The Gestural Input System for Living Room Digital Devices

Wen-Shan Chang and Fong-Gong Wu

Department of Industrial Design, National Cheng Kung University, 1, Ta-Hsueh Road, Tainan, Taiwan 70101, Taiwan
Michelle33.tw@yahoo.com.tw

Abstract. The focus of this research is to help users learn gesture symbols through semantic perception. With the semantic perception as the basis, qualitative and quantitative analysis will be conducted to analyze digital homes – as exemplified by the gesture symbols in the 3D space of the living room in order to deduce the design principles of different perceptive semantics. Samples of the case studies will be constructed in accordance with this design principle. Inspections and assessments will also be conducted to demonstrate the accuracy and feasibility of this principle. The findings in this research shall serve as reference for the design of interfaces and gesture recognition systems in the comprehensive surrounding. It shall also serve as the design standard for relevant future designs that involve semantic perception and design gesture symbols.

Keywords: home audiovisual multimedia, gesture recognition, gesture symbol, cognition.

1 Introduction

At present, many input devices are available in the market and input techniques are improving by the day. The existing gesture input recognition techniques include: DataGlove and HMM 3D gesture recognition (CMU), Video Camera, and Data-Driven Clustering in gesture recognition (MIT Media Lab, 1995). As the recognition efficiency increased, their usage also became more extensive. They are also gradually applied in our daily lives. Gesture inputs differ from those of hardware inputs as gesture inputs do not pass through other interfaces, making them more intuition-driven and human. However, since they are more difficult to learn, users often feel frustrated or confused the first time they use them. At the moment, a number of researchers have expressed their intent to help users learn through cognition. (Lin Jing, 2006) Based on the users' cognition of 2D symbols and in-depth researches on the present symbology and related applications [1], in this research, the existing gesture input technology shall serve as the basis, and the various home appliance commands in the living room will be integrated based on traditional semantic cognition. Through analysis, they will be converted into gesture symbols in the 3D space. The gesture symbol language in the space will be redesigned, and the users' cognition, recognition, response time, and preference level of 3D space gesture symbols will be probed into.

J.A. Jacko (Ed.): Human-Computer Interaction, Part II, HCII 2009, LNCS 5611, pp. 151–160, 2009.

2 Literature Review

2.1 Visual Cognition

When we see the form of an object, we start to associate it with other things through our sense of 'sight', 'touch', 'smell', 'sound', and 'taste.' Among these senses, associations of image through the sense of sight are the most evident. Baxter (1995) proposed that human visual message processing could be analyzed in two stages. The first stage is the rapid scanning of the overall visual image to obtain the 'form' and 'shape of the image.' This process is fast and not laborious. The second stage of overall visual image processing involves attention given to details of visual components. The visual cognition process in the first stage is prioritized in terms of the overall image. One feature is that the object we see is integral rather than components of the visual image. It is prioritized over the attention-based second stage.

This priority signifies that the visual image obtained in the first stage tends to dominate or affect the visual cognition process in the second stage. Seeing the 'details' of the object, the initial visual image stays in the mind. If we want to look at other images, we must take our eyes off the image we were looking at, close our eyes, or pay attention to other matters. Once a new image pops up, the attention will lead us to engage in detailed recognition of tasks. Users' performance is enhanced though the design displayed, which allows the users to attain information while completing the visual search. Apparently, the 'meaning' of the information can be sent out through visual search. In the visual method, meanings are frequently sent out through our perception ability. Based on these reasons, the interactions between perception ability and interpretation will become more and more important in design displays in the future. Therefore the learning design in this research will allow users to obtain the visual images and related information. Moreover, through cluster classifications, the users will be able to deduce the types of correlations between images and functions.

In turn, it will help the users understand the meaning of individual graphics in the cluster including reduced work loads and enhanced visual search and comprehension.

2.2 Gesture Prototyping

Although attempts have been made to search for ways to recognize graphics independently, the overlapping between perception and recognition has become more and more evident. Many interfaces form image clusters through clustering and designating locations. In addition, research findings show that clustered graphics not only helps users find suitable functions, they also rapidly and effectively allow users to learn about the correlation between graphic functions. In the aspect of perception, it is perceived that the addition of colors and frequently used shapes in conveying messages is rather important. All these are indications of the extensive use of frequently used visual methods when graphics are developed or spread through languages. Therefore, designers need to have the ability to effectively observe these frequently used methods. Although it is true that not all traditional designs are good, it is still an issue to consider. Martha (2006) proposed that the graphic model setup was rather important because it served as basis in design. Users may start to recognize graphics from the graphic prototype to alter the complex and tiny details. The design elements of the graphic prototype tend to lessen the load of the users while learning and assist them to strengthen their memory in terms of cognition.

