INTERNATIONAL JOURNAL OF COMPUTERS
Issue 2, Volume 2, 2008

A usability study of multimodal interfaces for the presentation of internet search results

A. Ciuffreda and D. Rigas

***Abstract*—**This article describes a usability evaluation of three interfaces that presented results from queries through Internet search engines. One interface used purely text in order to communicate attributes of retrieved results; the remaining two interfaces made use of different combinations of text and metaphors of visual and aural nature. This usability study was based on executions of Internet search activities with specific sets of keywords and on users' views. Five criteria were considered for measuring the degree of usability of the three interfaces. The results obtained from this investigation have shown that a combination of text, graphical objects and short speech messages increased the level of usability in interfaces of this class.

***Keywords*—** Interface, multimodality, search engines, usability.

I. Introduction

THE amount of information contained in the World Wide Web has increased exponentially in the recent years. Internet search engines, used by millions of people on daily basis, represent the main tool for obtaining information in the World Wide Web. The methodology adopted by the majority of these search tools for the presentation of retrieved results is based on simple textual lists. Results are often displayed in a list textually and ranked according to their degree of relevance. Common attributes of results include their title, a brief summary and their web address.

Due to the exponential growth of documents stored over the Internet the process of searching and obtaining relevant documents has become more difficult and tedious. As result in recent years there have been several efforts aimed to provide more feedback to the users. However the communication of larger sets of information to users becomes particularly difficult in typical text-based interfaces, as users' visual channel becomes easily overexploited, thus causing the browsing experience of information to be often frustrating. This problem has been pointed out in the past years, and different attempts have been made in order to provide interfaces that could offer better visualisations methodologies of the retrieved data. However, this issue has not been addressed fully, as no new methodologies have shown to be particularly suitable in this context. In this context the idea of using multimodality to convey information of results was investigated. It is expected that with the use of additional metaphors, users are expected to perceive a larger amount of information for each retrieved document in an efficient and effective way.

A. Ciuffreda is with the University of Bradford, Bradford, UK (phone: +44(0)7910934133; e-mail: A.Ciuffreda@bradford.ac.uk).

D. Rigas is with the University of Bradford, Bradford, UK (phone: +44(0)1274235131; e-mail: D.Rigas@bradford.ac.uk).

II. Previous work

A large number of experimental works have been performed in the past few years to provide an alternative to traditional text-based interfaces for the browsing of results retrieved from query activities. These research studies were mostly based on the use of visualisation methodologies of 2D or 3D nature. Research based on 2D visual methodologies include Kartoo [1] and Ujiko [2], which used different interactive 2D maps to show relationships between results and topics, and Grokker [3], which adopted interactive round maps to display retrieved results in a hierarchical order according to a directory-based classification. Another application that explored 2D graphical objects was Insyder [4], [5], which made use of a multiple view methodology based on different 2D graphs, such as bar charts and scatter plots, to display retrieved results according to different levels of abstraction.

Research based on three dimensions includes Periscope [6], based on an AVE methodology [7], which used a series of 3D interface models of holistic, analytical and hybrid nature in order to represent results in different level of abstractions, and NIRVE [8], [9], which implemented a 3D space window to display the retrieved documents in groups of box-shaped clusters, which communicated the average of concepts in the documents included with the use of bar charts. Another application of this class is SmartWeb [10], which adopted the concept of virtual reality to represent retrieved results as buildings scattered in virtual 3D landscapes where users could move within and explore, and Tafiti [11], which used a rotating tree of 3D nature to accommodate textually the title of retrieved results.

In interfaces other than Internet search engines, a series of experimental studies have demonstrated the successful use of speech sound and non-speech sound, such as auditory icons and earcons, as means to communicate information. Auditory icons represent entities and actions within the interface with the use of environmental sound [12]. They have been implemented in different applications (e.g. SonicFinder [13]). Earcons are short musical messages that have attracted

 Manuscript Received Jan. 15, 2008; Revised received June 17, 2008

research attention in recent years and have been used successfully in graphical interfaces to communicate information [14]-[16]. They have demonstrated to be particularly useful in interfaces for visually impaired users [17]-[19]. The use of speech, often of synthetic nature, has been successfully adopted in a wide range of applications, such as screen readers [20], talking faces [21] and linguistic tools [22].

