FLEXPHORES: A FLEXIBLE INTERACTION FOR WEB BASED PERSONAL DIGITAL PHOTO RETRIEVAL SYSTEM

N.A. Ismail[1], A. O'Brien[2]

[1]Faculty of Computer Science and Information Systems
University Teknologi Malaysia
81300 Skudai, Johor

[2]Research School of Informatics,
Loughborough University, LE11 3TU, United Kingdom.

Email: [1]azman@utm.my, [2]A.O-brien@lboro.ac.uk

Abstract: This paper describes the development of multimodal user interface for web based personal digital photo retrieval (FlexPhoReS) prototype. FlexPhoReS is an experimental system that enables digital photo users to accomplish photo retrieval tasks (browsing, keyword and visual example searching (CBIR)) using either mouse and keyboard input modalities or mouse and speech input modalities. It extends input modalities of web based photo retrieval technologies by offering alternative input modalities through a multimodal user interface in the World Wide Web environment. Our user study with 20 digital photo users showed that the prototype user interface for web based personal digital photo retrieval system is acceptable to the users.

Keywords: Photographs, multimodal user interface, multimodal interaction, personal digital photo, speech interface.

1. INTRODUCTION

In everyday life, people already have large collections of printed personal photos and the new digital photo technology helps collections grow further. This technology has resulted in more people having large personal collections of digital photos and sharing them with others, especially with their family and friends. Web based photo galleries are one of the methods of sharing that allow photo users to publish a collection of digital photos online in a centralised and organised way [1]. A recent survey conducted by InfoTrends/CAP Ventures, showed that 26% of Internet users had photos posted to an online photo service in 2004, a rise of 19% from 2003. In 2004 there were over 825 million photos stored at online photo services, and sharing and printing activities continue to increase [2]. It is therefore more and more

collections of digital photos are now connected to the internet and this has encouraged the development of digital photo systems to provide user friendly and flexible user interfaces to retrieve the growing libraries of large personal photo collections [3]..

In web applications, graphical user interfaces (GUIs) have become the user interface of choice. For many years, they have provided the user with a common look and feel, visual representations of data and direct control using mouse and keyboard input modalities as standard input devices. However GUIs only become possible to implement when computer hardware can produce accurate bitmap displays and can interactively manipulate accurate screen presentations to the users [4]. Moving beyond mouse and keyboard, multimodal interfaces are expected to be more transparent, flexible, efficient and powerful for human-computer interaction [5].

Numerous theoretical and empirical studies have investigated the potential of multimodal interfaces. The two most relevant studies are by Rodden and Wood [6] and by Käster [7]. Although these two studies have some similarity, Rodden and Wood [6] used speech annotation but did not include speech retrieval in a study of personal digital photos. Käster [7] make used of speech retrieval for stand-alone content based image retrieval for expert digital image users. Speech also has been mainly used for annotation in recent implementations. This is in an attempt to simplify the process of adding text to images for organisation purposes. Searching, however, is done by typing in the term to match the speech annotation [8]. Apart from Käster [7] no system has been developed and evaluated that uses speech for searching personal digital photos on the web which provides this research with an opportunity to create such a system.

In this paper we describe the development of multimodal interface for web based personal digital photo retrieval (FlexPhoReS) prototype which enables digital photo users to accomplish photo retrieval tasks (browsing, keyword and visual example searching (CBIR)) using either mouse and keyboard input modalities or mouse and speech input modalities. It extends input modalities of web based photo retrieval technologies by offering alternative input modalities through a multimodal user interface in the World Wide Web environment.

2. METHODOLOGY

We derived our approach in three stages. The first stage involved developing the model for FlexPhoReS for personal digital photo collections. The second stage concentrated on the development of the prototype. The third stage evaluated the prototype to find its acceptability with the digital photo users.

2.1 Modelling the FlexPhoReS

For the modelling stage, identifying users' needs and requirements were addressed. The aim was to understand as much as possible about the users. Along the review of related literatures, a data gathering exercise involving structured interviews was carried out with a small group of digital photo users on the subject of how they organise and retrieve their digital photos. The data gathering exercise conducted was used to gather information and to provide additional input into the construction of the model. A total of seven digital photo users, individuals between twenty and forty years of age who had more than 100 personal digital photos were selected to participate in the study. Both the literature review studies and interviews formed the basis of the conceptual model, through task description of the proposed system model. The model produced, in turn, was used to design the prototype. The initial design was sketched on paper and then mocked-up for an interactive version of a low fidelity prototype.

