Perceptual User Interface

GINNE.M. JAMES[1], B.S.ASHWIN PRABHAKAR[2], S.VAISHNAVI[3]
Assistant Professor, Dept. of Computer Technology, Sri Krishna Arts and Science College, Coimbatore, Tamil Nadu,India[1]
Student, Dept. of Computer Technology, Sri Krishna Arts and Science College, Coimbatore, Tamil Nadu, India[2]
Student , Dept. of Computer Technology, Sri Krishna Arts and Science College, Coimbatore, Tamil Nadu, India[3]

Abstract—The main use of perceptual inputs is an emerging area within HCI that suggests a developing Perceptual User Interface (PUI) that may prove advantageous for those involved in mobile serious games and immersive social network environments. Here we will see about the evolution of the interaction between man and machine and also introduce some new terms like perceptual user interface, multimodal interfaces, etc. Perceptual interface includes various modes of interaction which we will be using in our future like gesture technology, speech recognition, eye tracking and much more. This just covers the outline of the above mentioned topics.

Keywords—*perceptual user interface, multimodal system, cognitive science, hand gesture, virtual objects, recognition.*

Introduction

Think of typical situation where sit in front of our computers and enter the passwords and usernames, then we click icons by icons drag the mouse pointers and use the irregularly arranged keys to type a question and eventually we get the answer to get what we want or imagine you have just logged in to your new computer, and it is displaying some of its extravagant features. It then begins asking you a series of questions. You are in a hurry to get to your email, but it pops up with yet another start-up window to set some option that is not necessary to set up now. You breath, glare, growl something under your breath, and proceed to type with a little more speed.

The above scenario from the computer's point of view may be the affective or emotional response. But most of the time it'll be frustrating for the user, i.e., for us. This is where multimodal systems come in.

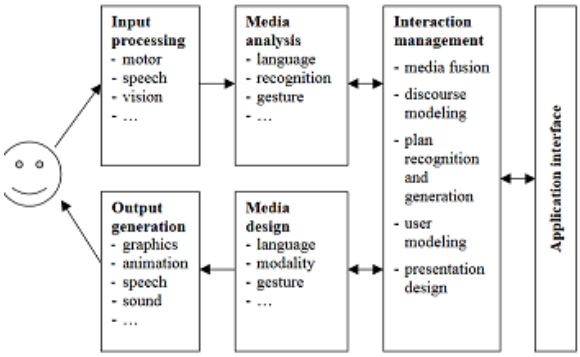
As a foundation for advancing new multimodal systems, proactive empirical work has generated predictive information on human-computer multimodal interaction, which is being used to guide the design of planned multimodal systems. Major progress has occurred in both the hardware and software for component technologies like speech recognition, pen, and vision. In Finally, real applications are being built that range from map-based and virtual reality , to field medic systems for mobile use in noisy environments, to Web-based transactions and standard text-editing applications.

The computer hardware and software development vector indicates movement away from a traditional windows, icons, menus and pointer (WIMP) and desktop paradigm based on 2D content and (traditional GUI) interaction. Input and manipulation of this variety is mature at present, with the mouse and keyboard as 1D or 2D non-perceptual interfaces that have been widely accepted for some time by the general public and developers. The extension of our interactive experience to mobile computing has brought touch screen and 2D gesture language that the majority of users are comfortable with and believe enhances their experience.

Architecture of Multimodal systems

Maybury and Wahlster describe a high-level architecture of intelligent user interfaces. As their idea of intelligent user interfaces includes multimodal interaction, this model can be used for modeling multimodal interfaces. Specifically, the highly important modals in a multimodal interface are user and discourse modals. There can be one or more user models in a system. If there are several user models, the system can also be denoted as an adaptable user interface. The discourse model handles the user interaction in an extreme level, and uses media analysis and media design processes to understand what the user wants and to present the information

with the appropriate output channels.