In [23] for the first time the combination of 2D graphs and speech and non-speech sound was used for the presentation of retrieved results from Internet queries. An experimental browsing platform, named AVBRO II, consisting of five interfaces based on different metaphors and sensory channels, was developed. Empirical studies based on the use of this platform demonstrated the higher level of efficiency, effectiveness and users' satisfaction of specific multimodal interfaces when compared to a traditional textual interface. The work presented in this paper represents a continuation of this experiment.

III. AVBRO II PLATFORM

A modified version of the initial AVBRO II browsing platform was considered for this new empirical investigation.

In addition to title, summary and URL, each result was represented by the times of typed keywords occurring in the document and its country of origin. The interface of this browsing tool was based on Adobe Flash® technology and developed using Adobe Actionscript® language only. The audio was based on recorded files of digitised speech and musical stimuli. For the processing of Internet queries, this platform made the Google API® web service. Based on this technology, the platform directly communicated with the Google® server. Keywords entered via the platform were sent to the Google® server, which performed a query with these words and returned the results back to the platform.

Three interfaces with a combination of different modalities were implemented in this new version of AVBRO II (see Table 1). One modality was based on text only. The remaining two modalities were based on an audio-visual approach to convey information. For example, spoken messages, icons, graphs and colours were used. The next sections describe these modalities in more detail.

A. Textual Interface

This type of interface was similar to other interfaces of typical search engines. It presented the whole set of retrieved results textually, as shown in Figure 1. The retrieved results were displayed in a ranked-based order from top to bottom in the central part of the interface.

B. Speech/Graph Interface

The interface (see Figure 2) presented each keyword using a specific colour. The first keyword was displayed in *red*

Interface	Display Method	Title, Summary, URL	Keywords Occurrences	Country of Origin
Textual	textual list	text	text	text
Speech/Graph	visual icons	text	2D graph	speech
Speech/Earcons	visual icons	text	earcons	speech

Table 1. Methodologies used for the representation of information in the three interfaces.

Fig. 1. An example of Textual interface.

characters, the second in *blue*, the third in *green*, the fourth in *orange*, the fifth in *purple* and the sixth in *cyan*. Different types of graphical objects and speech messages were used in addition to text for the communication of the retrieved results. Once the search process was completed by Google, the interface presented the 30 most relevant results as interactive circles displayed in an ordered tail sequence. Each circle was displayed in black colour with a grey edge and had a specific size, according to its rank. When users placed the mouse pointer over a circle, the circle expanded to display the ranking number of the corresponding document. Above the group of icons, a text panel displayed the title, summary and URL of the document selected. On the right side of the interface one of two graphs of 2D nature were displayed.

These graphs, named *CellGraph* and *HexaGraph*, were used to communicate visually the occurrences of used query words.

The *CellGraph* was based on a set of columns horizontally aligned which was made up of cells. The number of columns corresponded to the number of query words used. The colour of the cells corresponded to the keyword typed in the text field of the search query. The number of cells for each column corresponded to the occurrences of the represented keyword. Each cell represented ten occurrences. For example, the first column would have three red cells aligned vertically if the

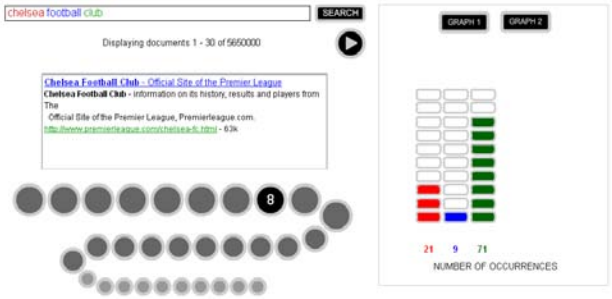

Fig. 2. An example of Speech/Graph interface displaying the CellGraph graph.

Fig. 3. An example of Speech/Earcons.

first keyword entered by the user had occurred twenty-seven times. If there were more than one hundred occurrences of a keyword only ten cells would have been displayed.