2.2 Developing the FlexPhoReS

For the second stage, an interactive version of the prototype was built based on the task descriptions. A use case task description was used to describe user tasks with the prototype. It focused specifically on the interaction between the user and the prototype system. Based on the task description, the FlexPhoReS paper prototype (low fidelity prototype) was developed. The intention was to clarify requirements and enable draft interaction designs, screen designs and learning how to use the system to be simulated and tested. Rough screen design was used for the initial design of each individual screen. Microsoft PowerPoint with audio output was used in designing the low fidelity FlexPhoReS prototype. Three subjects participated in the low fidelity prototype informal evaluation. All of them were digital photo users aged between 17 and 40, computer literate and experienced in using Internet and web applications. The participants were asked to carry out a realistic task. Each participant simulated pointing and clicking using a pencil, simulated typing by writing on paper and used their voice for speech input while the researcher clicked an appropriate screen for viewing the outcomes of the user task and explained what happened for each selected task. After completing the tasks, they were asked to give feedback related to the content and structure, screen design and learning how to use the system through a structured interview. The informal evaluation was input to the redesign and development of the high fidelity prototype.

2.3 Evaluating the FlexPhoReS

The third stage of this research concentrated on the evaluation of the prototype. The purpose of the evaluation was to find prototype acceptability with the digital photo users. Data was gathered after the participants interacted with the prototype and filled in an evaluation form of acceptability questionnaire modified and developed based on Questionnaire for User Interaction Satisfaction (QUIS) [9]. The set of questions concerning acceptability was divided into two categories namely (i) suitability and (ii) flexibility. It measured acceptability attributes on a 9-point scale.

The major evaluation was carried out with a total of 20 participants. All of the participants had digital photo collections and were experienced in using some of web based digital photo retrieval features (e.g. keyword searching and browsing). They also had considerable experience with computers.

3. RESULT AND DISCUSSION

3.1 The proposed model design

Our key findings from these modeling sessions were that:

- Browsing or searching with text via a set of user-oriented categories would give more specific access points and therefore better retrieval;
- There was an enthusiastic response to the prospect of effective automatic content based retrieval;
- The indication was very clear that the respondents were positive towards using speech interface as an additional to the graphical user interface in the photo retrieval application;
- A secure World Wide Web environment which allow users to access their photos at anytime and anywhere would give more flexibility tasks and transparent to the users.

The outcomes of the interviews informed the FlexPhoReS prototype development and task description. The descriptions of user tasks with the 'use case' were developed to express and

envision the user tasks for photo retrieval. A 'use case' for retrieving digital photos with chosen interaction modes in the World Wide Web based environment is as follows:

1. The system prompts user for a valid user id and password.
2. The system provides user with the interaction modes option.
3. The user logs in into the system with chosen interaction modes.
4. The system checks user authentication.
5. The system prompts the user with the retrieval strategies.
6. The user retrieves photos using the chosen retrieval strategy with chosen interaction modes.
7. The system searches and displays the search results based on user input.
8. The user browses the retrieval results.
9. The user refines the search if necessary.
10. The user logs out of the system.

Alternative courses:

4. If the user id or password are invalid.
4.1 The system displays an error message and speech prompt.
4.2 The system returns to step 1.
8. If inappropriate photos are found and displayed
8.1 The user can do retrieval refinement.
8.2 The system returns to step 5.

Figure 1.0 shows the 'use case' diagram for the proposed prototype system showing five use cases and one actor.

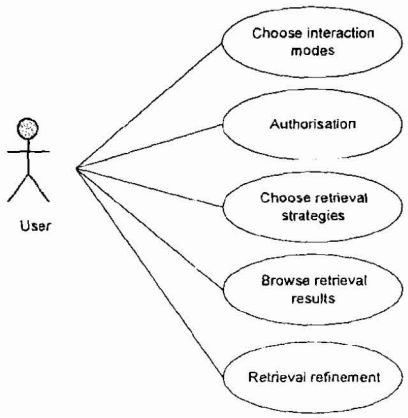


Figure 1.0: 'Use case' diagram for the photo retrieval system

The tasks description and the 'use case' diagram for the proposed photo retrieval system serve as the basis for the process model of FlexPhoReS. The process model (Figure 2.0) is used to describe the structure of the prototype system and represents the process flow.

