

An Empirical Investigation of Multimodal Interfaces for Browsing Internet Search Results

DIMITRIOS RIGAS, ANTONIO CIUFFREDA

Department of Computing
University of Bradford
Bradford
UNITED KINGDOM

Abstract: - An empirical investigation, comparing the level of usability between multimodal and non-multimodal interfaces within Internet search engines is presented. The experiment was based on Internet search activities through the use of five different interfaces provided by AVBro2, an experimental platform developed for this investigation. A total of fifty users participated in this experiment. Each user performed three tasks with every interface. The evaluation of the obtained results according to different usability criteria showed a higher degree of usability in multimodal interfaces when compared with a traditional text-based interface for browsing Internet search results. In addition a detailed analysis of the level of functionality of visual and audio metaphors implemented in these interfaces is discussed.

Key-Words: - Multimodality, Internet browsing, visual metaphors, audio metaphors, earcons.

1 Introduction

Internet search engines are an important and essential tool for exploring resources on the World Wide Web. Recently, the growth of the Internet has caused the amount of online resources to increase exponentially, creating more difficulty for users to browse and select appropriate documents. This has led to an increasing demand of feedback regarding retrieved results. It can become difficult for users when all information is conveyed via text, most common in search engine interfaces. As users read lines of text and scroll the results page continuously, their mental workload may increase easily.

2 Using Audio and Visual Metaphors

The presentation of information in computer interfaces has been subject to a series of studies in recent years. Different alternative interfaces have been presented as solutions to improve presentation of the documents and the browsing process as consequence. These interfaces are based on different visual metaphors of 2D or 3D nature. In addition the use of speech and non-speech sound in previous experiments has proved to be effective in conveying different types of information in different contexts.

Regarding interfaces of 3D the use of 3D models of holistic, analytical and hybrid nature for representing results was adopted by Periscope [1], with the use of AVE methodology [2]. Clusters of documents on a 3D space window were used by

NIRVE [3]. The concept of virtual reality for representing the set of retrieved results in different 3D landscapes has been implemented in ViOS [4] and SmartWeb [5]. These interfaces through the use of 3D metaphors can improve significantly the capability of representing and organize simultaneously sets of documents. However, most of these metaphors are not user friendly or inappropriate for the considered tasks. As a result, these interfaces may be very difficult to understand and use by the majority of people.

Different search engines have implemented 2D visual metaphors in their results interfaces. Interactive maps of various natures have been implemented by Kartoo [6], Ujiko [7] and Grokker [8]. The use of synchronized multiple views of 2D graphs for documents displayed have been adopted by VisMeB [9] and Insyder [10]. Although the visual metaphors used in the visualization methodologies of these search applications are more intuitive in comparison with the 3D metaphors previously discussed, limitations exist in the amount of information to deliver in these interfaces.

Different studies have proved the successful implementation of sound of non-speech nature as a mean to communicate information in system interfaces. Auditory icons [11] originally implemented in the SonicFinder application [12] have been used in a wide range of application interfaces. Earcons have been implemented in graphical interfaces for communicating information

Interface	Display Methodology	Title, Summary, URL	Keywords Occurrences	Country of Origin	Suggested Keywords	Related Pages
Textual	List	Text	Text	Text	Text	Text
Speech/Graph	Visual Icons	Text	2D Graph	Speech	Text	Visual Icons
Speech/Earcons	Visual Icons	Text	Earcons	Speech	Text	Visual Icons
Earcons	Visual Icons	Text	Earcons	Visual Icon	Text	Visual Icons
Graphical	Visual Icons	Text	2D Graph	Visual Icon	Text	Visual Icons

Table 1. Methodologies used for information representation in the five interfaces