The *HexaGraph* had a hexagon shape. The graph was divided in six different areas of equal size, similar to triangles. Each of these areas corresponded to a typed keyword. The occurrences of a keyword were represented as lines in the specific area. Each line represented ten occurrences using the colour of the corresponding keyword. The graph visualization could be changed by the user by clicking the related button placed above the displayed graph. Below the displayed graph the occurrences were also displayed textually. Each number was displayed with the colour of the keyword that represented. A speech message, spoken by a female voice, communicated the country of origin of the selected document. The action of pressing the button of the mouse over an icon would cause the represented document to open in a new page.

C. Speech/Earcons Interface

Most of the metaphors adopted in Speech/Graph interface were re-used in this new interface. The single aspect that differentiated this interface from the Speech/Graph interface (see Figure 3) was the method used to present keywords occurrences. The 2D graphs used in the Speech/Graph interface were replaced by a set of musical notes played sequentially in a rising-pitch order after the communication of the country of origin of the result with spoken messages. Each keyword (with its specific colour) was associated to a specific instrument. The musical instruments used were piano, guitar, saxophone, drum, the organ and bass. These instruments were chosen based on guidelines from [24]. Each note played from a musical instrument represented ten occurrences. For example, the occurrences of thirty-five times of the query word used in a webpage would have resulted in four rising pitch notes played sequentially by the guitar. Only ten notes were played if the number of occurrences exceeded one hundred.

IV. Experiment

A. Usability Criteria

The following criteria were chosen to measure the level of usability of the three interfaces: *learnability*, *errors rate*, *efficiency*, *memorability* and *users' satisfaction* [25]. The learnability of the interfaces was measured using the number of questions made by users regarding the nature of the interfaces and their metaphors after being explained to them, their opinions and the time spent to perform an initial search operation. The error rate was measured using the number of pages visited from a set of search operations. The level of efficiency was determined with the measurement of the browsing time spent to perform these tasks. The level of memorability of the interfaces was measured through the users' successful recognition of the information communicated by the metaphors used in the interfaces, after a specific period of time from their participation in the experiment. Finally users' satisfaction about the interfaces was evaluated by obtaining users' views of each individual metaphor and interface after being used.

B. Experimental Methodology

A total of 24 users participated in this empirical experiment. The sample consisted of 17 male and 7 female users. 19 had an engineering or informatics background. Typically, the age range of the sample was between 18 and 35 years old. The experiment was subdivided into three phases, one phase for each interface. In each phase, users were first introduced to the specific interface and metaphors used. Users were free to ask questions regarding any aspect of the interface which was not clear. On completion of this presentation, users had to perform a *training task* in order to become familiar with this interface. On completion of the training session, users were required to perform two search operations (*usability tasks*) with the same interface. The number of query words used in these tasks was two, three or four.

During the browsing process, users were instructed to try to use all the information provided in the results page. The number of pages accessed and the time spent for the browsing time required to complete the task was recorded. After the performance of the two tasks with an interface, users were asked to enter the information used in the two browsing processes, rate the level of utility of the same information and make comments about the suitability of the medium used to

communicate this information. A rotation scheme was used in order to guarantee an equal number of executions of a task in the three interfaces. In the final stage of the experiment, users were asked to rate the interfaces. Parameters used for these ratings were the level of learnability, pleasance and frustration felt when using these interfaces. Users were asked to select their preferred interface, and provide an explanation for their choice. These ratings were used to analyse the level of users' satisfaction of the interfaces used. Finally, four weeks after their participation in the experiments, users were asked to correctly identify the message conveyed by the metaphors used in the two experimental interfaces (Speech/Graph and Speech Earcons). These metaphors were the icons for representing retrieved results, the keywords occurrences graph, the earcons and the speech sound. As previously stated, the results from these questions were used to define the level of memorability of these two interfaces.

V. Results and Analysis

A. Learnability

Only a single question was formulated about the nature of the Textual interface; five and six questions meanwhile were formulated for the Speech/Graph interface and Speech/Earcons interface respectively.

Less time was spent with the Textual interface than with the Speech/Graph and the Speech/Earcons interfaces for the completion of the training tasks, as Table 2 shows. A total of 452 seconds were spent with the Textual interface; the amount of seconds spent with the Speech/Graph interface and with the Speech/Earcons interface was 481 and 582 respectively. A one-way ANOVA computation showed that this difference of time spent among the interfaces was not statistically significant (see Table 3).

