu & Kim (2011), there are two kinds of IVR systems that are commonly used: telephone data entry (TDE) and automated speech recognition (ASR). TDE is widely used with voice menus when people run errands via phones. With IVR systems, mobile phone users can access information by simply dialing a telephone number to create connectivity and communication with a server or internet (Inam;Azeta;& Daramola, 2017). Users can access voice information using a touch-tone interface or ASR. These systems may be difficult to use and have few recognized limitations. These problems will be discussed later in this paper (Kim;Liu;& Kim, 2011).

Basic technologies, as presented by Inam et al. (2017), that are used for implementing IVR systems that are capable of interacting with people are:

- Voice Extensible Markup Language (VoiceXML): allows interacting with the internet with a speech recognition system via speech browser or telephone or speech and touch-tone applications.
- Text-to-Speech (TTS): used for converting text input into voice or audio output or file.
- Automatic Speech Recognition (ASR): converts speech from a pre-recorded audio to text or voice.

There are some features that must be considered when designing IVR systems (Inam;Azeta;& Daramola, 2017):

- Multilingual/Single (MS): Supporting multiple language offers great advantage for users in IVR systems. Most common language in single language systems is English Language.
- Dialog Menu (DM): Dialog menus are good way to create interaction between system and user. The dialog menu can offer options about the call and work as a router of the call.
- Local Language (LL): Offering local language with IVR systems enables more effective interaction with the system.
- Feedback Mechanism (FM): Feedback is effective way to limit errors in interacting with the systems by requesting confirmation for the user’s input to system.

VoiceXML allows to create vocal application using the “classic” model for the Internet application development. This means that content and the business logic are separated from the presentation layer. It uses a W3C standarded language and is based on XML, which makes it easy to use for developers familiar with Web technologies (Corno;Pireddu;& Farinetti, 2007).

A dynamic voice response system platform (Karademir & Heves, 2013), consists of six modules *(Fig. 1)*. Voice XML Generator creates parametric routines used by Engine Unit for all incoming, outgoing or internal IVR calls. IVR Unit assigns a port to all incoming calls and creates a section ID for a call. This ID will be sent to Voice XML Generator with other call data. Engine unit receives the call data from Voice XML Generator and loads all necessary scenarios from database and sends the information received from IVR Unit to the Record Unit. Record unit saves all scenarios from Engine Unit in case of lost communication between Record Unit and Database. Central Unit consists of a Graphical User Interface (GUI), that can be used for designing and testing parameters, scenarios and rules for the system.

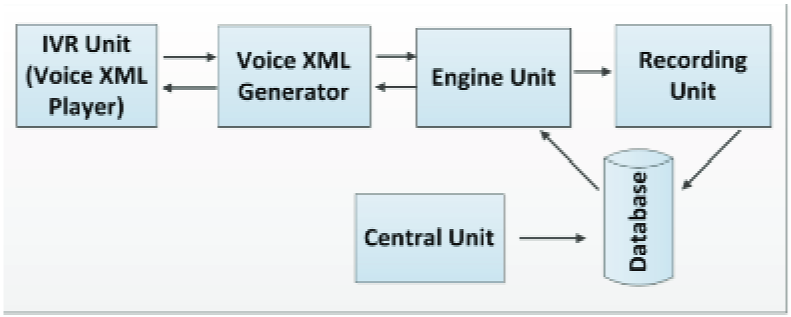

Fig. 1 Dynamic IVR Platform (Karademir & Heves, 2013)

Dynamic Voice Response System uses pre-designed and pre-programmed building blocks, called actions, to create customized flows. These actions are written using Voice XML. For example, Voice XML Generator actions are used for audio playback or transferring call, but Engine Unit Actions are mostly for data handling. Some of these actions and one example of work flows are shown in Fig. 2 (Karademir & Heves, 2013).