2.3 Real-Time Fingertip Tracking

An infrared video camera is used to detect the temperature of the fingertip and locate the position of the fingertip. The advantages of using the infrared video camera are that it is not subject to effects of background colors and the brightness of the surrounding light. It accurately positions the fingertip under the normal body temperature range. The principle of fingertip tracking surveys is the position of the previous image based on the fingertip. A Kalman Filter is used to predict the position of the fingertip in a particular image. This process is adopted to distinguish the location of each fingertip. The distinguished fingertip tracking data detected is able to assemble more gesture symbol arrays than the traditional palm tracking.

The symbolic gesture recognition is the same as most gesture recognition systems in which the graphic gesture recognition system adopts the Hidden Markov Model (HMM). The two factors for HMM inputs to recognize fingertip tracking include: (1) No. of fingertip tracking; and (2) 16 discontinuous encodings used to describe the average direction of fingertip shifts.

2.4 Intelligent Home Information Appliances

The origin of the ‘IA’ concept is dated back to 1978. Jef Raskin, a computer engineer of Apple that created the name ‘Macintosh’, had suggested the term ‘Information Appliance’ at the time. Raskin established Information Appliance Inc. in 1982. On the other hand, with regard to the concept of ‘Information Appliance, IA, Sondre Bjella wrote in ‘The Intelligent Home of 2010’ in 1987 that there will be an internet gateway in every future home. It connects to the external telecommunication network and provides home telecommunication functions. It connects internally to the home devices to monitor the home appliances. Due to information and electronic technology advancement, this concept has gradually become a reality. Man’s information applications are now complex and versatile. In IA related discussions, it is often defined as a ‘PC-based intelligent home device.’ Some people also call it the internet application platform-based internet appliance or IA (digital appliance). Simply, it is defined as information-based home appliances. As the demand for global intelligent home appliances grows, all home appliances will be integrated in the intelligent home network in the future. The management platform will connect the external service networks. According to a research report on In-Stat cited by Industrial Technology Research Institute, 448,000 units of smart appliances were sold around the globe in 2002. It is speculated that 3,810,000 units will be sold by 2007. The annual compound growth rate from 2002 to 2007 is 53.4%. Due to technological advancements, home appliances have become prevalent. The information-based home appliance industry and the electronic-based and information-based products will result in the intelligent living system’s becoming an important part of home living facilities. The development of living application technologies that are humanity-based, user-friendly, and convenient will be the trend of the future.

3 Research Planning

This research is divided into two stages, lexical analysis experimentation and gesture symbol design. In lexical analysis related experiments, the commands of audiovisual devices frequently used in home living rooms will first be collected. Subsequently, a usability assessment will be conducted to select the final commands. The selected commands will then be clustered to choose the final 9 functional clusters and 6 product commands in the experiment. After integrating the results above, the theories, principles and current development technologies stated in review literatures as well brainstorming and expert discussions will be adopted to construct new gesture symbols. Finally, they will be compared with the simulated actual use conductions. The results will be evaluated for analysis and discussions. The advantages, disadvantages, and effects of the new gesture symbols will be summarized to provide reference for gesture recognition researchers and interface design staff.

3.1 Lexical Analysis Experiment

In this research, the collection of vocabularies and usability survey analysis within the scope of the research will first be conducted. The research scope and framework will then be narrowed and confirmed. First, the frequently used audiovisual home appliances in living rooms will be surveyed to find out their prevalence, i.e., command function surveys. The use instructions of the 7 products will serve as reference, and the listed command names will be classified. Then, the 57 subjects who have had over 3 years home input system use (34 males and 23 females) will be requested to collect and compile the home input device commands. Among the various home digital devices, the usability questionnaire allows us to further derive at the commands used by most people.

In the usability survey undertaken by the subjects, the subjects are asked in the questionnaire to choose 5 commands in order of frequency of use. These 5 commands are rated on a scale of 1-5. The command of higher use is selected. In the questionnaire survey, representative products are first selected. Select 6 products with the highest use frequency and importance. As to product function, select the representative commands in order of their scores.

3.2 Command Clustering

After selecting the commands, since the product types and commands are complex, they need to undergo command simplification to reduce the load of the user during learning. After that, command clustering will then be conducted. The various basic types and norms obtained from previously compiled literature data and semantic analysis discussions will be adopted along with the 10 design experts in undergoing new gesture symbol command clustering. (Target group members – those that have had over 4 years of design background and have received professional training in design education)

Table 1.

