

Human Factors and Design Issues in Multimodal (Speech/Gesture) Interface

C. J. Lim^{*Corresponding author}, Younghwan Pan, Jane Lee

Dept. of Game and Multimedia Eng., Korea Polytechnic University, 2121 Jungwang-dong,
Ahihung-si, Gyeonggi-do, 429-793 Korea

scjlim@kpu.ac.kr, peterpan@kookmin.ac.kr, inibest@kaist.ac.kr

Abstract

Multimodal interfaces are the emerging technology that offers expressive, transparent, efficient, robust, and mobile human-computer interaction. In this paper, we described the speech/gesture based multimodal interface systematically from the human factors point of view. To design more practical and efficient multimodal interface, human factors issues such as user modeling, usability studies, speech and gesture interaction/integration, and redundancy are discussed. This paper can be helpful to the researchers in the field of study that enhances the performance of multimodal interaction.

Keywords

Multimodal Interface, Human Factors, Usability, Design Guideline

1. Introduction

In the human centered paradigm, new type of Human Computer Interface (HCI) which is more intuitive and natural is needed as computers are more powerful. It is intuitive that everybody can use without manual or training and natural that human and computer can communicate like as humans communicate each other. This is caused by the change from WIMP (Windows, Icons, Menus, Pointer) environment to the advanced computing environment. HCI technology is the bottleneck in the communication between human and computer because its advance can't catch up the computing environment (Jacob, 1993).

In particular, it is perceived that mouse or keyboards are very difficult input devices to aged or handicapped people. New computing environment (mobile, virtual reality, and ubiquitous) that enables diverse task

beyond existing desktop environment requires more intuitive and natural interface. Of particular interest are multimodal interfaces that use speech and natural gestures to communicate with computers, the same mechanism that humans use to communicate with each other (Sharma et. al., 1998).

Multimodal interfaces are composed of the emerging technologies that offer expressive, transparent, efficient, robust and mobile human-computer interaction. They also are strongly preferred by users for a variety of tasks and computing environments. The following shows that future computing environments that cause the change of human work methods and environments.

Mobile Computing Environments

Mobile computers are computer systems that are portable and wireless such as palm sized personal computer, personal digital assistance (PDA), hand held personal computer, lap-top computer. These are expanding quickly into the big electric appliance market. Mobile computing environments based on the development of telecommunication technology and minimization technology in electronic parts enable users to access to the needed information at anytime, anywhere. These have related with the following ubiquitous computing environments.

Ubiquitous Computing Environments

Ubiquitous computing, or calm technology, is a paradigm shift where technology becomes virtually invisible in our lives. Instead of having a desk-top or lap-top machine, the technology we use will be embedded in our environment. From the ubiquitous computing group in Xerox PARC, we have the following description: imagine a world with hundreds of wireless computing devices of different sizes in the same room. In order to bring this type of computing out into the environment, among the things we need to rethink are user interfaces, displays, operating systems, networks, and wireless communications.

This rethinking demands a radical departure from the tradition of putting machines out for our use, and having us adapt to them. Instead, in the world of ubiquitous computing, technology will be implicit in our lives, built in to the things we use, including the space. The proponents of this technology hold that this type of computing will be a more natural tool, and thus a more powerful and effective one for us to use.

The trend towards embedded, ubiquitous computing creates a need for HCI forms that are experienced as natural, convenient, and efficient. Natural actions in human-to-human communication such as speech and gesture, seem more appropriate for what Abowd and Mynatt (2000) have named everyday computing, and which should be support the informal and unstructured activities of everyday life.

Virtual Reality Environments

Users being experienced in virtual reality (VR), naming artificial reality, cyberspace, virtual world, virtual environment, artificial environment, and augmented reality are surrounded by three dimensional display that computer generated, can move and grasp virtual objects in virtual space beyond conventional two dimensional display. Virtual reality environments represent advanced HCI systems, communicating over several channels of information. The popular view of VR emphasizes the sensorial experience provided through high quality feedback devices, neglecting the user input component.

In reality, a typical platform for developing virtual reality environments has both a computer providing visual and auditory feedback plus an input device controlled by the user. Most of the information is received through the visual channel, similar to the real world. The need for increased immersion and interaction motivates the designers to explore the integration of additional modalities and to take advantage of cross-modal effects. Adding several communication channels comes at the price of system complexity, cost, and of integration/synchronization problems.

Future computing environments are coexistence of mobile computing (beyond desk-top), virtual reality (beyond space, time, risk, and cost), and ubiquitous computing (pervasive, and more intuitive and natural). These mean the extension of traditional computing environments. Novel speech/gesture (including pen) interfaces are candidate for use in future mobile, virtual reality, and ubiquitous applications.

Current widely used computing systems are unimodal in nature, and interaction with such systems is not always satisfactory. They lack robustness and accuracy, as they use only one mode of interaction. If two or more interaction modalities are combined, then

the strict restriction on recognition and accuracy are loosened (Sharma et al., 1998). This also facilitates redundant input and makes the use of such systems easier. Another advantage, with redundant input, is that multimodal interfaces enable people with special needs to access computers easily and efficiently (Bellik and Burger, 1994). Since communication takes place on a number of levels, people with special needs can still interact with standard computer systems. Multimodality will decrease the need to build special purpose interfaces for individuals with disabilities (Mynatt, 1995). Wang, Shahnavaz, Hedman, Padapdopoulous, and Watkinson (1993) found that redundancy between speech output and text display enables the user to shorten the learning time of the interface. Multimodal systems have a shorter learning time than the conventional WIMP systems.