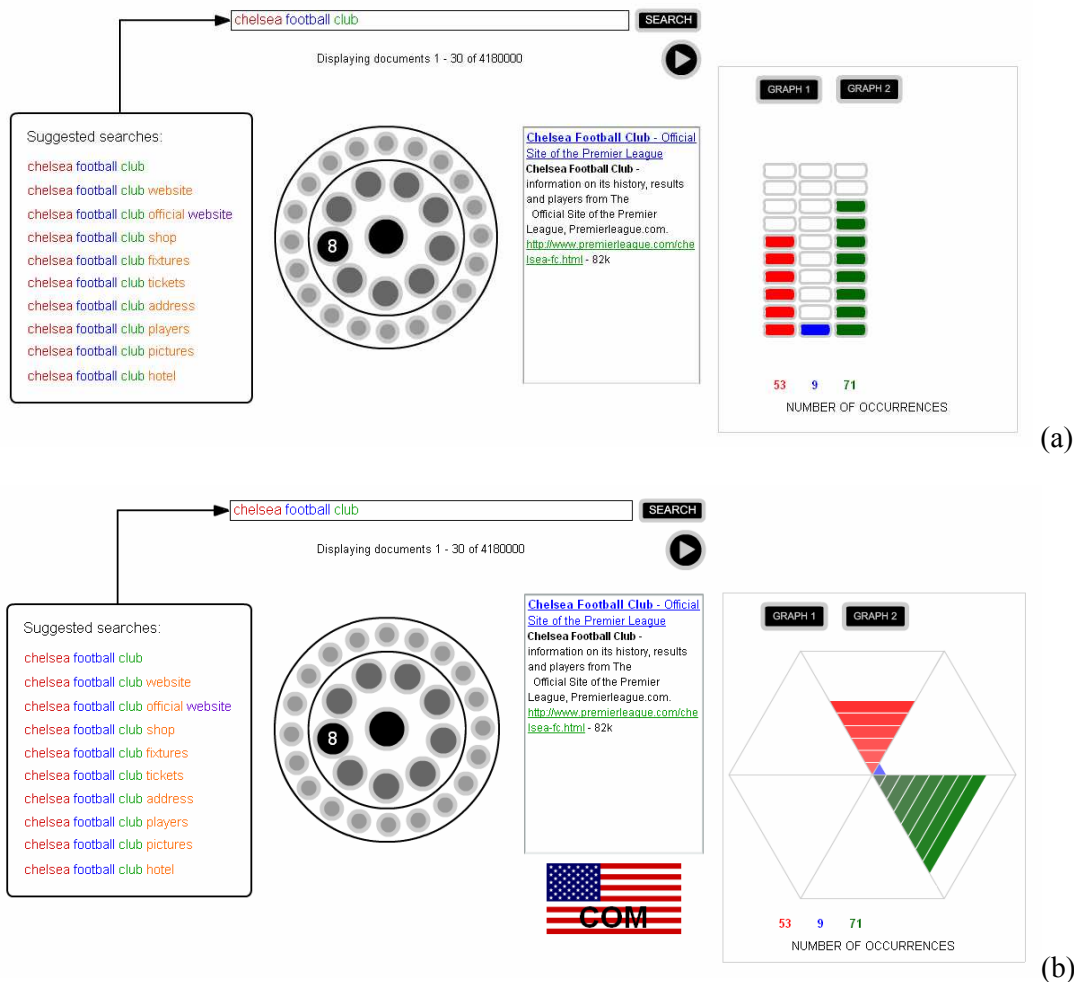


Figure 1. An example of Speech/Graph Interface displaying the CellGraph graph (a), and Graphical Interface displaying the HexaGraph graph (b).

[13-15]. Additionally, this type of musical stimuli has demonstrated to be useful in interfaces for visually impaired users [16-19].

In recent years, the use of Text-To-Speech technology [20] has been widely opted for a wide range of applications, due to technical improvements made into the naturalness and fluency of the sound. Recorded speech can represent a valid tool for delivering information, thanks to the level of naturalness and intelligibility that can be obtained, although the manual recording needed for each single instruction limits considerably its use.