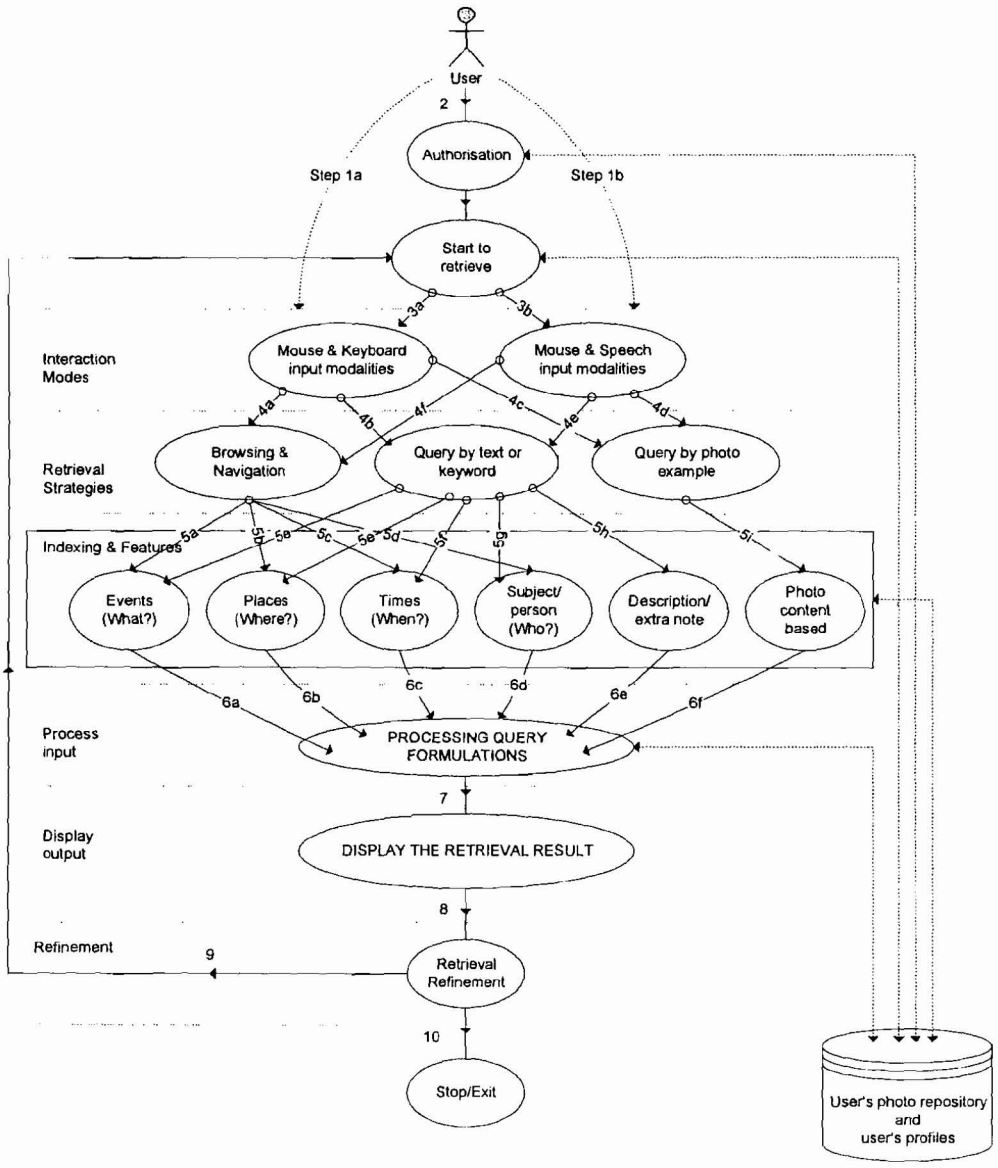


Figure 2.0: Process model of FlexPhoReS

Different users with different profiles retrieve their photos through the user interface. This stage consists of a set of user tasks and interaction modes which define the communication

between the user and the photo retrieval system. Within the multimodal interaction modes, users can interact with the system either using mouse and keyboard or using mouse and speech input modalities. They can also switch between these input modalities to suit their style and interest. With the chosen interaction mode (step 1a or 1b) initially the user logs into the system with a valid user id and password (step 2).