Multimodal interfaces are also needed in situations where users operate under hands-busy or eyes-busy restrictions. Such instances occur in areas of industrial visual inspection, air-traffic control systems, bio-medical tasks, nuclear power plant monitoring/control systems during emergency diagnosis and care, and in other data collection tasks. Oviatt (2003) observed that in applications involving direct object manipulation, like object simulation or map manipulation (requiring both speech and hands), users prefer multimodal interaction to unimodal interaction.

In particular, application areas of emerging technology such as wearable computers, interactive game, and human robots require multimodal interfaces to achieve more suitable interaction. Multimodal interfaces, new type of intelligent human computer interface, are fusion technology combined by artificial intelligence, computer vision, human information processing, human body recognition, cognitive science, and human factors. Sharma et al. (1998) described mapping of different human action modalities to computer sensing modalities for HCI. The hand movement is exploited in the design of numerous interface devices—keyboard, mouse, stylus, pen, wand, joystick, trackball, etc. The next level of action modalities involves the use of hand gestures, ranging from simple pointing through manipulative gestures to move complex symbolic gestures such as those based on American Sign Language. A multimodal framework is particularly well suited for embodiment of hand gestures into HCI. In addition to hand gestures, a dominant action modality in human communication is the production of sound, particularly spoken words.

Oviatt et al. (2003) described that the advances in multimodal interfaces are dependent on hardware advances in new media, the construction of new concepts for multimodal prototype systems, substantial

empirically oriented research with human participants, and the development of appropriate metrics and techniques for evaluating alternative multimodal systems designs. Human factors discovers and applies information about human behavior, abilities, limitations, and other characteristics to the design of tools, machines, systems, tasks, jobs, and environments for productive, safe, comfortable, and effective human use (Chapanis, 1985). Human factors approach to HCI seeks to match the characteristics of the human user with those of the technology in question in the best way possible, so that the interface exploits the capabilities of each while accommodating any limitations. Although advances in speech and gesture processing technology and theory are steadily being made, there are still limitations. As most problems are, one or two researchers can not solve HCI problem because it is interdisciplinary particularly. Problems are solved after all when one can find solution with integrating the result of diverse senior researchers. In this paper, we considered the speech/gesture based multimodal interface from human factors point of view.

In spite of the relatively early beginnings of research in the field of multimodal interaction (Bolt's "Put-That-There" in 1980), the current technical feasibility and clear advantages of multimodal interfaces and the many possible input modalities, the field of multimodal interaction is still young and immature, mostly restricted to laboratory prototypes. As Oviatt and colleagues (2003) mention in their review of the field of speech and pen-based gesture interfaces, one of the factors that is current limiting in multimodal interaction is the lack of substantial evaluation to guide the

interactive development and optimization of these systems.

A well-designed multimodal interface that permits flexibility can potentially leverage people's natural ability to use modes accurately and efficiently. When a new type of speech/gesture input device is applied to a specific task, we have to design the system in the view of human factors to improve user performance although it is well known that users have a strong preference to interact multimodally. Exactly, more human factors research will be needed: usability evaluation considering user preference, performance, and human error behavior; human centered design considering human capacity in the view of cognitive science and engineering theory; human factors design guideline and checklist to check out in the multimodal interface design process. There are a few researches about user test of multimodal interface but they have some limitations that the results are not sufficient to generalize.

This paper does not present new research results through experiment. However, we discussed about the speech/gesture based multimodal interface systematically from human factors point of view. In Section 2, we summarize the application area, the type of speech and gesture of the emerging multimodal interfaces. In Section 3, we discuss the many human factors issues in multimodal interfaces. Finally, in Section 4 we list the research challenges of human factors that remain to be addressed. This paper provides a good beginning point for anyone interesting in starting research in this area.