3. AVBro2: An Experimental Platform

In this context the use of multimodality for transmitting information in Internet search engines was presented, after an accurate investigation regarding usability problems related to traditional text-based results and possible alternatives. The implementation of this methodology aimed at increasing the amount of information of retrieved results that could be conveyed, thus preventing the rise of users' mental workloads. To empirically compare the level of usability of interfaces of different nature, AVBro2, a new Internet browsing

platform based on five interfaces of different modality was developed.

AVBRO2 (Audio Visual BROWser), named after the previously developed AVBRO platform [21], made use of Google API [22] technology to perform Internet queries through Google and retrieve related results. This platform provided five different interface modalities for communicating a defined set of information of retrieved results. In addition to the title, summary and address information, for every result the number of occurrences of used keywords within the document, its country of origin, a set of suggested keywords to improve the search and a set of pages related to the specific document were included. Three of the five interfaces were based on a multimodal approach to communicate information of results. They made use of text, visual metaphors such as graphical objects and colours, and audio metaphors of non speech or speech nature. These interfaces were named Textual, Speech/Graph, Speech/Earcons, Earcons and Graphical.

The Textual interface, similar to interfaces from popular search engines, presented information of results in a textual approach. Retrieved results were displayed textually in a list in the central part of the interface according to their degree of relevance. The colour of the entire set of information displayed was black.

In the Speech/Graph interface the first six keywords typed were displayed in red, blue, green, orange, purple and cyan respectively. Once the searching process was completed results were presented as black interactive circles in the central part of the interface. The size and location of these circles changed according to their degree of relevance of the represented document. Primary information of a document (title, summary and URL) were displayed textually in a text area on the right of these circles. The country of origin of a selected document was communicated through a female voice. The keywords occurrences of used keywords within a specific document were displayed on the right side of the interface with one of two 2D graphs available. The 2D graphs represented with different colour lines the occurrences of keywords used within the document selected. The graphs visualization could be switched by the user by clicking the related button placed above the displayed graph.

Related pages were displayed as interactive circles below the cluster of circles representing retrieved results. As in the previous interface, on the left side

of the interface a panel including the list of groups of suggested keywords was provided. Words included in a group had distinct colour from the rest.

The Speech/Earcons interface shared a similar approach for conveying information with the Speech/Graph interface. The only aspect differentiating this interface from the previous was the way keywords occurrences were presented. Instead of 2D graphs, a set of musical notes was played sequentially in a rising-pitch order after the country of origin was conveyed through voice.

In the Earcons interface the country of origin of retrieved results was displayed visually with a flag of the specific country. The interface also made use of earcons to provide occurrences of used keywords. Primary information, related pages and suggested query words were communicated visually as in the previous two interfaces.

The Graphical interface was based on visual metaphors only. The audio channel was not used. In this interface the country of origin was communicated visually through a flag. Other data, such as primary information, related pages and suggested keywords were transmitted as in the previous three interfaces with visual metaphors and text.

4. Experiment

In order to test the usability level of the five interfaces three important criteria were considered: effectiveness, efficiency and users' satisfaction.

The level of effectiveness was measured through the success rate of search operations of tasks. Regarding the level of efficiency of the interfaces, three measures were adopted: number of pages visited, seconds spent and the number of mouse clicks. Users' satisfaction was measured through questions made to users in order to obtain feedback about the interfaces and individual components used in these interfaces.

Fifty users participated in the programmed experiment, together with an investigator. The majority of participants were students from the University of Bradford coming from a scientific background, following courses in the areas of Engineering and Informatics. Most of the subjects used Internet ten or more hours a week. Thirty-six of those were male, while the remaining fourteen were female. The age of the majority of users ranged from 18 to 35 years.