The login information entered is compared with the information stored in the authorisation database which includes detailed information about users such as name, user identification (id), password and numbers of photos in the collection. To retrieve photos, users not only must understand how to use the interaction devices but also give careful consideration to understanding the retrieval strategies. Photo browsing and navigation, query by text or keywords and query by visual example are among the typical retrieval strategies employed in the retrieval process. With the chosen input modalities (step 3a or 3b), the photo retrieval tasks starts with the selection of appropriate retrieval strategies which represent the user query formulations (step 4a to 4f and step 5a to 5i). These query formulations are triggered by the user's information need.

Once the search begins, the users are expected to wait until the search process is completed (step 6a to 6f). Then the retrieval results are displayed which enable users to view (step 7). Users can stop retrieving or exit if they are satisfied with the retrieval results (step 10). However, in some cases, users need to reformulate the search statement and perform a new search (step 8 to 9). To support this retrieval process, all of the digital photos must be indexed based on their photo features.

3.2 The FlexPhoReS

FlexPhoReS is an experimental system for web based personal digital photo retrieval. The system user interface use multimodal interface which blend speech and graphical user interface (S/GUI) to control and retrieve their digital photo collection. Currently it has 30 digital photos and has been annotated manually by 'what?' (event), 'where?' (place), 'who?' (subject/people) and 'when?' (time). It allowed user to use either mouse and speech input modalities or mouse and keyboard input modalities to perform the same photo retrieval task. Therefore the user could select the input method that best suits the tasks. Figure 3.0 gives an overview of the FlexPhoReS user interface which has the following abilities:

- Browse photos by event (what?), by place (where?), by people/subject (who?) and by time (when?) from user photo repository web database.

- Control or navigate through the system. For example to logout from system, go to main page, retrieve system help, and go to next page and previous pages.
- Search photos from user photo repository web database by text or keywords.
- Search photo by visual example from user photo repository web database.

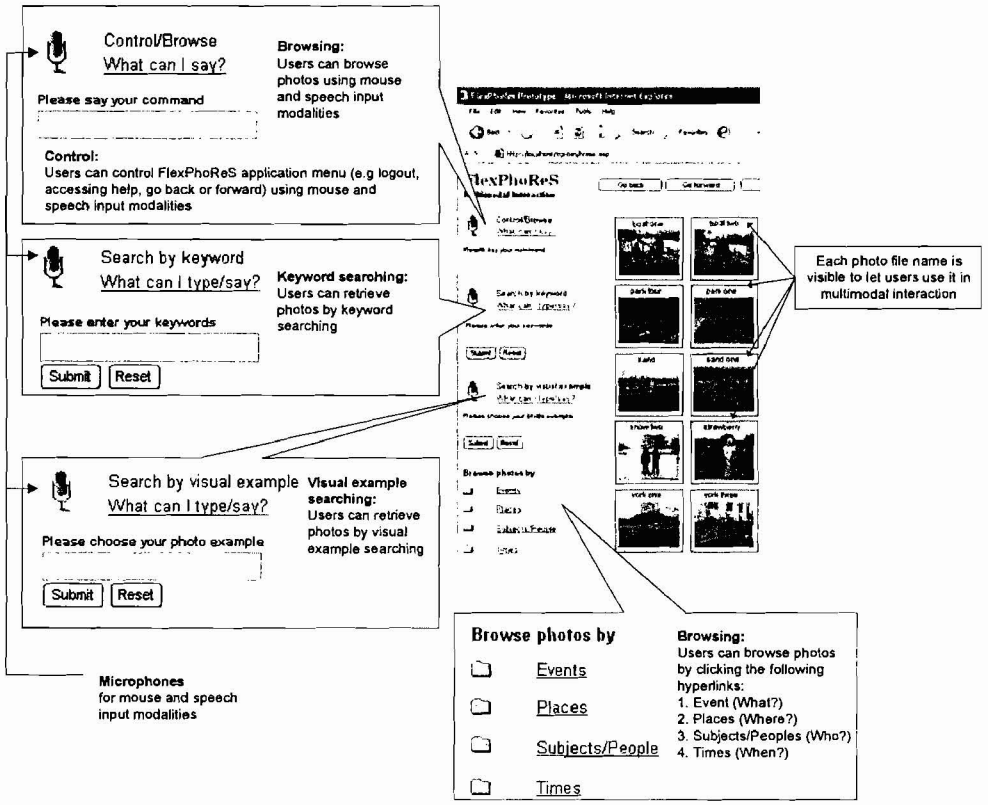


Figure 3.0: Overview of the FlexPhoReS user interface

Figure 4.0 shows the schematic diagram of the web based FlexPhoReS system. The system architecture consists of two sections, namely, client and server. All FlexPhoReS programs and data files including the user's photo repository, profiles, dialogues, grammars, prompts and retrieval engine are stored and located in the web server. No information is kept on the client side.