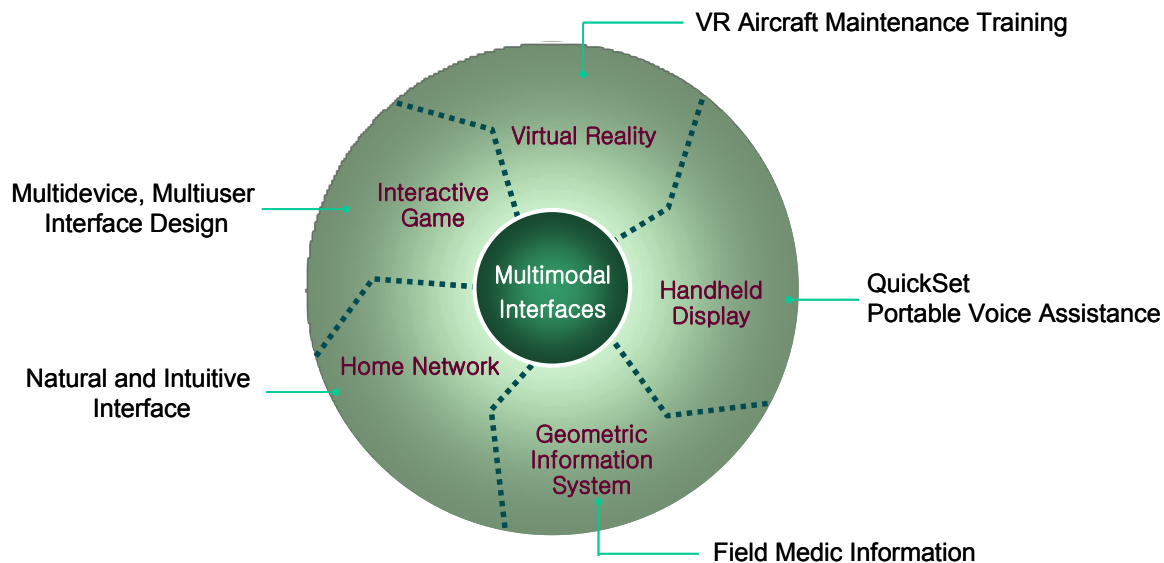


Figure 1. Multimodal Interface required in Advanced Computing Environments

2. Emerging speech/gesture or speech/pen input devices

Representative multimodal input devices (speech/gesture or speech/pen) are OGI's QuickSet system, IBM's Human-Centric Word Processor (HCWP), Boeing's Virtual Reality Aircraft Maintenance Training Prototype (VRAMTP), NCR's Field Medic Information System (FMI), BBN's Portable Voice Assistance (PVA), and PSU's interactive Map (iMAP) system. We can classify emerging speech/gesture based multimodal interfaces according to the application area, type of speech, and type of gesture. Table 1 summarizes the various types of speech/gesture based multimodal interfaces. In multimodal interfaces, one explicit goal has been to integrate complementary modality in a manner that yields a synergistic blend, such that each mode can be capitalized upon and used to overcome weaknesses in the other mode.

A multimodal interaction system, called crisis management (XISM) was completed in 2000 by Penn State Univ. (PSU) and simulated an urban emergency response system for studying speech/gesture interaction under stressful and time-constrained situations. The XISM system was the first natural multimodal speech/gesture based interface to run on a single processing platform holistically addressing

various aspects of the human computer interface design and development issues. The iMAP and XISM systems were developed as part of the Federal Laboratory on "Advanced and Interactive Displays" funded by the U.S. Army. More recently, under a grant from the National Science Foundation, PSU is developing a system called dialog assisted virtual environment for GIS (Geometry Information System). For more details about current multimodal interfaces, refer to the Oviatt et al. (2003) and Sharma et al. (2003).

3. Human factors and design issues

3.1 Multimodal Interaction Modeling Using Cognitive Theory

At the HCI level, we can describe a model of multimodal interaction as shown in Figure 2. The Human Action Modalities (HAM) and Computer Sensing Modalities (CSM) define the input flow, while the Computer Output Modalities (COM) and Human Perception Modalities (HPM) define the feedback flow in the view of interaction. The input flow refers to control and the feedback flow refers to perception at the human side. The information flow can be modeled as a sum of cross-talking perception

Table 1. Application areas of Multimodal Interfaces

<i>Industry</i>	<i>Product (Developed)</i>	<i>Type of Speech</i>	<i>Type of Gesture</i>
Education and Training	Virtual Reality Aircraft Maintenance Training Prototype (Boeing, Duncan et al., 1999)	Grammar-based moderate vocabulary	Pen-based multiple gesture
Medical Service	Field Medic Information System (NCR Co., Holzman, 1999)	Grammar-based moderate vocabulary	Pen-based deictic selection
Office Automation	Human-Centric Word Processor (IBM, Lai, et al., 1997)	Statistical language processing	Pen-based deictic selection
Emergency Management	Multimodal Crisis Management System (PSU, Sharma et al., 2000)	Natural Speech Command	Free hand deictic pointing
Broadcasting	Weather Narration (PSU, Sharma et al., 1998)	Prosody-based Co-analysis	Co-verbal Strokes
Geometry Service	QuickSet (OGI, Cohen, et al., 1997)	Grammar-based moderate vocabulary	Pen-based multiple gesture
Traveling Service	TravelMATE (SRI, Julia et al., 1999)	Simple command	Natural deictic pointing

and control communication channels. A good theoretical insight in the human computer communication system is provided through channel capacity and bandwidth analysis (Schomaker et al., 1995). Capacity and bandwidth requirements for different communication channels vary with the human capacity of information processing (visual, auditory, haptic, etc.) and with the human motor performances (eye motion, body motion). Here, we believe an interdisciplinary approach (including, but not limited to cognitive systems engineering, distributed cognition, activity theory, cognitive ergonomics) to learning about the work domain will assist in the development of multimodal design guidelines.

The channel capacity, bandwidth as well as other parameters (delays, inferences, etc.) specific to each channel are important factors when integrating modalities with redundant representation. Wickens (1992) proposed a model of human information processing and multiple resource theory partly evaluate human cognitive difficulties. We can partly evaluate human cognitive difficulties in the situation of parallel task using this modal and theory. Due to the complexity of human information processing channels, however, no suitable model is available for describing detailed multimodal integration. Therefore improving the design of multimodal interfaces relies mostly on feedback data from human factors experiments.