Measure	Interface	Textual	Speech/Graph	Speech/Earcons	Earcons	Graphical
Pages Visited	Textual		+2.14	+2.45	+2.09	+2.28
	Speech/Graph	-2.14		+0.46	+0.05	+0.23
	Speech/Earcons	-2.45	-0.46		-0.39	-0.22
	Earcons	-2.09	-0.05	+0.39		+0.17
	Graphical	-2.28	-0.23	+0.22	-0.17	
Seconds Spent	Textual		+0.99	+2.67	+1.72	+2.11
	Speech/Graph	-0.99		+1.85	+0.83	+1.29
	Speech/Earcons	-2.67	-1.85		-0.92	-0.38
	Earcons	-1.72	-0.83	+0.92		+0.47
	Graphical	-2.11	-1.29	+0.38	-0.47	
Mouse Clicks	Textual		+6.09	+6.13	+5.9	+6.22
	Speech/Graph	-6.09		+0.36	+0.05	+0.42
	Speech/Earcons	-6.13	-0.36		-0.3	+0.04
	Earcons	-5.9	-0.05	+0.3		+0.35
	Graphical	-6.22	-0.42	-0.04	-0.35	

Table 2. Values obtained through t-test calculation when comparing the number of pages visited, seconds spent and mouse clicks performed with the five interfaces used. Values are displayed in bold if statistically significant.

Each participant was asked to perform a total of fifteen tasks. Each interface had to be used for the completion of three tasks only. In order to guarantee the performance of each task by an equal number of interfaces throughout the process, the interfaces were assigned through a rotation scheme.

Tasks were based on finding specific information by using a given set of specific keywords. Two, four and six keywords were given to three tasks respectively. The tasks were evenly distributed in three groups according to their difficulty degree of completion. The tasks were based on findings of information belonging to a wide range of topics which did not appeal to the majority of people. As result it was guaranteed certain impartiality among users when trying to answer the given questions.

When browsing retrieved results, users were advised to try to make use of all information provided in the AVBro2 results interfaces. Simultaneously, the investigator was required to observe users' behaviour and provide data for each performed task in order to analyse the level of effectiveness and efficiency of the used interface. The data provided by the investigator was the effective completion of the assigned task, the number of pages accessed, the time spent and the number of mouse clicks performed to complete a task. After making use of an interface participants were required to rate its level of difficulty and learnability, together with the usefulness of the information used provided with different modalities.

5. Results And Analysis

The level of effectiveness of the five interfaces used was high. The total number of participants was able to complete the fifteen tasks assigned. Users were

not given help during their task performances and were free of any decision and action during their browsing activities.

According to the obtained data, the number of pages visited when using experimental interfaces was notably lower in comparison with the Textual interface. The lowest number of pages accessed was obtained by the Speech/Earcons interface. A total of 277 pages were visited over 150 tasks performed with this interface based on text, graphs, speech and non-speech sound. Similar results were gained with the Graphical, Earcons and Speech/Graph interface with 282, 286 and 287 pages accessed respectively. A higher number of pages visited instead, 343, was obtained by the Textual interface.

As with pages visited, less time was spent for the performance of the assigned tasks by using the four experimental interfaces. A total of 13722 seconds was spent to perform 150 tasks with the Speech/Earcons interface, meanwhile 14260, 14983 and 16218 seconds were needed to perform the same tasks with the Graphical, Earcons and Speech/Graph interfaces respectively. Users using the Textual interface spent more time to perform these tasks, as the total amount of time required was 17840 seconds.

A large number of mouse clicks was performed via the use of the Textual interface compared with the four experimental interfaces. The total number of mouse clicks performed with this interface for the completion of the 150 tasks was 488. The numbers of mouse clicks recorded for the Speech/Graph, Speech/Earcons, Earcons and Graphical interfaces were similar to each other, being 291, 283, 290 and 282 respectively.