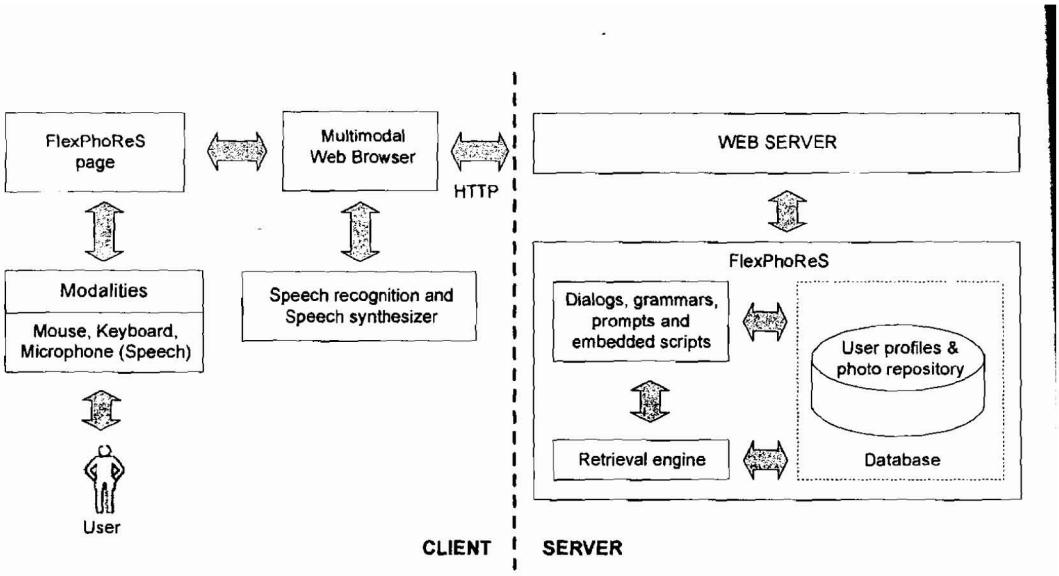


Figure 4.0: Schematic of FlexPhoReS

The client machines run the web browser and the server machine runs the web server. Microsoft Internet Information Services (IIS) web server and MATLAB web server were used to deploy FlexPhoReS in the World Wide Web environment. Microsoft Internet Information Services (IIS) web server is a platform to enable information publishing on the Internet while the MATLAB web server is used to power the visual retrieval function in FlexPhoReS. It comprises a combination of M-files, Hypertext Markup Language (HTML) and Active Server Pages (ASP). The system is initiated when the user enter a FlexPhoReS URL in a web browser; the Web server opens the FelxPhoReS application's default page. The Web server sends HTML, SALT, and JavaScript to the client machine. SALT markup in the pages that the Web server sends to the client can trigger the speech recognition and text-to-speech synthesis engine. For text-to-speech synthesis, the prompt element is used to specify the content of the audio output. Speech recognition, or speech-to-text, involves capturing and digitizing the sound waves. Then FlexPhoReS processes the words and compares them with the FlexPhoReS grammar (XML tag suite) which is a structured collection of words or phrases that the FlexPhoReS recognises and attempts to match human patterns of speech. In FlexPhoReS, the listen element is used for speech recognition. Listen element contains one or more grammar elements, which are used to specify possible user inputs. Figure 5.0 illustrates the listen behaviour of speech recognition events timeline that used in FlexPhoReS prototype system.