Psychological research has provided relevant empirical results and theoretical grounding on how humans deploy their attention and coordination two or more modes during the execution of complex task, especially in cases where one mode is visual and the other auditory (Wickens, 1992). More research about this area make attentive interfaces possible which are user interfaces that optimize the attention of system and user within a multitasking situation using measurements and models of the user's attention.

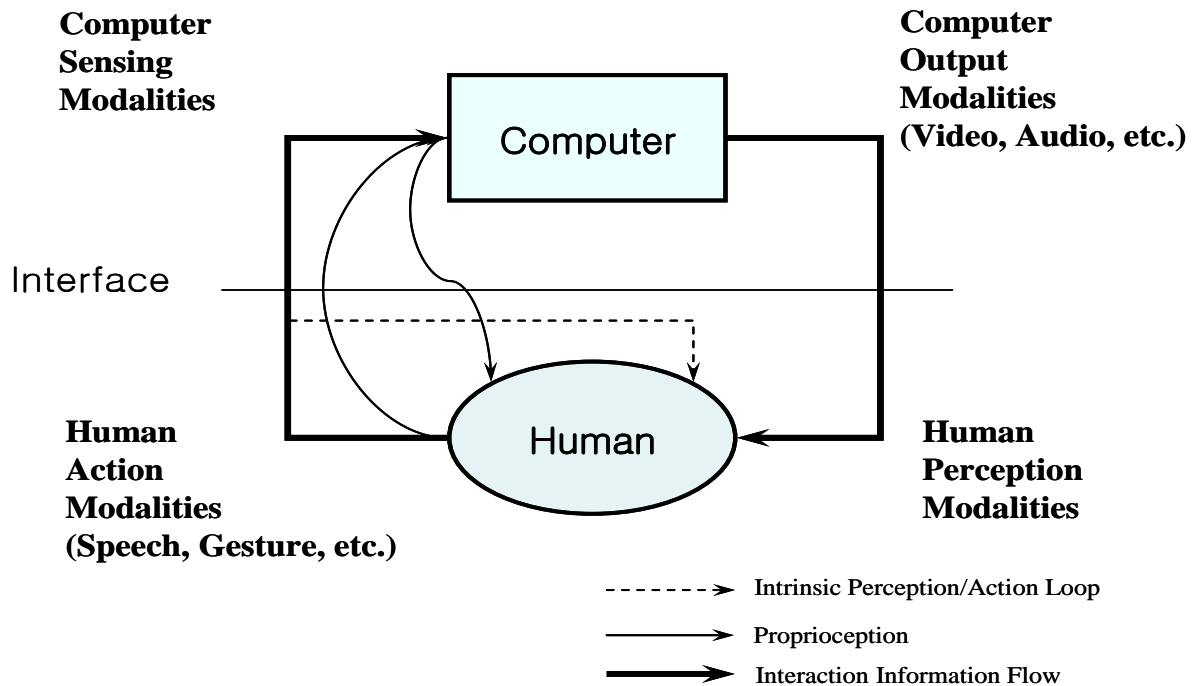


Figure 2. Human Computer Interaction Model; revised from (Schomaker et al., 1995)

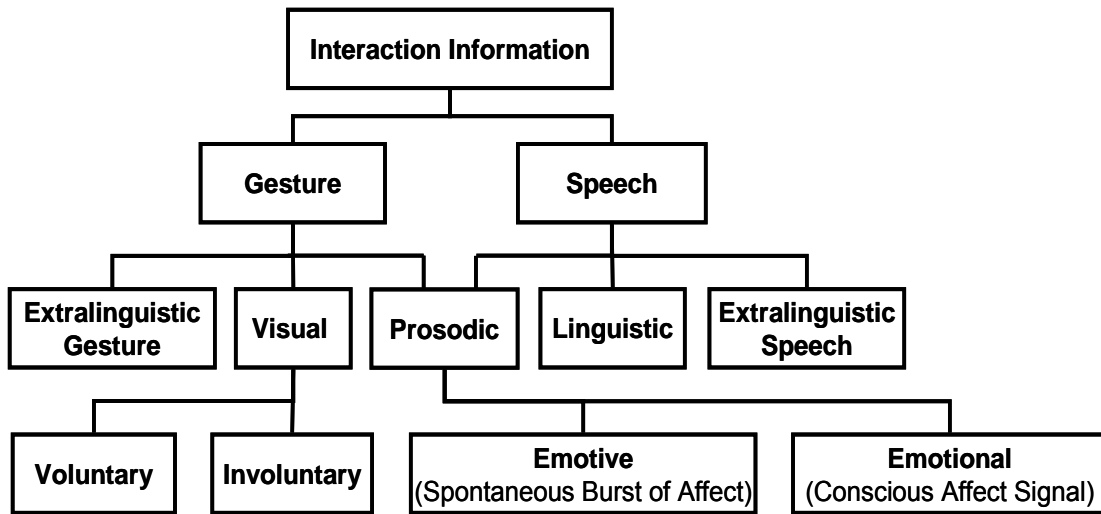


Figure 3. Decomposition of Interaction Information in Multimodal Interfaces

To develop theory-based predictive evaluation techniques, we need more understanding about user model and multimodality through collecting behavioral data and analyzing patterns. Extended GOMS (Goals, Operators, Methods, and Selection rules and CCT (Cognitive Complexity Theory) models are supportive to do that.