Name	Function	Name	Function
Function 1	Switch, receive/disconnect	Function 9	Menu, emergency mode, mode switch
Function 2	Play, dial	Product 1	Image display
Function 3	Number input	Product 2	T.V.
Function 4	Pause, hold, timer	Product 3	Telephone
Function 5	Stop, eject	Product 4	Light
Function 6	Frequency (left/right), redial	Product 5	Door
Function 7	Amplifier,call,beep	Product 6	Air-conditioning
Function 8	Fast/slow.motion, frequency (up/down), volume,brightness, temperature (high/low),wind speed (fast/slow),wind direction (left/right)		

Lexical clustering consists of 6 products (image display, T.V., telephone, light, door, air-conditioning, etc.) and 9 command functional clusters. (switch, play, number input, pause, eject, frequency (left/right), amplifier, fast/slow motion, menu, etc.) shall serve as reference for design in future researches. It is also of great referential value for converting commands into gestures.

3.3 Gesture Symbol Design

In the gesture symbol design stage, word-concept association and brainstorming are adopted. In order to demonstrate how users detect and assess the issues, they are required to incorporate their design concept or features in the descriptive text. The users' feelings are gathered to assess design features and concepts. By means of collective idea formulating, the users brainstormed for different ideas. The 10 design experts undergo the design of new gesture symbols in this research. (Target group members: those that have had over 4 years of design background and have received professional training in design education)

3.4 Symbol Analysis Assessment

In the gesture analysis and assessment stage, the gesture command symbols obtained and distributed previously are based on the current technologies and one's understanding and analysis of the products. In addition, all the lexical associations obtained in the pre-test and various suggestions shall also serve as important references. They are regarded as the 'meaning' of commands to the users. The gesture symbols that are more difficult to implement technically are deleted or combined if they are identical. Finally, the gestures are divided into two groups, single-handed gestures and two-handed gestures. The 72 single-handed gestures and 76 two-handed gestures make up the 148 formulated ideas.

3.5 Semantic Differential Ratings/Symbol and Set Preferences

After analyzing and assessing the symbols, compile the gesture symbols and conduct a Semantic Differential Ratings/Symbol and Set Preferences. Through the Semantic Differential Ratings/Symbol and Set Preferences, the subjects' preference level for the gesture symbols will be found, which shall serve as reference in determining the gestures and designing the symbols. After compiling the analysis, a subjective assessment on the gesture symbols is conducted. A total of 41 subjects filled out the questionnaire. The scores range from 0-6. The questionnaire covers the gesture design pictures and association design elements. The gesture symbols with higher scores are selected to undergo significance analysis. After the Semantic Differential Ratings/Symbol and Set Preferences, the results obtained by the experts, and the objective factors such as recognition difficulties due to limitations of technology will be taken into consideration to determine the final single-handed and two-handed gesture design system.

4 Experimental Analysis

Foley [4] believes that the quality of the graphic interface design can be assessed through three main criteria including: response time–the users' time spent to complete a task; recall rate–the recall ratio of the user's completing a certain task; and preference–the users' preferences during use. Zwaga [21] expounded two factors, experience and age, and the effects they have on the subjects when assessing the graphics. Cued Response Assessment is the assessment method adopted. The cued response matrix places the illustrations and their respective meanings on the upper and lower part of the questionnaire arrayed in random numbers. The subjects are then asked to fill in the meanings they know below the illustrations. The results are then made into a cued response matrix. In the matrix, the numbers of subjects who answer correctly are recorded on the boxes corresponding to the X-axis (graphic) and Y-axis (meaning of the graphic) and the rest are subjects who have given wrong answers. This shall serve as reference for assessing the cued response of the graphics assessed. If the rate of correct answers is high, the graphic cued response is said to be low; on the contrary, if the rate of correct answers is low, the graphic cued response is said to be high. This research probes into the recall rate through Symbol Identification Test. The cued responses are assessed through cued response tests, and the user' preference of use is found through Semantic Differential Ratings/Symbol and Set Preferences.

4.1 Experimental Method

1. Cued Response Tests: after reading the symbols, the subjects select the gesture commands believed to be the most suitable matches (placed on the same page) to test the cued response.
2. Symbol Identification Tests: All gesture commands are arrayed randomly. The subjects are asked to demonstrate the gesture symbol expressed. The operation error rate and response time are then measured.

3. Semantic Differential Ratings/Symbol and Set Preferences: at the end of the symbol identification tests and cued response tests, fill out the subjective scales to find out the users' preferences.

Ritu (1996) proposed that in order for users to understand the functions of the information system, a good pre-education mode has to be found because the users' software interfaces eventually become a fixed set of operation behaviors (single-handed mode) when using related software in the future. Thus, in reference to related literatures, a set of gesture learning methods are set up including the learning mode and the learning content.