Multimodal systems are completely different from that of the present GUIs mainly because of the input they receive from the user. In GUIs, the inputs are certain and controlled whereas in Multimodal systems the inputs are most of the time uncertain and come with disturbances. Let's take speech recognition for example, the words we pronounce may not be the same every time and also the background noise is also taken with the input. This means that the system needs to be probabilistic every time. Secondly, whereas standard GUIs assume a sequence of discrete events, such as keyboard and mouse clicks, multi- modal systems must process two or more continuous input streams that frequently are delivered simultaneously. The challenge for system developers is to create robust new time-sensitive architectures that support human communication patterns and performance ,including processing users' parallel input and managing the uncertainty of recognition-based technologies. A general approach to reducing or managing uncertainty is to build a system with at least two sources of information that can be fused. For example, various efforts are under way to improve speech recognition in noisy environments by using visually-derived information about the speaker's lip movements, called "visemes", while interpreting "phonemes" or other features from the acoustic speech stream. Multimodal systems that interpret speech and rim motion integrate signals at the level of visage and phoneme features that are closely related temporally. Such architectures are based on machine learning of the visage-phoneme correlations, using multiple hidden Markov models or temporal neural networks. This feature-level architectural approach generally is considered appropriate for modes that have same time scales. A second architectural approach, which is appropriate for the integration of modes like speech and gesture, involves fusing the semantic meanings of input signals. The two input signals do not need to occur simultaneously, and they can be recognized independently. This semantic fusion architectural approach requires less training data, and entails a simpler software development process. As an illustration of the semantic fusion approach, we describe the Quick Set multimodal architecture and information processing flow.

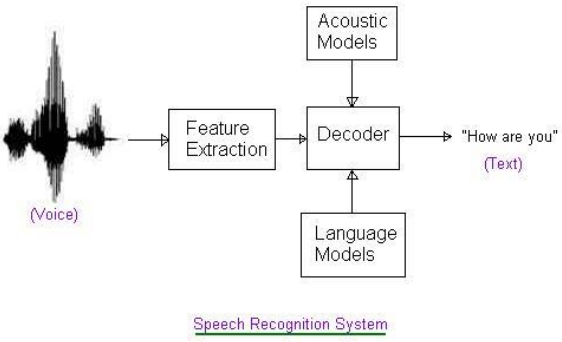


Fig 1. Basic structure of speech recognition system

Design of multimodal systems

Designing Multimodal Input and Output. The cognitive science literature on inter sensory perception and intermodal coordination has provided a foundation for determining multimodal design principles . To optimize human performance in multimodal systems, such principles can be used to direct the design of information presented to users, specifically regarding how to integrate multiple modalities or how to support multiple user inputs (for example, voice and no. Some of the general guiding principles that are essential to the design of effective multimodal interaction are:

Maximize human cognitive and physical abilities. Designers need to establish how to support intuitive, streamlined interactions based on users human information processing abilities (including attention, working memory, and decision making) for example:

• Avoid unnecessary presenting information in two different modalities in cases where the user must simultaneously attend to both sources to comprehend the material being presented; such redundancy can increase intellectual load at the cost of learning the material.

• Maximize the advantages of each modality to reduce user's memory load in certain tasks and situations, as illustrated by these modality combinations:

° System visual presentation coupled with user manual input for spatial information and parallel processing;

° System auditory presentation coupled with user speech input for state information, serial processing, attention alerting, or issuing commands.

Assimilate modalities in a manner compatible with user preferences, context, and system functionality. Supplementary modalities should be added to the system only if they improve satisfaction, efficiency, or other aspects of performance for a given user and context. When using multiple modalities:

• Match output to acceptable user input style (for example, if set grammar is constrained by the user, do not design a virtual agent to use unconstrained natural language);
• Use multimodal cues to improve collaborative speech (for example, a virtual agent's gaze direction or nod can guide user turn-taking);
• Ensure system output modalities are well synchronized for a short duration of time (for example, map-based display and spoken directions, or virtual display and non-speech audio);

There are two basic things which inspired in designing and organizing the multimodal systems. First, the cognitive science literature on inter sensory perception and intermodal coordination during production is beginning to provide a foundation of information for user modeling, as well as information on what systems must recognize and how multimodal architectures should be formulated. For example, the cognitive science literature has provided knowledge of the natural integration patterns that satisfy people's lip and facial movements with speech output. Given the complex nature of users' multimodal interaction, cognitive science has and will continue to play an essential role in guiding the design of robust multimodal systems. In this respect, a multidisciplinary perspective will be more central to successful multimodal system design than it has been for traditional GUI design. Secondly, high-fidelity automatic simulations also have played an important role in prototyping new types of multimodal systems. When a new multimodal system is in the planning stages, design sketches and low-fidelity mock-ups may initially be used to visualize the new system and plan the sequential flow of human-computer interaction. These tentative design plans then are rapidly transitioned into a higher-fidelity simulation of the multimodal system, which is available for proactive and situated data collection with the intended user population.