Interaction information of multimodal interface consists of gesture and speech. Gesture information is decomposed to extralinguistic, visual, and prosodic actions. Speech information is decomposed to extralinguistic, linguistic, and prosodic utterances. Visual gestures contain voluntary and involuntary actions. Prosodic information contains spontaneous burst of affect (emotive) and conscious affect signal (emotional). Figure 3 shows the described information in multimodal interaction.

Speech/Gesture driven multimodal interfaces demand careful design of information flow between

users and those subsystems that manage various knowledge and data sources. Human Factors specialists have to focus on the multimodal dialogue design beyond the traditional menu design.

3.2 Human Factors in Speech

Two possibilities exist for the use of speech in multimodal interfaces. The first involves the speaking of an artificial command language such as “MOVE” or “QUIT” from the menus of a typical graphical user interface. The other is a natural spoken language such as “Connect to the white house website” or “Wake me on 7 in the morning”.

Figure 4 shows that several speech commands can be used with deictic gestures in XISM (Multimodal Crisis Management System).

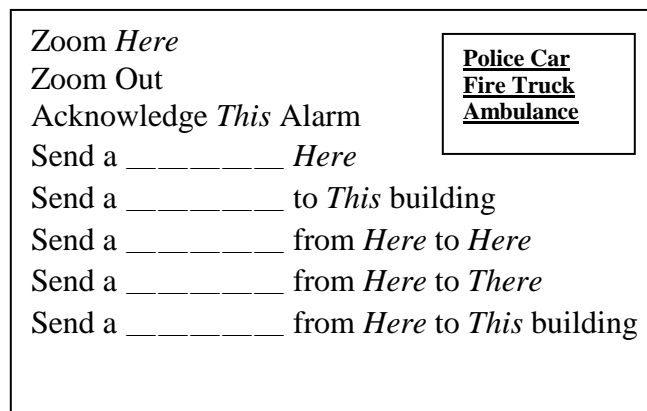


Figure 4. Speech Commands Used in XISM (The italic part of the speech command means that deictic gestures are to be co-occurred)

One issue that needs further study concerns the effect that human short-term memory has on the design of speech-based interaction. That is to devise an approach that allows dialog with users that has the structural advantages of menu interaction and the directness of free from spoken natural language. Other issue concerns the discovery of the optimal “granularity” for spoken commands. Designers need to know what signals need to be sent, what signals need to be recognized and how crucial signal placement is. In particular, human sensitivity to the location of conversational signals within a dialog schema needs to be measured. More thought and a considerable body of practical experience will be required before human factors practitioners can identify classes of application that can make the best advantage of speech recognition and speech synthesis technology. Human speech often exhibits complex patterns of amplitude and frequency variation. In addition, the duration of phonemes change and imbedded silence may or may not be present. These patterns define the “prosodics” of language. Prosodic cues are the area where further investigation could lead to much more natural human/machine communication. The incorporation of prosodic features in speech recognition and automated speech understanding has led to significant improvements in recognition accuracy (Lea, 1980). Furthermore, prosodic cues can also help speed up the recognition process by significantly reducing the search space. In agreement with a review of the ergonomics of automatic speech recognition interfaces (Hapeshi and Johns, 1988), the major ergonomics problems when designing voice driven applications are: 1) The determination of the most effective mode for feedback, 2) How to combine audio and visual feedback, 3) The type of error correction and its implementation, 4) The adaptability of the system (speaker adaptive recognition system, focus shifting, adaptation to the recognition performance), 5) Expert and naïve user models, 6) Modification and selection of the vocabulary by the user (the user should be able to define new terms in the course of the interaction), 7) User adaptation using feedback (e.g. to deal with the Lombard effect), 8) Error collection for future analysis.

3.3 Human Factors in Gesture

Gesture recognition is the process of inferring gestures from the captured motion data. Depending on the motion acquisition method, gestures are parameterized in a variety of ways. Articulated model-based visual approaches or magnetic trackers usually yield time varying joint angles, while pen based

approaches we should be interested in the way people use and respond to hand gestures in conversation, particularly in the specific workplace. Still, gesture varies more from individual to individual than does language itself. It's not possible to point to a particularly American way of gesturing or even a Pennsylvanian or female or ethnic way of gesturing. Researchers are still trying to understand just how much gestures are a function of culture. From these points, we can use only a deictic gesture in multimodal interfaces though the advanced image technology can capture our natural gesture pattern. Multimodal co-analysis of kinematics of hand movement and speech signals provide an attractive means of improving continuous gesture recognition. Gesture-word co-occurrence analysis has been found effective in improving the recognition rate of pen and had gestures. To extend the deictic gesture to natural gesture, more research about perceptual mapping and feedback problems are needed.