1. ERI Method (Zhai & Kristensson, 2003): Learners must demonstrate the actions to strengthen the learners' memory provided that correct symbols are not displayed. If the learner demonstrates the correct symbols, they will appear twice after determining the order. If the correct symbols cannot be demonstrated, the original order of array will continue.
2. Top-Down Method (Lee & Zhai, 2004): The collective visual space of the alphabet and memory technology establishes an array familiar with the commands. It is more favorable than repeated training. The teaching procedures set up in accordance with the literatures, the screen images and verbal explanations are adopted. The subjects take on the teaching processes in two experiments. The experiment is carried out in two sessions with three tests for each experiment conducted one day apart. The experiment is completed in a span of three days.

4.2 Teaching Content of Gestures

1. Concreteness: Clearly inform the subjects regarding the implications of the gestures. (Gilhooly & Logie, 1980; Paivio et al, 1968)
2. Semantic distance: Describe the correlation of gesture semantics and graphic functions. Semantic linking fosters a closer relationship between commands and graphics. (McDougall et al, 1999)
3. Complexity: Inform the subjects regarding the detailed constituents and complexity of the gestures.

5 Experimental Results

Cued Response Tests: After the second cued response test, the recognition rates for single-handed and two-handed are both 92% or higher. After learning for the second time, remarkable improvements are shown as compared to the first experimental results.

Symbol Identification Tests: In terms of the operation error rate in accordance with the second experiment, in the single-handed shows a P value of 0.267 based on the results of the 'Chi-Square Test'. In other words, the error frequencies of single-handed and two-handed show no significant differences. In terms of the response time in accordance with the results of The Mann-Whitney U Test, the 15 gesture symbols of the observed values, after the average time of the 30 samples of the original observed values are converted into classes, the classes of the single-handed average time total 212.00. The average value of the classes is 14.13. The classes of the two-handed average time total 253.00. The average value of the classes is 16.87. The P value is 0.412, showing that no significant differences exist between the average time of single-handed and two-handed.

Semantic Differential Ratings/Symbol and Set Preferences: Based on the results of The Mann-Whitney U Test, the 15 gesture symbols of the observed values, after the 30 samples of the original observed values are converted into classes, the P value is 0.683. In other words, there are no significant differences in preference level between single-handed and two-handed.

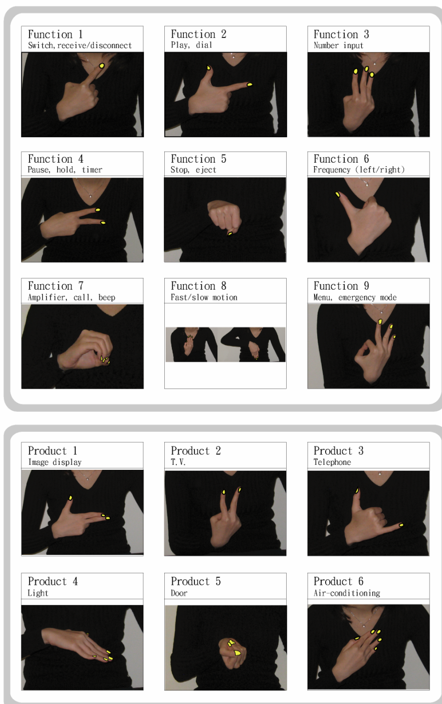

Fig. 1.

6 Conclusion and Recommendation

The single-handed and two-handed designs show no significant differences in terms of assessment results. All the gesture commands have reached the applications of a set of public label symbols stated in the public sign design standard operation procedures of the International Organization for Standardization. (ISOTC 145/SCI) The ISO recommends that a correct recognition rate of 67% or higher be reached. However, discussions can be conducted on the subjects' opinions and the operation conditions. Among the 24 subjects, 7 expressed that out of the single-handed and two-handed gestures, the two-handed gestures are preferred, 15 subjects prefer single-handed, 2 subjects expressed no preference over single-handed or two-handed gestures. Among the users that prefer single-handed gestures, the reasons that attribute to this preference include: single-handed gestures are easier to remember and are more convenient as the commands can be demonstrated with one hand. It is intuition-based that the gestures can be memorized using symbols. The actions are so ‘cool’ that they require no extra actions. The design logics of two-handed gestures have less uniformity. Among the subjects that prefer two-handed gestures, the reasons that attribute to this preference include: they find it more difficult to relate to single-handed gestures. They do not know what to do with the other hand. When matched with actions, they are able to comprehend and remember the gestures. The design disparity is also greater and the gesture variations are more versatile. On the other hand, single-handed commands are less ‘friendly.’