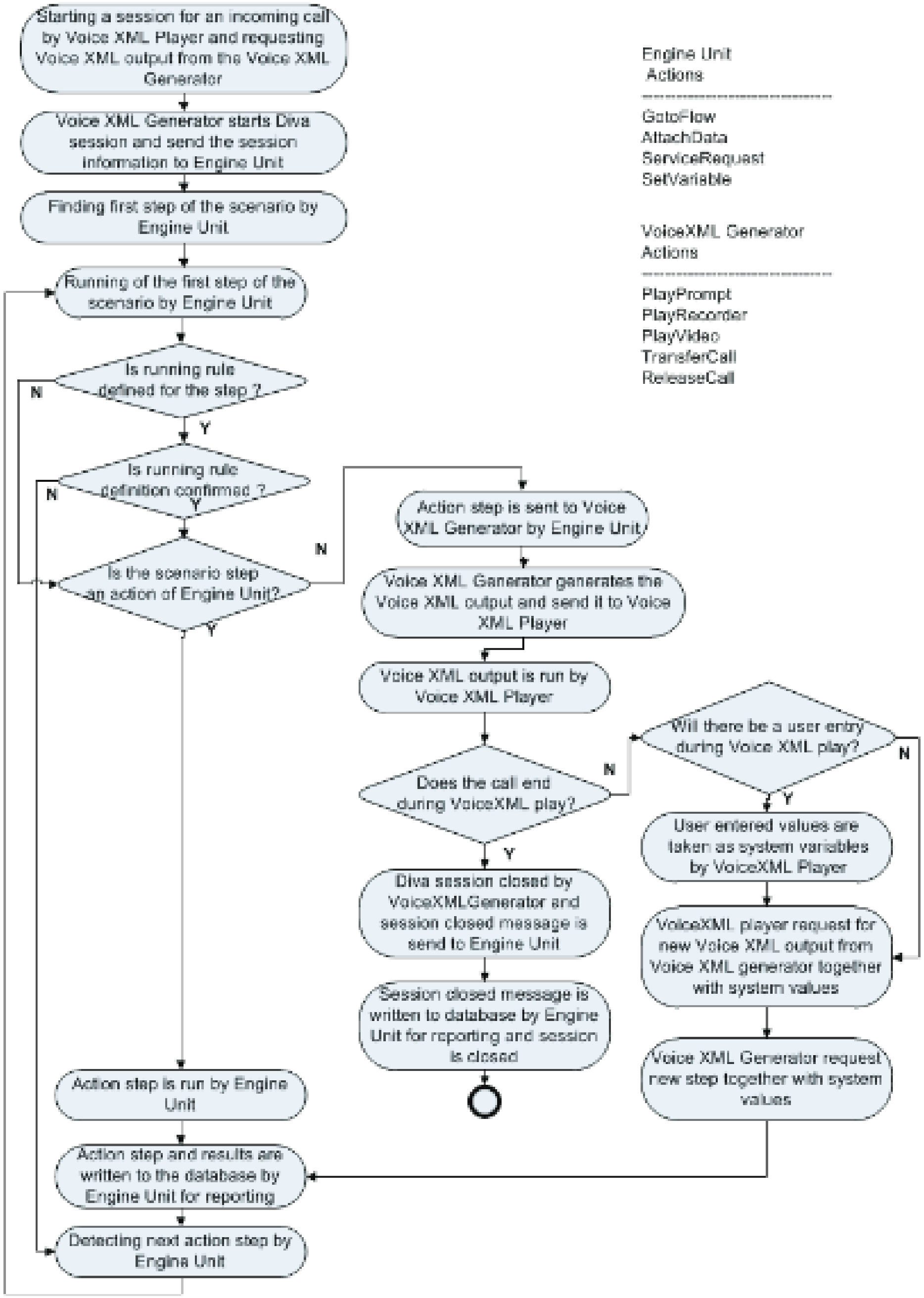
Starting a session for an incoming call by Voice XML Player and requesting Voice XML output from the Voice XML Generator
Voice XML Generator starts Diva session and send the session information to Engine Unit
Finding first step of the scenario by Engine Unit
Running of the first step of the scenario by Engine Unit
Is running rule defined for the step ?
N
Y
Is running rule definition confirmed ?
N
Y
Is the scenario step an action of Engine Unit?
N
Y
Action step is sent to Voice XML Generator by Engine Unit
Voice XML Generator generates the Voice XML output and send it to Voice XML Player
Voice XML output is run by Voice XML Player
Does the call end during VoiceXML play?
N
Y
Will there be a user entry during Voice XML play?
N
Y
User entered values are taken as system variables by VoiceXML Player
VoiceXML player request for new Voice XML output from Voice XML generator together with system values
Voice XML Generator request new step together with system values
Diva session closed by VoiceXMLGenerator and session closed message is send to Engine Unit
Session closed message is written to database by Engine Unit for reporting and session is closed
Action step is run by Engine Unit
Action step and results are written to the database by Engine Unit for reporting
Detecting next action step by Engine Unit
Engine Unit Actions
GotoFlow
AttachData
ServiceRequest
SetVariable
VoiceXML Generator Actions
PlayPrompt
PlayRecorder
PlayVideo
TransferCall
ReleaseCall