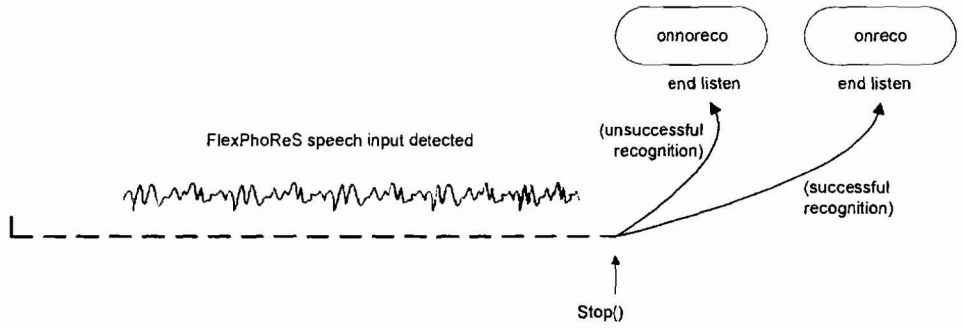


Figure 5.0: FlexPhoReS speech recognition event.

ıe FlexPhoReS has five following key components:

- User profiles, it includes detailed information about users such as name, address, email, user identification (id) and password.

- Repository and access control, a repository is a user's photo database, while access control manages the user security aspect of the system model. The access control issue is handled by authorisation of the client during the login into the system. Access into the system is managed by requiring the user to enter a valid username and password. The information entered is compared with the information stored in the authorisation database. Any attempt to access the FlexPhoReS web page will result in a login being displayed on the client's multimodal web browser. The client is required to enter a user identification (id) and password. The information is passed back to server program on the web server. The user id and password are then compared against the authorisation file. The FlexPhoReS starting page which contains valid user's photo repository is displayed when the user id and the password are confirmed valid. Otherwise a warning message is generated and display to the client. This in turn will determine the authorised user of the photo repository.

- Retrieval strategies which includes query by text or keywords, query by image content, browsing and navigation.

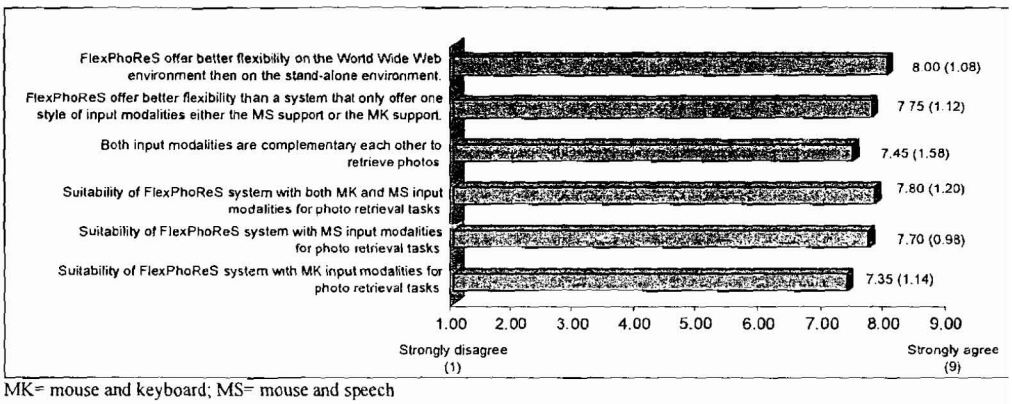
- Photo features. To support retrieval strategies, photo features of the FlexPhoReS are based on text features and visual features. Text features including browsing and searching by text keywords using Structure Query

Language (SQL) which allow users to retrieve their photo by people/subject (who?), by place (where?), by event (what?), by time (when?) and by description/extra note. Whilst, the visual indexing is limited on the aspect of a visual retrieval algorithm. Microsoft Internet Information Services (IIS) web server and MATLAB web server are use to deploy the FlexPhoReS in the World Wide Web environment. Microsoft Internet Information Services (IIS) web server is a platform to enable information publishing on the internet while the MATLAB web server is used to utilise visual retrieval function in FlexPhoReS. It comprises a combination of M-files, Hypertext Markup Language (HTML) and Active Server Pages (ASP).

- Multimodal interface refers to the style of interaction which enables users to use a speech and graphical user interface in photo retrieval tasks. Internet Explorer version 6.0 with speech Add-on is used as a multimodal web browser which allow users to run speech technologies along with keyboard and mouse for multimodal interaction. In speech recognition, the application processes the words and compares the user's speech input with the application grammar (XML tag suite) which is a structured collection of words or phrases that the application recognizes.