3.4 Integration of Speech and Gesture

We believe that to develop a successful multimodal interface it is important to focus on the development of a synergistic integration principle, supported by the synchronization of the multimodal information streams on temporal coherence principle. A probabilistic evaluation of all possible speech/gesture combinations promises a better estimation of users' intent than either modality alone. The conditional probabilities of observing certain gestures given a speech utterance will be based on several factors. Speech utterances will first have to be analyzed keyword classes such as typical deictic keywords (e.g., “this”, “that”). These keywords can then be associated with corresponding deictic gestures. The association needs to take gesture and utterance component classes into consideration and maintain the appropriate mapping between speech and gesture components.

3.5 Usability of Multimodal Interfaces

The International Standard Organization (ISO) has published an emerging standard numbered ISO 9241. Its full name is ‘ergonomic design for office work with visual display terminals (VDTs)’. Part 9 of the standard, called ‘requirements for non-keyboard input devices’, addresses the evaluation of performance, comfort and effort (ISO, 1998).

Several experimental studies have adopted the recommendation of this standard. Various experiments from the past, have conducted for studying and

verifying multimodal interfaces. To generalize the result of user studies, more theory based predictions and formal experimental design are needed. For multimodal interfaces, we also have to adapt and follow reflective HCI practice proposed by Schön (1983).

Most researchers know that there are a lot of difficulties in evaluation of multimodal interfaces: 1) There are no standard benchmark databases. 2) Multimodal interactions depend on user's behavior (not reproducible) and current hardware/software. 3) Evaluation criterion is frequently unclear. 4) Evaluation of qualitative aspect is not reliable. 5) There is meta-cognitive gap. 6) There are many things we don't know about multimodalities.

We suggest some challenging research issues: 1) Performance evaluation approach from low level to high level. 2) Combination of user-based empirical evaluation and theory-based predictions. It is meaningful only when signal level analysis is integrated with higher-order cognitive task analysis and user-

centered design practice. Moreover, we need theory-based predictions to generalize the result of user tests. To compensate for the weakness of experimental result effectively, the second is necessary.

Shneidermann (1992) recommends a more formalized approach for advanced system interface design in order to identify the range of critical usability concerns. Figure 5 shows that an iterative user test process for multimodal interfaces. However, a formalized approach for a multimodal system does not yet exist; therefore, we must piece together elements from several approaches and draw upon a suite of methods for addressing questions about individual and collaborative human work with computer systems.

In many usability experiments, researchers suppose that there are no performance differences between mediated conditions, so they use only indirect and end result measures such as completion time or error rate. To systematize interaction evaluation for multimodal interfaces, we need more objective and quantified in-process measures.

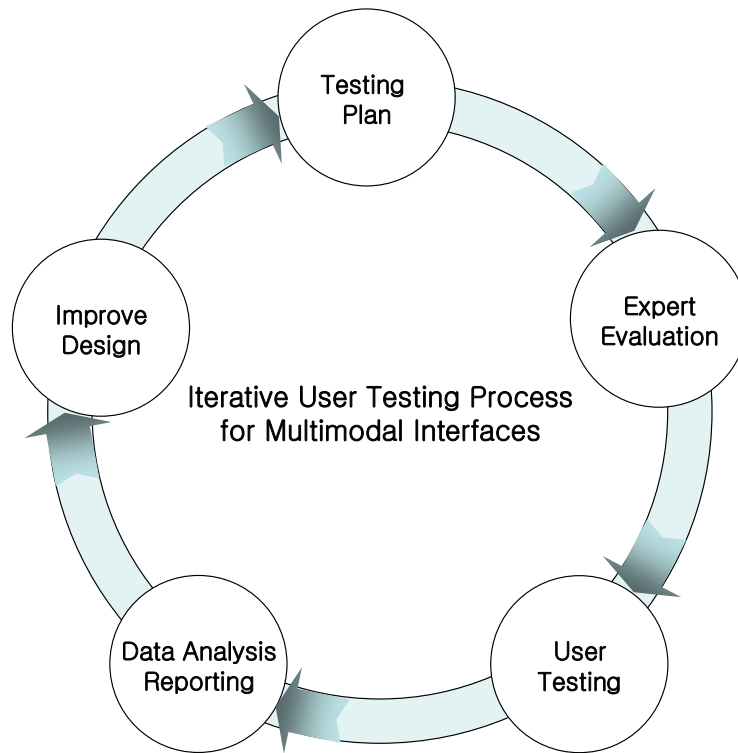


Figure 5. Iterative User Testing Process for Multimodal Interfaces

3.6 Redundancy

When providing redundant feedback it will be essential to consider the limitations of human information processing (Card et al., 1983) to avoid sensorial overload. Multimodal interface designers thus need to establish the most-effective command coordination schemes for their specific tasks.

Redundancy as a phenomenon in multimodal interfaces is analyzed in the light of various theories and models concerning human cognition. The main concern in this part of the work is the human ability to process several messages simultaneously. On the basis of existing research, a suitable theoretical framework is sought for the efforts to avoid combinations of output elements in which messages interfere with the processing of each other. When sensorial redundancy is provided to users, it is essential to consider the design of the integration of these multiple sources of feedback. One means of

addressing this integration issue is to consider the coordination between sensing and user command, and the transposition of senses in the feedback loop (Stanney et al., 1998).