Fig. 2 Flow diagram of IVR Platform (Karademir & Heves, 2013)

2.2 Problems in IVR

Kim et al. (2011) have identified four problems inherent to IVR systems:

- Transience: Human listening is controlled by short-term memory. This means that the user usually remembers only few seconds when listening voice menus and prompts. Because of this transience of memory, it is difficult to remember menus and prompts that are too long. This should be considered when designing IVR systems.
- Linearity: A linear interaction type means that user can only navigate information in one given order. Speech and audio interfaces must be sequential, while visual interfaces can be simultaneous. Using linear interface is much more time consuming than graphical user interface (GUI) that can display all messages simultaneously.
- Ambiguity: Sometimes speech can be difficult to recognize or understand. This is a problem in IVR because user can't interrupt the speech and tell it to repeat. In most cases user can resolve the meaning of a word by the context of the sentence, but it gets more difficult if only one word is spoken.
- Minimal Feedback: The lack of visual feedback in IVR systems can make user feel less in control, because user doesn't necessarily know the current state of the system. This is especially problematic if the system has tasks that can take long time.

3. Eye-tracking technology

3.1 Basics of eye-tracking technology

Eye tracking can be divided into two categories, intrusive and non-intrusive. Intrusive eye tracking methods include electromagnetics embed to eye with contact lenses or electrode straight into the eye. Both methods can be potentially harmful to body and especially eye. Non-intrusive eye tracking means taking advantage of Video Oculographic (VOG). In this method cameras are used to get sequences of eye images and obtain and analyze data from the images. Non-intrusive method based on VOG can also be divided to two parts: IR camera and monocular camera. Algorithms based on IR cameras involve Pupil Center Cornea Reflection (PCCR). This is used to get the coordinates of pupil by differentiating light and dark image of the eye almost simultaneously when glint is only difference between these two images. This way the glint can be easily tracked and identified in the images. In monocular camera based VOG algorithms include Hough Transform that uses gray-scale or gradient-scale images (Zhao;Wang;& Yan, 2016). Hough Transform will be presented later in this paper.

As presented by Zahir, Hossen, Al Mamun, Amin & Ishfaq (2017), to improve the usefulness of the eye-tracking with camera both indoors and outdoors, infrared filter should be used for reducing or eliminating other electromagnetic radiation interference. The effect of using IR filter can be seen in fig 3.

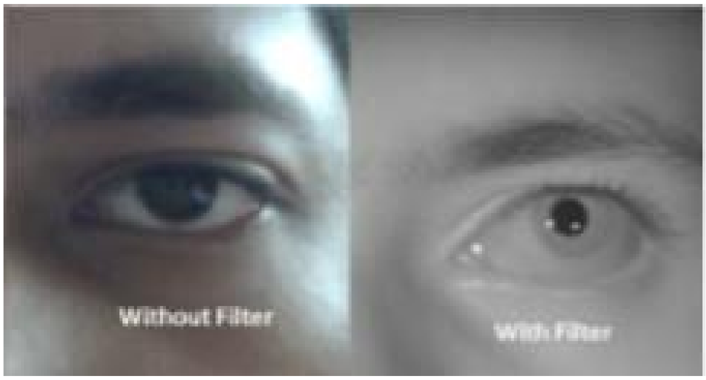

Fig. 3 Eye image of without [center] and with [right] IR filter by Zahir et al. (2017)

Eye-tracking technology in this context means the combination of digital image processing and cameras focusing on the movement of the eye. One or more cameras can be mounted on glasses, or remotely in front of the user. The key features of the eye are captured from the image and the relative positioning of these features are calculated to extrapolate the eye gaze. This kind of image processing requires lot of computing power. (Hiley;Redekopp;& Fazel-Rezai, 2006).

The general structure of eye-tracking device consists of camera, USB Interface, data for position of the eye, and database of pupil position. There are multiple methods and algorithms for how the pupil is detected from captured image. (Sambrekar & Ramdasi, 2015)

3.1.1 Pupil detection

Commonly used technique in pupil detection is Hough transform (HT). It is a segmentation algorithm that locates different shapes in images. In this case, it can locate the circular area of the pupil in the eye (Sambrekar & Ramdasi, 2015). One way to isolate the pupil from background is to use the difference image between bright-eye, and dark-eye images (Hiley;Redekopp;& Fazel-Rezai, 2006).