During high-fidelity simulation testing, a user interacts with what she believes is a fully-functional multimodal system. During the interaction, a programmer assistant at a remote location provides the simulated system responses. As the user interacts with the front end, the programmer tracks her multimodal input and provides system responses as quickly and accurately as possible. To support this role, the programmer makes use of automated simulation software that is designed to support interactive speed, realism with respect to the targeted system, and other important characteristics. For example, with these automated tools, the programmer may be able to make a single selection on a workstation field to rapidly send simulated system responses to the user during a data collection session. High-fidelity simulations have been the preferred method for prototyping multimodal systems for several reasons. Simulations are relatively easy and inexpensive to adapt, compared with building and iterating a complete system. They also permit researchers to alter a planned system's characteristics in major ways (e.g., input and output modes available), and to study the impact of different interface features in a systematic and scientific manner (e.g., type and base-rate of system errors). In comparison, a particular system with its fixed characteristics is a less flexible and suitable research tool, and the assessment of any single system basically amounts to an individual case study. Using simulation techniques, rapid adaptation and investigation of planned system features permits researchers to gain a broader and more principled perspective on the potential of newly emerging technologies. In a practical sense, simulation research can assist in the evaluation of critical performance tradeoffs and in making decisions about alternative system designs, which designers must do as they strive to create more usable multimodal systems. The most recent high-fidelity simulation tools have been designed to collect data and prototype new multimodal systems that support collaborative group interactions. They also are beginning to support real-time processing of the paralinguistic aspects of users' natural speech and pen input signals, such as changes in user amplitude that indicate intended addressee during multi-person exchanges, which is needed to develop new adaptive multimodal systems. An example of a dual-wizard high-fidelity simulation environment designed to prototype collaborative multimodal interfaces and also adapt to changes in users' speech and pen amplitude. This particular simulation collected speech, visual, and digital pen and paper data from students during 3-person collaborative meetings while they used a computational assistant to

solve mathematics problems. Two wizards were required in this simulation to process key user data in real time involving the linguistic content of users' requests, and amplitude of their speech or pen communication. Specialized simulation software and wizard training both were needed to support adequately fast and error-free teamwork between the two wizards. To support the further development and commercialization of multimodal systems, additional infrastructure that will be needed in the future includes: simulation tools for rapidly building and reconfiguring multimodal interfaces, automated tools for collecting and analyzing multimodal corpora, and automated tools for iterating new multimodal systems to improve their performance.

Hand Gesture

Nowadays, the majority of human-computer interaction (HCI) is based on mechanical devices such as keyboard, joystick or gamepad. In recent years there has been a growing interest in methods based on computational vision due to its ability to recognise human gestures in a easy way . These methods use the images acquired from a camera or from a setto pair of cameras as input. The main goal of these algorithms is to measure the hand contour at each time instant. To facilitate this process, uniquely coloured gloves or markers on hands or fingers are used. In real time also, using a controlled background makes it possible to locate the hand efficiently work . These two conditions impose restrictions on the user and on the interface modals. One of the most fascinating application of multimodal systems is gesture recognition systems.The impendent of virtual environments brings in a whole new set of problems for user interfaces. The computer vision devices can be implemented and upgraded to the new input devices in the future. It gives the input command to the computer rather than to make a function of taking photos or record videos. We can do more discharge to transform the computer vision devices to become an input command device to reach the function as keyboard or mouse. One of the ways to give signal to computer vision devices is by using hand nod. More specifically hand gesture is used as the input signal to the computer. Certain signal can be recognized by computer as an input of what computer should work. These will benefit the entire user without using a direct device and can do what they want as long as the computer vision device can sense it. The user feels easier to work with the mouse or keyboard.