4. Further research topics in the evaluation of multimodal interface

We can divide technologies for multimodal interface into the following core technologies as Figure 6. To design more practical, efficient multimodal interface, more analytic human factors area such as cognitive theory, human information processing and usability evaluation are indispensable. We believe that the following lists are the further issues related to the evaluation of multimodal interface.

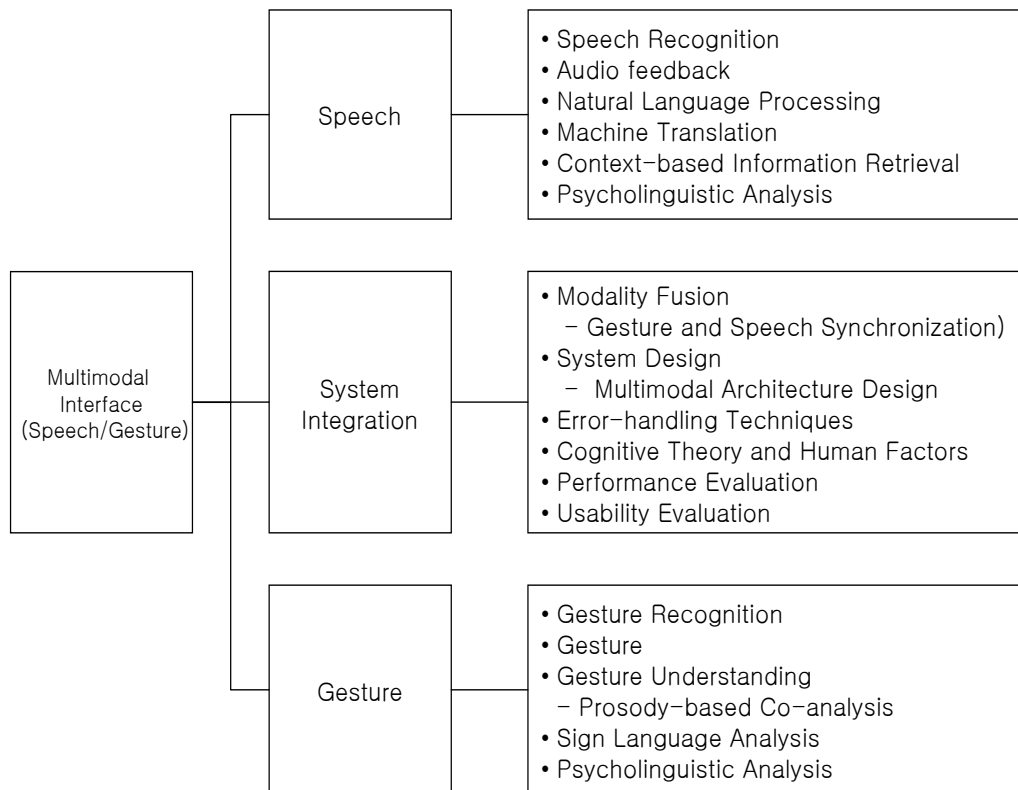


Figure 6. Technology Tree of Multimodal Interface

Understanding Human Information Process

- Linguistics and performance characteristics of natural communication modalities.
- How humans deploy their attention and coordinate two or more modes during the execution of complex tasks (e.g. how to use multimodal input in dual pointing tasks).
- Detailed modeling of human information and sensory process.

Design Guideline of Multimodal Interface

- Development of appropriate measures and methodologies for evaluating alternative multimodal system designs (more objective and quantified in-process measures).
- Suggest an appropriate multimodal interface to a specific application such as graphical manipulation of geometric information system (GIS), wireless handheld PC (mobile and wearable), multimodal watch station of nuclear power plant (NPP) and aircraft cockpit.

Performance Evaluation of Multimodal Interface

- Comparative analysis between unimodal and multimodal interfaces in the view of task performance.
- Determine the impact of a specific multimodal interface on cognitive load as well as take subjective measures such as discomfort and user preference.
- Empirical usability test of a specific multimodal interface with human participants.
- Suggest a systematic methodology for connecting culture, usability and esthetics for designing natural multimodal interfaces.

To solve the further research issues, human factors researchers have to include the following concept of the context.

Complementarity

- What purposes can be served by complementary representations expressed in different media and complementary input operations using different modalities?

User knowledge

- To what extent should metaphors and interfaces rely on users possessing prior knowledge of the context within which the metaphor is instantiated?

Integration, cross-referencing, combination

- To what extent should metaphors be combined and integrated? Do they become more/less effective when such dependencies are introduced?

Cross cultural factors

- Can metaphors and interfaces limit the problems associated with cultural differences or do they exacerbate the problem? Is it possible to design universal metaphors?

5. Conclusion

In this paper, we presented further research topics related with human factors and design issues in multimodal interface. To design more practical, and efficient multimodal interface, interaction modeling using cognitive theory, command policy integrating speech and gesture, understanding and co-analysis of speech and gesture, user test and performance of multimodal interface are to be considered.

Human factors researchers may apply human information theory, user knowledge, task analysis, cross cultural factors, and experimental methods to solve the design issues of multimodal interface.