The Hough transform algorithm, presented by Paul Hough in 1962, detects features for particular shapes in digital images. When detecting pupil in the eye, HT aims to detect the best-fitting circle for the eye pupil contour. When pupil radius is unknown, HT detects the pupil edge and its center coordinates in the digital image. If the pupil radius is known, a pair of coordinates are used as search parameters. The biggest challenge for HT is that the pupil is not always a perfect circle. There are many factors that affect the shape like pupil position, illumination, corneal reflection, and blinking. Also, the accuracy is relative to the running time of the algorithm and the performance of the processing unit (Bozomitu;Păsărică;Cehan;Rotariu;& Barabaşa, 2015).

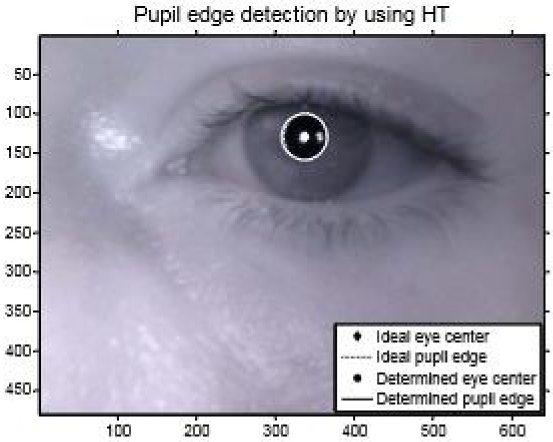

Fig. 1 Pupil in center

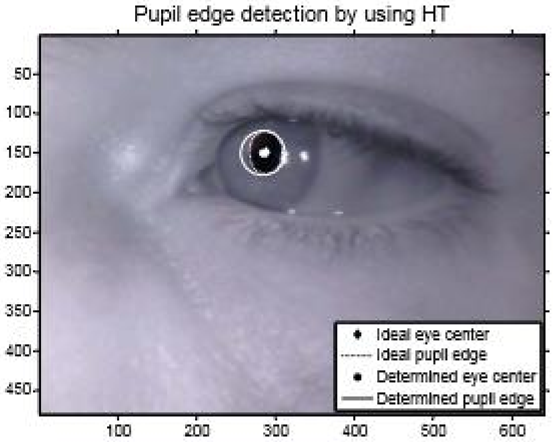

Fig. 2 Pupil on the edge of sclera

In ideal case, the pupil edge and the determined pupil edge created with HT algorithm are overlapping (Fig. 1). This means that the relative error of the pupil center positioning is minimal. This ideal case occurs when the pupil is in the center of the sclera. When pupil is not in the center of the sclera, the shape of the pupil changes from circle to ellipse, which minimizes the radius. In this case the relative error of the pupil center positioning increases (Bozomitu;Păsărică;Cehan;Rotariu;& Barabaşa, 2015).

3.1.2 Glint detection

Glint is a small and intense dot in the pupil. It is produced by a reflection of light from the corneal surface (Sambrekar & Ramdasi, 2015). Although glint is easier to process from the image than pupil because of the high contrast to the background, it is vulnerable to false positives and fragmenting caused by external factors (Hiley;Redekopp;& Fazel-Rezai, 2006).

One glint detection method proposed by Zhao et al. (2016) is to use Harris corner detection. This method is one of the most common for detecting features in images. It uses gradient images that can be pre-filtered for more accurate results. This method uses monocular camera without IR assistant. The glint of the eye is extracted by cross-correlation between pre-adaptive threshold algorithm and pre-filtered Harris corner detection algorithm. Although the coordinates of the glint can be defined accurately, there are conditions where the contrast between pupil and glint are nearly ignorable, resulting to false extraction of the glint (Zhao;Wang;& Yan, 2016).

3.1.3 Eye positioning and gaze

After detecting the pupil, the positioning can be calculated by computing the deviation of current position and average position. Average position means the position when the eye is looking forward. The position can be rendered as 2-axle coordinates (Wanluk;Visitsattapongse;Juhong;& Pintavirooj, 2016).

Eye gaze means either the gaze direction, the distance between pupil and fixed points such as glint inside pupil (Zhao;Wang;& Yan, 2016), or the point of gaze of an eye related to the head for determination of person's line of sight or point of fixation (Sharma & Abrol, 2015).

Eye-gaze-based algorithm produce the estimation of the corresponding Region of Interest (ROI). This means that the eye gaze aims to define the point where the user is looking. For this estimation, pupil detection and glint detection are required. The estimated relative position of these two detected features are mapped as reference points for glint vector and the center of the pupil (Sharma & Abrol, 2015).