Although single-handed and two-handed gestures show no significant differences in the experimental results, after the experiment and expert discussions, the single-handed gestures are believed to have higher development potentials than two-handed gestures in the future. Single-handed gestures have limitations in terms of the design of actions. Nevertheless, convenience will remain an important factor for future gesture designs. In single-handed commands, the actions of function 5, function 7, and function 3 with 0 in number are more similar. Confusions are likely to result in recognizing them. Overtly similar gesture symbol designs should therefore be avoided during designing.

Furthermore, ‘the effects on culture and language’ are also an important factor. Take this research for instance, the switch function of the T.V. and telephone is expressed as ‘ON/OFF’ in English, but in Chinese, it is translated as ‘to open’, ‘to play’, etc. There are misunderstandings and confusions in cognitive integrations. Future researches may probe into whether or not semantic cognitions differ in countries of different languages and cultures. Also, the concept of gesture inputs not only enhances the convenience in life, the use rights of minority groups such as the visually impaired, verbally impaired, and seniors have to be taken into consideration with more care and creative ideas. Finally, the inclusion of this user group as test subjects will make the research more comprehensive and user-friendly.

References

1. Christian, A.S., Tombre, K.: Architectural Symbol Recognition Using a Network of Constraints. Pattern Recognition Letters 22(2), 231–248 (2001)
2. Fisher, J.: Testing the Effect of Road, Traffic sing Information Value on Driver Behavior. Human Factors 34, 231–237 (1992)
3. Fiske, J.: Introduction to Communication Studies (1995) ISBN 9573225204
4. Foley, J.D., Wallace, V.L., Chan, P.: The Human Factors of Computer Graphics Interaction Techniques. IEEE Computer Graphic and Application 4, 13–48 (1984)
5. Gittins, D.: Icon-based Human-Computer Interaction. International Journal of man-Machine Studie 24, 519–543 (1986)
6. Horton, W.: The Icon Book Visual Symbols for Computer Systems and Documentation. John Wiley & sons, Inc., Chichester (1994)

7. Neilson, J.: Top Ten Mistakes in Web Design. Sunsoft Distinguished Engineer (1996); Caplan, S.: Using focus groups methodology for ergonomic design. Ergonomic 33(5), 527–533 (1990)
8. Neilson, J.: Usability Engineering, pp. 97–155. Ap Professional Press, California (1993)
9. Nespoulous, J.L., Perron, P., Lecours, A.R.: The Biological Foundations of Gestures: Motor and Semiotic Aspects. Lawrence Erlbaum Associates, Hillsdale (1986)
10. Nordin, A. (ed.): Taxonomy and phylogeny of Buellia species with pluriseptate spores (Lecanorales, Ascomycotina). In: Symb. Bot. Ups., vol. 33(1), pp. 1–117 (2000)
11. Oka, K., Sato, Y.: Real-Time Fingertip Tracking and Gesture Recognition University of Tokyo Hideki Koike University of Electro-Communications, Tokyo (2002)
12. Piaget, J.: Construction of reality in the child. Routledge & Kegan Paul, London (1954)
13. Schlenzig, J., Hunter, E., Jain, R.: Vision based gesture interpretation using recursive estimation. In: Asilomar Conference on Signals (1994)
14. Starner, T.: Visual Recognition of American Sign Language Using Hidden Markov Model, Master's Thesis, MIT Media Laboratory (February 1995)
15. Starner, T.: Visual recognition of American Sign Language using Hidden Markov Models. Master's thesis, MIT Media Lab (July 1995)
16. Turing, M., Hannemann, J., Haake, J.M.: Hypermedia and Cognition: Design for Comprehension. Communications of the ACM 38, 57–66 (1995)
17. Wallschlaeger, C.: Basic Visual Concepts and Principles for Artists, Architects and Designers, Times Mirror Pte Ltd. (1996)
18. Weimer, D., Ganapathy, S.K.: Interaction techniques using hand tracking and speech recognition. Multimedia interface design, 109–126 (1992)
19. Whitaker, L.A., Sommer, R.: Perception of Traffic Guidance sing Containing Conflicting Symbol and Direction Information. Ergonomic 29(5), 699–711 (1986)
20. Wickens, C.D., Hollands, J.G.: Engineering Psychology and Human Performance, 3rd edn., pp. 513–550. Prentice-Hall Inc., Englewood Cliffs (2000)
21. Zwaga, H.J., Boersema, T.: Evaluation of a set of Graphic Symbols. Applied Ergonomic 14(1), 43–54 (1983)