According to users' ratings, all three interfaces were easy to learn. Considering the first two points of the Likert scale, 22 users (91.6%) agreed or strongly agreed to the Textual interface being easy to learn, meanwhile the level of learnability of Speech/Graph and Speech/Earcons interfaces was agreed or strongly agreed by 23 (95.8%) and 20 (83.3%) users respectively.

B. Error Rate

Users who made use of the two multimodal interfaces accessed a substantial inferior number of pages for the achievement of the usability tasks, in comparison with the traditional Textual interface. A total of 81 and 78 pages were accessed respectively when using the Speech/Graph and Speech/Earcons interface; meanwhile a total of 116 pages were visited with the Textual interface (see Table 4). A one-way ANOVA calculation (see Table 5) demonstrated that the difference of pages accessed among the three interfaces was statistically significant at .05 level, with the obtained F value (6.515) being higher than the critical value considered (3.06).

	Interface		
	Textual	**Speech /Graph**	**Speech/Earcons**
Σ	452	481	582
X	18.833	20.041	24.25
σ	20.4	16.8	21.7

Table 2. Total number of seconds spent (Σ) by users for the performance of training tasks, with relative mean (X) and standard deviation (σ).

	SS	**df**	**MS**	**F**	**P**
Between Groups	388.083	2	194.042	.497	.611
Within Groups	26940.7	69	390.446		
Total	27328.8	71			

Table 3. Outcome of the ANOVA computation for the comparison of means of seconds spent in the three interfaces for the performance of training tasks. F value is displayed in bold if statistically significant, being higher than 3.13, the critical value at .05 level.

	Interface		
	Textual	**Speech /Graph**	**Speech/Earcons**
Σ	116	81	78
X	2.416	1.687	1.625
σ	1.58	0.92	0.95

Table 4. Total number of pages visited (Σ) by users for the performance of usability tasks, with relative mean (X) and standard deviation (σ).

	SS	**df**	**MS**	**F**	**P**
Between Groups	18.597	2	9.299	**6.515**	.002
Within Groups	201.229	141	1.427		
Total	219.826	143			

Table 5. Outcome of the ANOVA computation for the comparison of means of pages visited in the three interfaces for the performance of usability tasks. F value is displayed in bold if statistically significant, being higher than 3.06, the critical value at .05 level.

C. Efficiency

As with pages visited, less time was spent by users when browsing retrieved results during the performance of usability tasks with the multimodal interfaces. A total of 1548 seconds and 1941 seconds were spent respectively when using the Speech/Graph and Speech/Earcons interface; meanwhile a total of 2153 seconds were spent with the Textual interface (see Table 6). A one-way ANOVA calculation, whose results are shown in Table 7, demonstrated however that the difference of time among the three interfaces, was not statistically significant at .05 level, with the obtained F value (2.907) being slightly lower than the critical value considered (3.06).

D. Memorability

Almost the entire majority of users were able to correctly recognise the information communicated by the metaphors of visual and aural nature used in the multimodal interfaces.

	Interface		
	Textual	**Speech /Graph**	**Speech/Earcons**
Σ	2153	1548	1941
X	44.85	32.25	40.43
σ	32.9	18.9	24.2

Table 6. Total number of seconds spent (Σ) by users for the performance of usability tasks, with relative mean (X) and standard deviation (σ).

	SS	**df**	**MS**	**F**	**P**
Between Groups	3926.514	2	1963.257	2.907	.058
Within Groups	95216.792	141	675.296		
Total	99143.306	143			

Table 7. Outcome of the ANOVA computation for the comparison of means of seconds spent in the three interfaces for the performance of usability tasks. F value is displayed in bold if statistically significant, being higher than 3.06, the critical value at .05 level.

Over the total number of users sample (n=24), the information communicated by the visual icons was correctly identified by 23 users and partially identified by the remaining user. The function of the 2D graph was correctly identified by 22 users and identified partially by the remaining two. A total of 21 correct answers and one partially correct answer were found accurate for the recognition of the information conveyed by the musical stimuli played in the Speech/Earcons interface. Finally, the message communicated by the speech played in the two multimodal interfaces was correctly identified by 21 users, while two answers were partially correct only.