3.1.4 Using ideograms as a communication method

One way to communicate with computers is using ideograms. This means that there are pictures or icons on computer screen that user can select. There are different methods for ideogram selection. First method is to select ideogram by maintaining users gaze over the ideogram for a certain time interval. In this method the gaze is being used similar to mouse cursor. This method raises the problem of "Midas touch" effect. This means that ideograms can be randomly selected because of users wandering gaze. This method can also be challenging if the user does not understand the meanings of all ideograms and needs more time to understand the action related to that ideogram. There also has to be a resting zone on the computer screen where the cursor can be placed without selecting any ideograms. Other method for selecting ideograms is using the blinking of an eye similar to clicking a mouse. The challenge with this method is to identify voluntarily blinking by counting the number of consecutive frames of blinking or using dynamic threshold for selection (Păsărică;Bozomitu;Cehan;& Rotariu, 2016).

4. Special aspects in interface design for disabled

4.1 Characteristics of Locked-in syndrome

Locked-in syndrome has been diagnosed as a condition in which a patient is completely aware of his/her surroundings but has lost ability to control all muscles in the body except for the eyes. Eye is controlled via six extraocular muscles and three cranial nerves. This makes the eye as one of the accurate and precision organs to be controlled in the body. The patient has full cognitive functions but is unable to respond to most stimuli. Due to this paralysis, eyes are one of the few mechanisms that allows the patient to interact with the environment. By offering means to interact increases both physical and mental welfare of the patient. (Boustany;Itani;Youssef;Chami;& Abu-Faraj, 2016).

4.2 Improving the quality of life for disabled

The focus of most interfaces designed for disabled is to improve the quality of life by increasing the level on interaction that the user can have with his/her environment. Nowadays when computers and are a big part of everyone's life, it is necessary that these devices are not restricted to only those who can easily handle the common peripherals like keyboard and mouse. Physical limitations often make it challenging to use traditional interfaces and systems like keyboard, mouse, or joystick. This means that new kinds of input and output methods must be created. (Wanluk;Visitsattapongse;Juhong;& Pintavirooj, 2016), (Sambrekar & Ramdasi, 2015).

Popular means for assistive technology in controlling computers are switches that are used for scanning through a matrix of letters, symbols, words or phrases. Each matrix entry can be used with a sequence of switch operations. Because of the matrix scan delays, there is a barrier in communicating at an effective rate with this technology. (Betke;Gips;& Fleming, 2002).

Electric wheelchair is one of the devices that improve the quality of life for many disabled people. For people that are physically restricted to control the wheelchair manually, researchers have developed several technologies such as speech recognition, head array, electroencephalography (EEG) based brain-computer-interfaces (BCI), electro-oculography (EOG) systems, and sip-and-puff (SnP) switches. All these technologies have limitations that prevent the use in daily life (Eid;Giakoumidis;& El Saddik, 2016).

4.3 Challenges in Designing Assistive Technology

Dubey,Mewara,Gulabani & Trivedi (2014) categorize assistive technologies in seven types. These types contain assistive devices with different applications (Table 1.)

Category (Domain)	Examples of applications
Medical Science	Braille embossers, screen readers, alternative keyboards, mice to touch screens, light signaler alerts, listening devices to closed captioning
Education/Organizational Skills	Reading tools, learning software programs, electronic pointing devices
Communication	Speech/voice recognition programs, Text-to-speech or speech synthesizers, keyboard filters
Daily living activities	Joysticks, trackballs, touch screens
Recreation, leisure, and adaptive play	Screen enlargers and magnifiers
Mobility	Wheelchair, ambulance aids, walkers, canes
Computer applications	Telecommunication device for the deaf (TTD) and teletypewriter (TTY) conversion modems, on-screen keyboards

Table 1 Categorizing of Assistive Technologies by Dubey et al. (2014)

Widely used design philosophy, universal design, emphasizes that the design should be used by all irrespective of their application domain and context of use. On the other hand, to maximize the usage, the design should be dependent of the context of the use. This means that some approach between universal and context dependent design should be used to maximize usability for large number of users and still support customization for individual needs. In the field of assistive technologies, one of the most important challenges is to consider the individual needs of the user. (Dubey;Mewara;Gulabani;& Trivedi, 2014)