6. References

- [1] Abowd, G. E., and Mynatt, E. D., "ACM Transactions on Computer-Human Interaction", Charting Past, Present and Future Research in Ubiquitous Computing, (2000), 7(1): 29-58.
- [2] Bellik, Y., and Burger, D., "Conference on Human Factors and Computing Systems", Multimodal interfaces: new solutions to the problem of computer accessibility for the blind, (1994), pp. 267 – 268.
- [3] Bolt, R.A., "Computer Graphics", Put-That- There: Voice and gesture at the graphics interface, (1980), 14(3): 262-270.
- [4] Bradford, J. H., "ACM SIGCHI Bulletin", The Human Factors of Speech-Based Interfaces: A Research Agenda, (1995), 27(2): 61-67.
- [5] Card, S. K., Moran, T. P., and Newell, A., "The Psychology of Human-Computer Interaction", Lawrence Erlbaum.
- [6] Casali, S. P., Williges, B. H., and Dryden, R. D., "Human Factors", Effects of Recognition Accuracy and Vocabulary Size of a Speech Recognition System on Task Performance and user Acceptance, (1990), 32(2):183-196.
- [7] Chapanis, A., "Proceedings of the Human Factors Society 29th Annual Meeting", Some reflection on progress, Santa Monica, CA: Human Factors Society, (1985), pp.1-8.

- [8] Damper, R. I., Speech as an Interface Medium: How can it Best be Used? In "Interactive Speech Technology: Human Factors Issues in the Application of Speech Input/Output to Computers", Taylor & Francis, (1993).
- [9] Hapeshi, K., and Jones, D., "International Reviews of Ergonomics", The Ergonomics of Automatic Speech Recognition Interfaces, (1988), 2: 251-290.
- [10] Jacob, R. J. K., "IEEE Computer", Eye-gaze computer interactions: What you look at is what you get, (1993), 26(7): 65-67.
- [11] Lea, W., Prosodic Aids in Speech Recognition, In "Trends in Speech Recognition". NJ: Prentice-Hall: Englewood Cliffs. (1980), pp. 166-205.
- [11] Mynatt, E. D., "Conference on Human Factors and Computing Systems", Transforming graphical interfaces into auditory interfaces, (2000), pp. 67 – 68.
- [12] Oviatt, S.L., Multimodal interfaces. In "The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications", J. JACKO AND A. SEARS, Eds. Lawrence Erlbaum Assoc., Mahwah, NJ, (2003), chap.14, pp. 286-304.
- [13] Schomaker, L., J. Nijtmans, A. Camurri, F. Lavagetto, P. Morasso, C. Benoit, T. Guiard-Marigny, B. Le Goff, J. Robert-Ribes, A. Adjoudani, I. Defee, S. Munch, K. Hartung, and J. Blauert., A Taxonomy of Multimodal Interaction in the Human Information Processing System. "Multimodal Integration for Advanced Multimedia Interfaces (MIAMI). ESPRIT III", Basic Research Project 8579. Available: <http://hwr.nici.kun.nl/~miami/> (1995).
- [14] Schon, D.A., "The reflective practitioner". Basic Books, New York. (1983).
- [15] Stanney, K.M., Mourant, R.R., and Kennedy, R.S., "Presence: Teleoperators and Virtual Environments." Human factors issues in virtual environments: A review of the literature, (1998), 7(4): 327-351.
- [16] Stanney, K., and Kennedy, R. S., "1996 SOUTHCON Conference Record, IEEE Operations Center", Human Factors Evaluation of Virtual Environments, Piscataway, (1996), NJ, pp. 316-321.
- [17] Sharma, R., Yeasin, M., Krahnstoever, N., Rauschert, Cai, G., I., Brewer, I., MacEachren, A., and Sengupta, K., In "Proceedings of IEEE.", Speech-Gesture Driven Multimodal Interfaces for Crisis Management, (2003), 91(9), pp. 1327-1354.
- [18] Sharma, R., Pavlović, V. I., & Huang, T. S., In "Proceedings of IEEE.", Toward Multi-modal Human-Computer Interface, (1998), 86(5): 853-869.
- [19] Shneidermann, B., "Designing the User Interface: Strategies for Effective Human-Computer Interaction.", Addison-Wesley. (1992).
- [20] Wang, E., Shahnvas, H., Hedman, L., Papadopoulos, K., and Watkinson, N., "Human-Computer Interaction: Software and Hardware Interfaces.", A usability evaluation of text and speech redundant help message on a reader interface. In G. Salvendy & M. Smith (Eds.), (1993), pp. 724-729.
- [21] Wickens, C., In "Proceedings of IEEE International Conference on Systems, Man, and Cybernetics", Virtual Reality and Education, (1992), pp. 842—847.

Authors' bio.

C. J. Lim is currently a Professor in the Department of Game and Multimedia Engineering at Korea Polytechnic University doing research in Virtual Reality and Human Computer Interaction.

Yoinghwan Pan is currently a Professor in the Graduate School of Techno Design at Kookmin University doing research in Human Centered Design and Usability.

Jane Lee is currently a manager in the LG Electronics doing research in Human Factors and System Reliability.