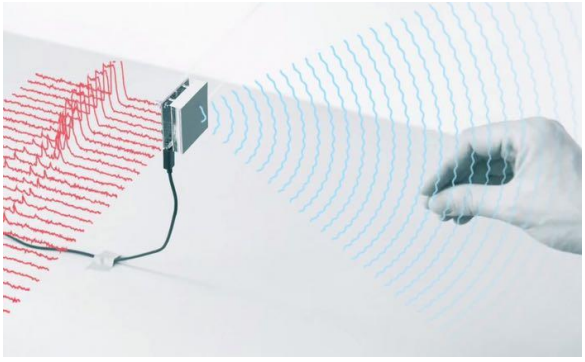


Fig 2. Gesture recognition device taking input from the user

Interaction between humans comes from different sensory modes like nod, speech, facial and body languages. The main advantage of using hand gestures is to interact with computer as a non-contact human computer input technique. The present research effort defines an environment where a number of challenges have been considered for obtaining the hand gesture recognition techniques in the implicit environment. Being an interesting part of the Human computer interaction hand nod recognition needs to be booming for real life applications, but complex structure of human hand presents a series of challenges for being tracked and understood. Other than the gesture complexities like variability and flexibility of structure of hand other challenges include the shape of nod, real time application issues, presence of background noise and variations in illumination conditions. The specifications also involve accurate form of detection and recognition for real life operations.

Fig 3. Apple watch controlled drone

The present research effort has a goal of developing an application using vision based hand gestures for handling of objects in virtual environment. Our application presents a more effective and user friendly methods of human computer interaction intelligently with the usage of hand nod. Functions of mouse like controlling the movement of virtual object have been

replaced by hand gestures. The complexity involved is with the apprehensive and concession phases of the simulated virtual application. The challenges encountered are noisy environment which creates a big violation on the detection and recognition performance of human hand gestures. The webcam is used for capturing hand as input to reduce cost. Manipulation of virtual objects has been done through modeling of some predefined command based hand nod. The future computer or laptop may eliminate the use of keyboard and mouse by substituting with a vision-based interpretation device.

Future Works

PUIs have been rapidly developing over the past few years. These multimodal systems are already into our smart phones as speech recognitions and face detectors and these are expected to reach our houses.

After these mindboggling technologies and cool gadgets, one might think of a more advanced interface i.e., catching the brain waves of the user and responding to the thought input of the user. This kind of interface is already under development where scientist try to catch and decode the brain waves of the user letter by letter in Japan.

References

[1]Sharon oviatt, Philip cohen "Multimodal interfaces that process what comes naturally" communications of the acm, march/2000 Vol. 3.

[2]Rosalind W. Picard "Affective perception "M.I.T. Media Laboratory E15-392, 20 Ames St., Cambridge MA 02139.http://www.media.mit.edu/~picard

[3]Sharon oviatt"Handbook of Human-Computer Interaction", (3rd ed., edited by J. Jacko), Lawrence Erlbaum: New Jersey, 2012.

[4] Turk, M., Ed. Proceedings of the 1998 Workshop on Perceptual User Interfaces; research.microsoft.com/PUIWorkshop; www.research.microsoft. com/PUI- Workshop

[5] A. Mazalek, S. Chandrasekharan, M. Nitsche, T. Welsh, P. Clifton,A. Quitmeyer, F. Peer, F. Kirschner, "Recognizing self in puppet controlled virtual avatars," In proceedings of the 3rd international conference on Fun and Games, 2010, pp. 66-73. http://dx.doi.org/10.1145/1823818.1823825

[6] R. Chua, D.J. Weeks, D. Goodman, "Perceptual-motor interaction: some implications for human-computer interaction," The human-computer interaction handbook, L. Erlbaum Associates Inc. Hillsdale, NJ, USA, 2003, pp. 23-34.

[7] R.J.K. Jacob, L.E. Sibert, "The perceptual structure of multidimensional input device selection," In proceedings of the ACM Human Factors in Computing Systems Conference(ACM CHI '92), 1992, pp. 211-218.

[8] E. Guy, P. Punpongsanon, D. Iwai, K. Sato, T. Boubekeur, "LazyNav: 3D ground navigation with non-critical body parts," In proceedings of the 2015 IEEE Symposium on 3D User Interfaces, 2015, pp. 43-50.

[9] R.A. Bolt, E. Harranz, "Two-handed gesture in multi-modal natural dialog," In proceedings of the ACM Symposium on User Interface Software and Technology (ACM UIST'92), 1992.