E. Users' Views of the Interfaces and Metaphors Used

Figure 4 shows the results of users' ratings about the three interfaces, together with their choice of preferred interface. As the graph shows, users found the Textual and Speech/Graph interface very pleasant to work with. Nineteen users (79.1%) users agreed or strongly agreed that the Textual interface was pleasant to use, while 20 of them (83.3%) agreed or strongly agreed with the same statement when considering the Speech/Graph interface. The Speech/Earcons interface was judged positively by users, although to a somewhat lesser extent, as only 15 of them (62.5%) believed or strongly believed that using this interface was a pleasant experience.

No interface was found frustrating to use, based on users' views. Eighteen users (75%) agreed or strongly agreed that the Textual interface was not frustrating to use. Nineteen (79.1%) and 17 (70.8%) of them agreed or strongly agreed that using the Speech/Graph and Speech/Earcons interface respectively was not a frustrating experience.

The most preferred interface, among the 24 users who took part at the experiment, was the Speech/Graph interface, as 12 users (50%) of the sample group preferred this interface. Within the rest, seven users (29.1%) selected the Textual interface as their most preferred interface, while the remaining five (20.9%) opted for the Speech/Earcons interface.

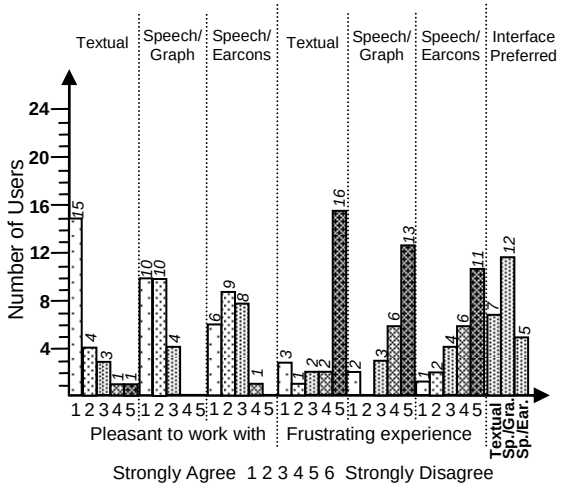

Fig. 4. Frequency of users' responses to the statement "the interface was pleasant to work with" and "using this interface was a frustrating experience" and their choice of preferred interface.

Experimental observations suggest users' positive attitude towards the keywords occurrences graph may have played an influent role in these ratings. Users' tendency to judge positively the Textual interface may be due to their habits of using Internet search engines with similar interfaces.

VI. CONCLUSION

This paper described an empirical investigation based on Internet search queries with the participation of 24 users. The aim of the investigation was to measure and compare the level of usability of a traditional text-based interface, named *Textual*, and two experimental multimodal interfaces, named *Speech/Graph* and *Speech/Earcons* respectively.

The outcome of the experiment demonstrated the higher level of usability of the multimodal interfaces in comparison with the traditional text-based interface for browsing retrieved results from the Internet. It was believed that the level of learnability and memorability of the Speech/Graph and Speech/Earcons interfaces could not be higher than the traditional Textual interface, due to users' habits to use this latter type of interface. The results, however, did not demonstrate a significant difference in terms of these two criteria. According to these results, users seemed to have learned easily to use the two multimodal interfaces and memorise their metaphors used on a long term basis.

None of the used interfaces was found unpleasant or frustrating to use, according to the results obtained from the users' questionnaire.

No statistical difference in efficiency was recorded among the three interfaces, although a noticeable less time was spent for the accomplishment of the requested tasks with the Speech/Graph interface, which resulted to be also the most preferred interface by the large majority of users.

Results however showed a significant difference of error rate between the two multimodal interfaces and the Textual

interface. The number of pages visited in the Speech/Graph and Speech/Earcons was clearly lower in comparison with the Textual interface.

The experiment also offered a strong encouragement in continuing the research towards the multimodality-based direction in the design of new Internet search engine interfaces.