5. Multimodality in HCI

5.1 Theory of Multimodality

Multimodality means that human can interact with computer or environment in many ways. Multimodality with environment is natural to human, but with computers it is not that common. Nowadays computers use only one or two interface devices, like keyboard and mouse. HCI systems today are not natural to use and challenging. The development of interface devices is based on mouse, joystick or keyboard. Some research has confirmed that users prefer different devices on different tasks, for example gestures for direct object manipulation and natural language for descriptive tasks. Wireless

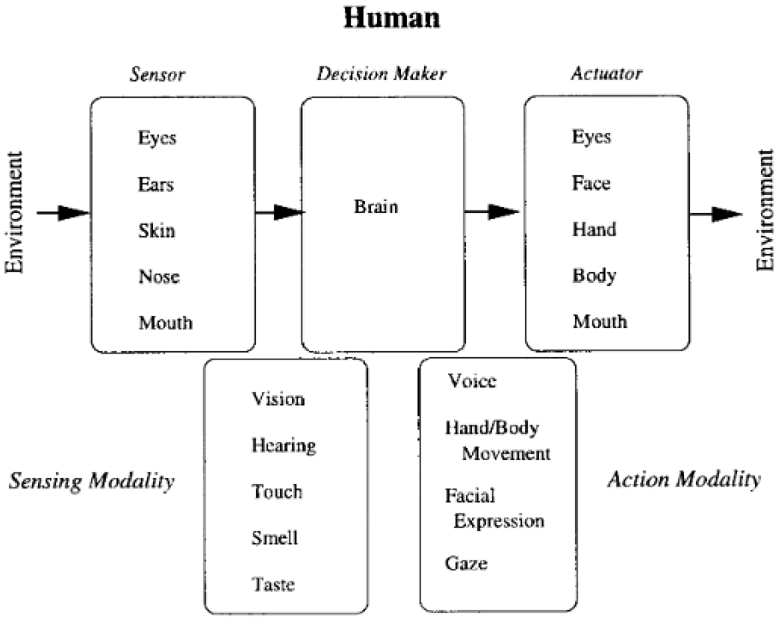

Fig.3 Modalities for human sensing and action (Sharma;Pavlovic;& Huang, 1998)

technology and rapid development in computer processors will aid the development of multimodal interfaces by making them more comfortable, accurate, and fast. Ideal case in HCI is that computer can interpret all natural human actions. This means that computer needs to be able to perceive all actions of humans. These actions include speech, gestures, eye gaze, and all body movements (Sharma;Pavlovic;& Huang, 1998).

Modalities in HCI can be divided into sensing modalities and action modalities (Fig. 3). Sensing modalities are the means for human to receive output from computer. For example, vision is used for reading from monitors. Action modalities are the means for creating input for computer, like hand movement for using mouse and keyboard. Hand movement is the most exploited modality in HCI because of dexterity of human hand, which allows accurate selection and positioning of mechanical devices with the help of visual feedback. All action modalities are potential candidates for multimodal interfaces (Sharma;Pavlovic;& Huang, 1998).

Multimodality in HCI should be improved, because human interaction is naturally multimodal. Human tends to look at the object he/she is talking about, for example. Also, when listening to someone talking, human usually perceives more than just the speakers voice. The limitations of interacting with computers restricts the effectiveness of harnessing the potential that the computer can offer. Different modalities complement each other. For example, gestures are ideal for object manipulation and natural language is better suited for descriptive tasks (Sharma;Pavlovic;& Huang, 1998).

There are more advantages in multimodality than just more enhanced efficiency. The flexibility of alternating between different input modes helps to avoid any physical overexertion. Multimodality can also enhance usability in noisy or mobile environments, where traditional systems are difficult or impossible to use (Oviatt, 1999).

5.2 Multimodal integration in HCI

As presented by Sharma et al. (1998), three levels of integration can be distinguished by using the theoretical and computational apparatus developed in the field of sensory data fusion: data fusion, feature fusion, and decision fusion. Data fusion is the lowest level of fusion. This fusion involves integration of raw observations and can appear only when the observations are of the same type. Because of that, this fusion rarely occurs in multimodal integration of HCI. One example, when it can occur is when two different cameras are used to gather visual information of one object. Data fusion is characterized by the highest level of information of the three fusion types. It also requires a high level of synchronization of the observations. Feature fusion assumes features that have been analyzed from each data streams. This type of fusion is usable, when the modalities are closely coupled or synchronized, for example speech and lip movement. This fusion contains less detailed information than data fusion but is less sensitive to noise. Nonetheless, feature sets can be large, which can result in high computational cost for this approach. Decision fusion uses individual mode decisions or interpretations of each modalities. This means that the synchronized semantic level decisions of different modalities are fused for interpreting. Highest level of information from a modality is estimated as a feature, then this feature will be interpreted as a decision and fused with another interpreted decision. This fusion type is the most common in HCI and it is the most robust and resistant to individual sensor failure. It requires a low data bandwidth and is usually less computationally expensive than feature fusion. (Sharma;Pavlovic;& Huang, 1998).