3.3 Users' acceptability of the prototype

Participants' acceptability scores with FlexPhoReS were collected. Means and standard deviations (in parentheses) of data collected through the acceptability questionnaires are shown in Figure 6.0. Reliability of the questions was assessed using Cronbach's Alpha, yielding a value of $\alpha = 0.882$, which indicates the questions were highly reliable.



MK= mouse and keyboard; MS= mouse and speech

Figure 6.0: Acceptability rating and (standard deviations) of FlexPhoReS participant's

Analysis of the acceptability of FlexPhoReS revealed that all of the related questions achieved positive scores (above 4.5). The most favourable responses among the participants was related to the flexibility of FlexPhoReS in the World Wide Web environment rather than in a stand-alone environment (Mean=8.00, standard deviation= 1.08). The least favourable response was related in answer to the question on the complementary nature of the two modalities (Mean=7.45, standard deviation=1.58).

Overall the results revealed that all of the participants agreed that mouse and keyboard input modalities by themselves are suitable for photo retrieval tasks. They also agreed that mouse and speech input modalities alone are suitable. An acceptability rate was also given when both input modalities were considered together and the majority of participants agreed that both input modalities are complementary to each other in retrieving photos. Several participants stated that both input modalities were user friendly, practical and easy to use. A number of participants noted that mouse and speech input modalities were more interesting and easier for retrieving photos instead of mouse and keyboard input modalities. They noted, however, that mouse and speech input modalities are very sensitive to noise. Among the suggestions made was that noise reduction is essential to improve FlexPhoReS system performance.

Exploration of the flexibility of the prototype showed that the majority of participants agreed that FlexPhoReS system offers better flexibility than a system that only offers one style of input modalities. In terms of system platform environment, the majority of the participants agreed that FlexPhoReS system offers better flexibility on the World Wide Web environment than on a stand-alone environment.

4. CONCLUSION

In conclusion, this paper has demonstrated that the prototype of flexible user interface for web based personal digital photo retrieval system is acceptable to the users. The prototype is successful in providing a flexible model of the photo retrieval process by offering alternative input modalities through a multimodal user interface. This interface provides the user with multiple and alternative modes of interfacing with a web based personal digital photo retrieval system beyond the traditional mouse and keyboard input modalities. It allows users to select the modality that best fits the situation and is not likely to replace traditional modes of input, but seems to have a useful place along with other types of input modalities in the system user interface. Therefore the FlexPhoReS flexibility of a web based personal digital photo retrieval

system user interface potentially could provide a usable system for more individuals in more various environments.

ACKNOWLEDGEMENTS

We would like to thank the Universiti Teknologi Malaysia (UTM) for their continuous support in the research work.

REFERENCES

[1] House, N., M. Davis., Y. Takhteyev., M. Ames., and M. Finn. 2004. From 'what?' to 'why?': the social uses of personal photos. <http://www.sims.berkeley.edu/.../vanhouse_et_al_2004a.pdf>

[2] O'Keefe, M. 2004. Online photo service usage continues to grow. <http://www.infotrends-rgi.com/home/Press/.../../11.17.2004.html>

[3] Shneiderman, B. and H. Kang. 2000. Direct annotation: a drag-and-drop strategy for labeling photos. Fourth International Conference on Information Visualisation. London. IEEE Computer Society, pp. 88-98.

[4] Roope, R. 1999. Multimodal human-computer interaction: a constructive and empirical study. PhD thesis. Tampere University.

[5] Oviatt, S. 2003. Multimodal interfaces. In: Jacko and Sears (eds.), The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications. Lawrence Erlbaum Associates, Inc, pp. 286-304.

[6] Rodden, K. and K. Wood. 2003. How do people manage their digital photographs? Proceedings of the SIGCHI conference on Human factors in computing systems. Florida. ACM, pp. 409-416.

[7] Käster, T., M. Pfeiffer., and C. Bauckhage. 2003. Combining speech and haptics for intuitive and efficient navigation through image databases. Proceedings of the 5th international conference on Multimodal interfaces. Vancouver. ACM, pp. 180-187.

[8] Chen, J., T. Tan., and P. Mulhem. 2003. Using speech annotation for home digital image indexing and retrieval. Content Based Multimedia Indexing Conference, CBMI 2003. Rennes, France. IEEE Computer Society, pp. 195-200.

[9] Shneiderman, B. 2005. Designing the user interface. Boston: Addison Wesley.