Proceedings of International Conference on Advances in Computer Technology and Management (ICACTM)
In Association with Novateur Publications IJRPET-ISSN No: 2454-7875
ISBN No. 978-81-921768-9- 5
February, 23rd and 24th, 2018

MULTIMODAL COMMUNICATION IN HUMAN COMPUTER INTERACTION

ROSHAN M. JAISWAL
MET Institute of Computer Science,Mumbai, India,
roshanj2007@gmail.com

ABSTRACT
With such a rapid advancement in powerful mobile devices and sensors in recent years, inclusion of multimodal communication is on the rise. In this research paper, importance of multimodal human computer interaction is being discussed. It talks about different gestures, body language, facial expression recognition and emotions in voice. Also the challenges and issues and its various applications in Multimodal Human Computer Interaction is also being discussed.
This research paper also includes comparison between unimodal and multimodal communication.

I. INTRODUCTION

Facial expressions are our interfaces in our emotional and social live. People are very much interested in automatic analysis of facial gestures in multimodal human computer interaction. Many researchers experiments with new hardware devices to explore new paradigms for human computer interaction. Optimization of performance of the interaction between human and computer is also very crucial, hence developing new models and new theories is important.
Multimodality is the ability of the system to communicate with a user to convey the meaning automatically.
Different senses and their corresponding modalities:
Table 1: Details regarding sensory organs

Sensory perception	Sense organ	Modality
Sense of sight	Eyes	Visual
Sense of hearing	Ears	Auditive
Sense of touch	Skin	Tactile
Sense of smell	Nose	Olfactory
Sense of taste	Tongue	Gustatory
Sense of balance	Organ of equilibrium	Vestibular

This research paper intends to provide the benefits and challenges in Multimodal Human computer interaction. Along with it applications of Multimodal system will also be seen. Multimodal model has its own architecture to understand.
Multimodal basically means multiple input devices or multi sensor interaction
There are two views on multimodal interaction. The first view focuses on the human view and control and the second view focuses on the computer view and control.

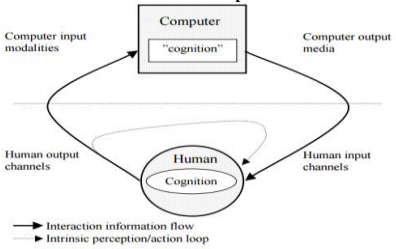

Fig 1: Relation between Human and Computer

The modalities can be divided into seven groups as follows:

1. Internal chemical (Blood oxygen, glucose, pH)
2. External chemical (taste, smell)
3. Somatic senses (touch, pressure, temperature, pain)
4. Muscle senses (stretch, tension, join position)
5. Sense of balance
6. Hearing
7. Vision

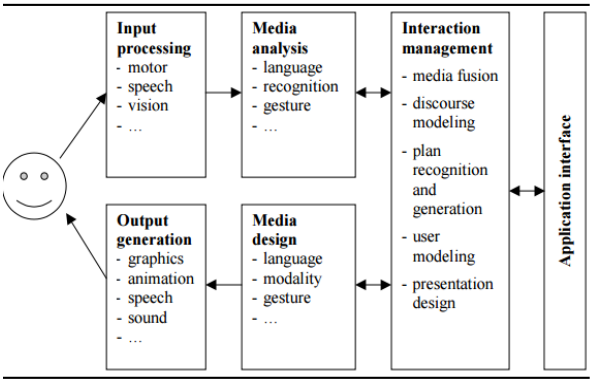

Fig 2: Input output interaction for human

II. Benefits of Multimodal system

1. Efficiency
2. Redundancy
3. Perceptibility
4. Naturalness
5. Accuracy
6. Synergy

III. Applications of Multimodal system

1. Mobile telecommunication

The increasing demand coming from large network operators in converging their different infrastructure for fixed and mobile communication networks, results in a couple of interesting questions that have to be resolved in near to midterm future. This consists in the use of different terminals and access modes for communication and access to information services. In fact, operators are interested in deploying services available in mobile networks like Short Message Service (SMS) to their fixed networks customers.
Similar services like SMS is possible in a computing environment for the management of large files and directories.
All these ways will bring a new way of communication for a customer and he doesn't need to remember a large set of numbers for cell phone or email etc.,
But the customer can contact the partner by simply using the name, letting the network decide dynamically which is the appropriate method for establishing the communication based on the context and presence information held in the network. With respect to

information services this requires the possibility to adapt the content to the capabilities of the terminal that is used in the current situation or even more the use of several devices interacting for a service, like the presentation of a video on a television set, while the mobile phone is used to as voice channel and as remote control.