5.3 Mechanisms for Decision-Level Fusion

Because decision fusion is the most commonly used type of fusion and it is the most relevant in the context of this paper, two mechanisms are considered, as presented by Sharma et al. (1998): frames and software agents.

Frames as a concept is commonly used in artificial intelligence literature. A frame is a unit of a knowledge source describing an object. Each frame has a number of slots associated with it. The slots include possible properties of the object, actions, or relationship between frames. Relationships are used in designing networks of frames for a particular context with linking contextual semantics. These are called semantic networks, which are uses when associating different modalities with individual frame slots. For example, speech can designate object's color, while gesture can designate object's color. (Sharma;Pavlovic;& Huang, 1998).

Software agent is a software entity that operates continuously and autonomously in a certain environment, sometimes resided by other agents and processes. Agents are supposed to operate long periods of time without human intervention. They also should learn from their experience and be able to communicate with other agents. They can be task oriented, flexible, adaptive, and can delegate modal interactions to different communicating subagents when integrating modalities. Open agent architecture (OAA) is highly suitable for multimodal fusion tasks. In OAA. each agent can handle single modalities, while the modal agents communicate with a central agent, known as a facilitator. Facilitator handles the interactions with other agents that need multimodal information. This architecture allows implementation for sensor discordances detection, evidence accruement, and contextual feedback. However, this architecture is more complex than some other integration techniques. This complexity in OAA can be facilitated with distributed computing, in which different agents can exist on different computers. (Sharma;Pavlovic;& Huang, 1998)

5.4 Examples of multimodal interfaces

Hatfield, Jenkins, Jennings & Calhoun (1996) propose a concept called The Eye/Voice Mission Planning Interface (EVMPI). This concept consists of system with integration of voice recognition and eye-tracking technology for pilots and operators using advanced military systems. This integration permits hands-free operation of cockpit displays in planes by using gaze on user interface items of interests and giving verbal commands for the system. The aim of this concept is to reduce the pilot's cognitive and manual workload.

Papaj, Pleva, Cizmar, Dobos, Juhar & Ondas (2007) present a project called MobilTel that provides a communicator that uses multimodal interface. Included modalitites are speech recognition, graphical interface, pen/touch screen interaction and keyboard. The communicator has two multimodal services: Railway Scheduler and Weather. These services were tested with different combinations of modalities.

Neßelrath, Moniri & Feld (2016) present a prototype in-car system for using some car features by speech, gaze and micro-gestures. This prototype uses SiAM-dp (Situation Adaptive Multimodal Dialogue Platform) for handling the interaction between car and driver. This system can handle car actuators like the windows, outside mirrors and the turning lights.

6. Conclusions

IVR and eye-tracking technology is not very much used combination in designing multimodal interfaces, according to the fact that not much research literature was found. One reason for this could be that the usefulness of this combination can be limited to only a small group of users. The advantages of using IVR are easy modification of voice menus and standardized programming language. IVR can also be used more privately when used with earphones compared to using the system via visual screen. The problems of IVR are limited length of voice menus because of human short-term memory, the required linearity of speech and audio interfaces, challenges in identifying speech, and lack of visual feedback. Eye-tracking technology is proven to be useful but is vulnerable to environmental influence like lighting and movement. One promising communication method with eye-tracking is using ideograms on computer screens. One challenge, especially with eye-tracking technology is the requirement for moderate computation capacity, which restrains the mobility and ubiquity at the moment. One strength of this combination is that when using IVR instead of visual interface, the eye-tracking is not interfered with the need to read from screen.

As said before, this combination of eye-tracking and IVR is not very common. The user group that would benefit from this combination the most, is people that communicate with other people with communication boards, like the people with Locked-In Syndrome. These communication boards are placed in front of the user, and he/she looks at different letters to form words. Using eye-tracking and IVR can remove the need for help from another person. This way the user is less dependent on personal assistance. In addition to communicating with other people, eye-tracking and IVR can ease the use of other applications like phone apps.

References

Betke, M., Gips, J., & Fleming, P. (2002). The camera mouse: Visual tracking of body features to provide computer access for people with severe disabilities. *IEEE Transactions on neural systems and Rehabilitation Engineering* (pp. 1-10). IEEE.