REFERENCES

[1] Kartoo, *http://www.kartoo.com*.
[2] Ujiko, *http://www.ujiko.com*.
[3] Grokker, *http://www.grokker.com/grokker.html*.
[4] H. Reiterer, G. Mußler, T. M. Mann, and S. Handschuh, "INSYDER—an information assistant for business intelligence," in *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2000, pp. 112-119.
[5] H. Reiterer, G. Tullius, and T. M. Mann, "INSYDER: a content-based visual-information-seeking system for the Web," *International Journal on Digital Libraries*, Vol. 5, 2005, pp. 25-41.
[6] W. Wiza, W. Walczak, and W. Cellary, "Periscope: a system for adaptive 3D visualization of search results," in *Proceedings of the Ninth International Conference on 3D Web Technology***,** 2004, pp. 29-40.
[7] W. Wiza, W. Walczak, and W. Cellary, "AVE-method for 3D visualization of search results**,"** *Lecture Notes in Computer Science*, Vol. 2722, 2004, pp. 219-246.
[8] J. Cugini, S. Laskowski, and C. Piatko, "Document clustering in concept space: the NIST Information Retrieval Visualization Engine (NIRVE)," National Institute Standards and Technology (NIST), June 2006, http://zing.ncsl.nist.gov/cugini/uicd/ccpaper.html.
[9] J. Cugini, S. Laskowski, and M. Sebrechts, "Design of 3D visualization of search results: evolution and evaluation," in *Proceedings of IST/SPIE's 12th Annual International Symposium: Electronic Imaging*, 2000, pp. 23-28.
[10] N. Bonnel, A. Cotarmanac'h, and A. Morin, "Meaning metaphor for visualizing search results," in *Proceedings of the Ninth International Conference on Information Visualisation*, 2005, pp. 467–472.
[11] Tafiti, *http://www.tafiti.com*.
[12] W. W. Gaver, "Synthesizing auditory icons," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1993, pp. 228-235
[13] W. W. Gaver, "The SonicFinder: an interface that uses auditory icons," *Human Computer Interaction*, Vol. 4, 1989, pp. 67-94.
[14] S. A. Brewster, P. C. Wright, and A. D. N. Edwards, "An evaluation of earcons for use in auditory human-computer interfaces," in *Proceedings of InterCHI'93*, 1993, pp. 222-227.
[15] S. A. Brewster, "Providing a structured method for integrating non-speech audio into human-computer interfaces", PhD Thesis, University of York, UK, 1994.
[16] D. Rigas, "Guidelines for auditory interface design: an empirical investigation", PhD Thesis, University of Loughborough, UK, 1996.
[17] J. L. Alty, and D. Rigas, "Exploring the use of structured musical stimuli to communicate simple diagrams: the role of context," *International Journal of Human-Computer Studies*, Vol. 62, 2005, pp. 21-40.
[18] D. Rigas and J. L. Alty, "The use of music in a graphical interface for the visually impaired," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1997, pp. 228-235
[19] D. Rigas and J. L. Alty, "The rising pitch metaphor: an empirical study," *International Journal of Human Computer Studies*. Vol. 62, 2005, pp. 1-20.
[20] C. Asakawa and T. Itoh, "User interface of a home page reader," in *Proceedings of the Third International ACM Conference on Assistive Technologies*, 1998, pp. 149-156.
[21] J. Ostermann and A. Weissenfeld, "Talking faces-technologies and applications," in *Proceedings of the 17th International Conference on Pattern Recognition*, Vol. 3, 2004, pp. 826-833.
[22] M.J. Hamel "FreeText: a "smart" multimedia web-based computer assisted language learning environment for learners of French," in *Proceedings of the Second International Conference on Multimedia and Information & Communication Technologies in Education*, 2003, pp. 1661-1665.
[23] D. Rigas and A. Ciuffreda, "An empirical investigation of multimodal interfaces for browsing internet search results" in *Proceedings of the 7th Conference on 7th WSEAS International Conference on Applied Informatics and Communications*, 2007, pp. 194-199.
[24] D. Rigas and J.L. Alty, "How can multimedia designers utilise timbre?," in *Proceedings of BCS-HCI 1998*, 1998, pp. 273-286.
[25] J. Nielsen, *Usability Engineering*, Morgan Kaufmann Publishers, 1993.