The overwhelming number of information presented today to individuals will continue to increase, from which the largest part is irrelevant in the actual situation to the user. In the current scenario more than 50% of the emails contain spam, while browsing on google website, one needs to visit several links before one gets a relevant information.

This need currently makes important to personalize the information that flows in and out. Also a virtual assistance is necessary to guide the user properly.

Hence the requirement of a function that boost the interaction between the user and the software. It could be done using speech and graphical interface with the support of enhanced multimedia features wherever possible in a cost effective manner.

Along with the above needs, the need for powerful application servers and processing units which uses intelligent switching techniques should be pushed for enhancing the network communication.

2. Using in automobiles

Automakers are trying to differentiate their brands through HMI technology, so they're adopting multimodal interfaces.

In the nearing future, drivers may use technologies like eye gazing, headdirecting device for proper direction and image tracking system to make vehicle control more smooth and easy.

Automotive engineers are searching for the perfect intersection between device complexity and user input and output.

A simulator study was conducted to compare 16 younger (mean age 22 years) and 16 older (mean age 68 years) drivers' ratings of workload (time, visual, psychological stress) and performance of navigation and button-pushing (identification of vehicle or road hazards) tasks under both high- and low-load driving conditions when simple or complex advanced traveller information (ATI) was presented visually only, or auditory or both can be used for display. For all participants, both the auditory and multimodality displays produced better performance in terms of response times, total number of correct turns and subjective workload ratings than those of using the visual-only display. Participants using the multimodality display also made the fewest errors related to push-button and navigation tasks, and controlled their vehicles properly. The visual display led to less safe driving, apparently because it imposed higher demands on the drivers' attention. An age effect was found in the present study, with younger drivers performing better and reporting less stress than older drivers. Notably, however, use of the multimodality display significantly improved the older drivers' performance in the button-pushing task.

Traditionally, automobile dashboards and centre consoles have been the domain of buttons, dials and knobs. But, those mechanical devices are about to go the way of running boards, hood ornaments and rumble seats.

In the near future, the automobile drivers may interact with their vehicles and various infotainment systems using various hand gesture, touch and voice based commands.

Advanced human-machine interface (HMI) features are already available on luxury vehicles produced by Audi, BMW, Cadillac, Lexus, Lincoln and Mercedes-Benz. Suppliers such as Visteon and Continental unveiled technology that pushes the limit even beyond the current status.

Fig 3: A person operating an virtual system

3. Interactive information panel or kiosk

User interface using multimodal system allow people to interact with the kiosk machine via speech, gestures to the standard input and output devices. Architecture of the kiosk contains key modules of speech processing and computer vision. An array of four microphones is applied for far-field capturing and recording of user's speech commands, it allows the kiosk to detect voice activity, to localize sources of desired speech signals, and to eliminate environmental acoustical noises. A noise robust speaker-independent recognition system is applied to automatic interpretation and understanding of continuous Russian speech. The distant speech recognizer uses grammar of voice queries as well as garbage and silence models to improve recognition accuracy. Pair of portable video-cameras are applied for vision-based detection and tracking of user's head and body position inside of the working area.