Boustany, G., Itani, A. E., Youssef, R., Chami, O., & Abu-Faraj, Z. O. (2016). Design and development of a rehabilitative eye-tracking based home automation system. *3rd Middle East Conference on Biomedical Engineering.* IEEE.

Bozomitu, R. G., Păsărică, A., Cehan, V., Rotariu, C., & Barabaşa, C. (2015). Pupil Centre Coordinates Detection Using the Circular Hough Transform Technique. *Electronics Technology (ISSE).* IEEE.

Corno, F., Pireddu, A., & Farinetti, L. (2007). Multimodal Interaction: an integrated speech and gaze approach.

Dubey, A. K., Mewara, H. S., Gulabani, K., & Trivedi, P. (2014). Challenges in Design & Deployment of Assistive Technology. *2014 International Conference on Signal Propagation and Computer Technology* (pp. 466-469). IEEE.

Eid, M. A., Giakoumidis, N., & El Saddik, A. (2016). A novel eye-gaze-controlled wheelchair system for navigating unknown environments: case study with a person with ALS. *IEEE Access*, 558-573.

Hatfield, F., Jenkins, E. A., Jennings, M. W., & Calhoun, G. (1996). Principles and guidelines for the design of eye/voice interaction dialogs. *Proceedings Third Annual Symposium on Human Interaction with Complex Systems HICS'96* (pp. 10-19). IEEE.

Hiley, J., Redekopp, A. H., & Fazel-Rezai, R. (2006). A low cost human computer interface based on eye tracking. *28th IEEE EMBS Annual international Conference* (pp. 3226-3229). IEEE.

Inam, I. A., Azeta, A. A., & Daramola, O. (2017). Comparative Analysis and Review of Interactive Voice Response Systems. *Conference on Information Communications Technology and Society.*

Karademir, R., & Heves, E. (2013). Dynamic Interactive Voice Response (IVR) Platform. *EuroCon 2013* (pp. 98-104). Zagreb: IEEE.

Kim, H. C., Liu, D., & Kim, H. W. (2011). Inherent usability problems in interactive voice response systems. *International Conference on Human-Computer Interaction* (pp. 476-483). Springer Berlin Heidelberg.

Neßelrath, R., Moniri, M. M., & Feld, M. (2016). Combining speech, gaze, and micro-gestures for the multimodal control of in-car functions. *12th International Conference on Intelligent Environments (IE)* (pp. 190-193). IEEE.

Oviatt, S. (1999). Ten myths of multimodal interaction. *Communications of the ACM, 42*(11), 74-74.

Papaj, J., Pleva, M., Cizmar, A., Dobos, L. U., Juhar, J., & Ondas, S. (2007). MOBILTEL-Mobile multimodal telecommunications systems and services. *2007 17th International Conference Radioelektronika* (pp. 1-4). IEEE.

Păsărică, A., Bozomitu, R. G., Cehan, V., & Rotariu, C. (2016, October). Eye blinking detection to perform selection for an eye tracking system used in assistive technology. *Design and Technology in Electronic Packaging (SIITME), 2016 IEEE 22nd International Symposium* (pp. 213-216). IEEE.

Sambrekar, U., & Ramdasi, D. (2015). Human computer interaction for disabled using eye motion tracking. *Information Processing (ICIP), 2015 International Conference on* (pp. 745-750). IEEE.

Sharma, A., & Abrol, P. (2015). Comparative Analysis of Edge Detection Operators for Better Glint Detection. *2015 2nd international Conference on Computing for Sustainable Global Development.* IEEE.

Sharma, R., Pavlovic, V. I., & Huang, T. S. (1998). Toward multimodal human-computer interface. *Proceedings of the IEEE, 86*(5).

Wanluk, N., Visitsattapongse, S., Juhong, A., & Pintavirooj, C. (2016). Smart wheelchair based on eye tracking. *The 2016 Biomedical Engineering International Conference.*

Zahir, E., Hossen, M. A., Al Mamun, M. A., Amin, Y. M., & Ishfaq, S. M. (2017). Implementation and performance comparison for two versions of eye tracking based robotic arm movement. *International Conference on Electrical, Computer and Communication Engineering* (pp. 203-208). IEEE.

Zhao, F., Wang, H., & Yan, S. (2016). Eye Glint Detection and Location Algorithm in Eye Tracking. *IEEE/CIC International Conference on In Communications in China (ICCC).*