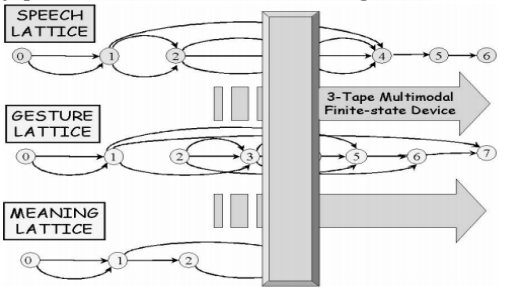

Fig 4: Multimodal device

IV. APPROACHES

1. Face localization
2. Facial feature detection
3. Facial expression recognition

4. Face tracking

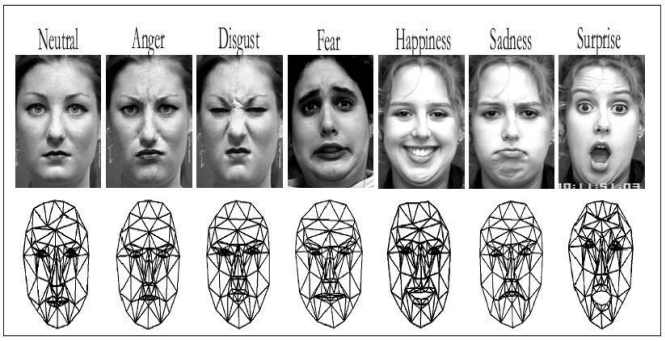

Fig 5: Different facial expressions

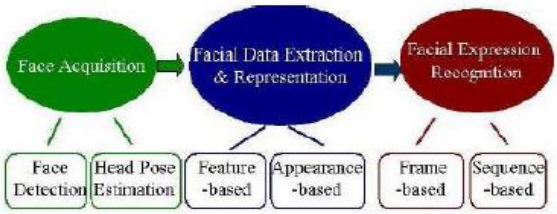

Fig 6: Relation between face acquisition and facial expression

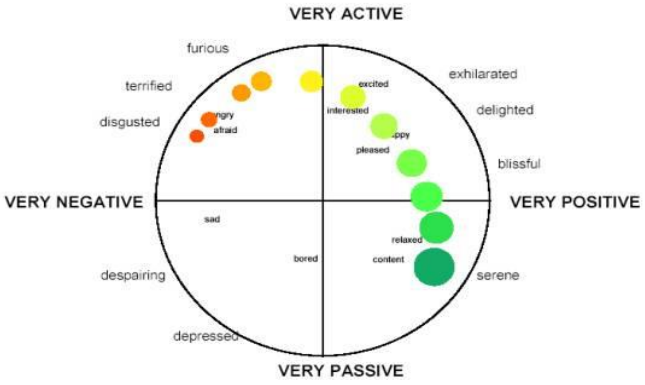

Fig 7: Different expressions

Problems with Face detection approaches

1. Pose: The images of a face may vary due to different relative position with camera
2. Presence or absence of structural components: Examples could be beards, moustaches, glasses etc.
3. Facial expression: Different facial expression directly affects the appearance of a face.
4. Obstruction: Face may be partially obstructed by other objects
5. Image orientation: Due to different camera orientation the face image may vary
6. Imaging conditions: Different camera characteristics like brightness, contrasts, saturation, lighting, background makes the face image appear different.

V. Multimodal vs Unimodal Physiological Control in Videogames

In the last few decades, videogames have evolved in a big way in terms of graphics
like 3D environments. With the inclusion of biofeedback interaction the multimodal system is more intense as compared to traditional unimodal system.

The games can be compared in terms of Ease of use, Fun, Originality and many more. Both unimodal and multimodal versions are known for their won strengths like : the unimodal version for its simplicity of use, and the multimodal for its various graphic effect.

With multimodal system, the player feels empowered by using all the abilities provided in the game. Also game can be made more difficult by demanding more physical efforts from the player.

VI. CHALLENGES IN MULTIMODAL COMMUNICATION

Significant progress on multimodal interaction systems are needed before it becomes a common name and inseparable part in computing. Each modalities in multimodal system (vision, speech, sound, touch etc.) requires continued development and on each one of them active research can be conducted. Different camera poses and various user appearances is a huge challenge.

CONCLUSION

There are several issues which still exist in smooth functioning of multimodal interaction. Each modalities needs individual attention to get multimodality efficient and popular. The mean of accuracy and error rates needs to controlled and should be kept under the threshold value.

REFERENCES

1. https://arxiv.org/abs/1406.0532
2. https://www.w3.org/2004/02/mmi-workshop/Sienel-alcatel.pdf
3. http://dx.doi.org/10.1080/00140130010011369
4. S. Oviatt, "Multimodal Interfaces," in *Chapter in Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications* (Lawrence Erlbaum Assoc. Mahwah, NJ, 2003), pp. 286–304.
5. 3E.R. Kandel and J.R. Schwartz, Principles of Neural Sciencies. Elsevier Science Publisher, 1981
6. http://ijcsiet.com

Roshan M. Jaiswalreceived MCA degree in Computer Application from Thakur College,Mumbai University, India in 2009. His current research area is Distributed and Cloud computing. Currently he is working as an Assistant Professor in Department of Masters of Computer Application(MCA), MET, Institute of Computer Science , Mumbai, India.

He is having 10 years of teaching and industry experience. He is a member of International Association of Engineers (IAENG) and presented 5 papers in International journal and conferences.