A series of t-test calculations were performed between values of pages visited, seconds spent and

mouse clicks of the five interfaces used. The t values generated by comparing the pages accessed between the Textual interfaces and Speech/Graph, Speech/Earcons, Earcons and Graphical interfaces were 2.14, 2.45, 2.09 and 2.28 respectively. These numbers were higher than 1.96, the critical value of t with a degree of freedom of 298 at .05 level of significance. Values obtained in t-test calculations among the four experimental interfaces were noticeably lower than the previously considered critical value of t , showing no difference in number of pages accessed from a statistical perspective. As with pages visited, t-test calculations showed a statistical significance between the time spent for performing tasks through the Textual interface and the Speech/Earcons interface and between the Textual interface and the Graphical interface. The t values obtained by these calculations, 2.67 and 2.11 respectively were higher than 1.96, the considered critical value of t . A high t value of 1.72 was also obtained by comparing the time spent between the Textual interface and the Earcons interface. However, this was lower than 1.96 and not statistically relevant. Other high t values obtained but not relevant from a statistic perspective were recorded when comparing the difference of time taken between the Speech/Graph interface and Speech/Earcons interface and between the Speech/Graph interface and the Graphical interface.

T-test calculations showed the differences of mouse clicks performed between the Textual interface and other interfaces used were significant at a statistical level. The t value obtained by these calculations, 6.09, 6.13, 5.9 and 6.22 were larger than their correspondent critical value, 1.96. No significant differences from a statistical perspective were recorded through t-test calculations when comparing the number of mouse clicks among experimental interfaces.

The level of users' satisfaction was positive in all five interfaces with no significant difference. With a rating scale of 1 to 6 (1 most easy and 6 most difficult) the easiest interface to use was the Graphical interface, with an average of 1.96. Earcons, Textual and Speech/Graph interfaces followed with ratings average of 1.96, 2.4 and 2.44 respectively. The Speech/Earcons interface averaged 2.78, proving to be the most difficult interface to use according to users' opinions.

Confidential results were obtained by users' questionnaires regarding the level of learnability. The difference of learnability among the five interfaces was very small. According to users the Graphical interface was the easiest interface to learn, average score of 1.82. Textual, Earcons and

Speech/Graph interfaces were ranked lower with an average score of 1.9, 2.06 and 2.08 respectively. The Speech/Earcons interface was the most difficult interface to learn, with an average score of 2.4.

The level of users' satisfaction towards the different information offered in the five interfaces varied. Users preferred the traditional list-based approach for displaying retrieved results. According to a rating scale ranging from 1 to 6 (1 most useful and 6 least useful) the average rate of text for documents representation was 4.78. Meanwhile, the average rate of the use of interactive icons was slightly lower, being 4.61.

The occurrences of used keywords were judged positively by participants. Keywords occurrences information was found more useful if delivered through 2D graphs. The average score of this metaphor for this purpose was 4.77 over 6. Text and Earcons were also positively rated, by gaining a score of 4.36 and 3.43 respectively.

The country of origin of documents received good ratings from users. Subjects found this information more useful if given through speech, rather than delivered through text or visual icons. The average rating of speech sound when used for this information was 3.77. Meanwhile, the rate of usage of visual icons and text for this functionality received an average score of 3.58 and 2.94 each.

Information of related pages obtained a score of 2.26 when communicated through interactive visual icons and 2.02 when conveyed through text. Suggested keywords, provided textually in all five interfaces, achieved a score of 2.45.

The use of colours in the four multimodal interfaces was positively judged by users. The average score for this feature was 4.52.

6. Conclusion

An empirical experiment based on Internet search activities through five interfaces of different nature was conducted involving fifty users. For the results retrieved the interfaces provided a large set of information. This was to aid users in the browsing process of these results. Information was conveyed from the five interfaces in different modalities. One interface was purely based on text, similar to other common search engines. Moreover, the remaining four interfaces made use of different media, such as graphs, icons, speech and non-speech sound to deliver same information to users. Results obtained from this experiment showed interfaces based on a multimodal approach improved the quality of Internet search activities, as their level of usability was higher than standard text-based interfaces. The

experiment also revealed keywords occurrences information was a very helpful feature for the majority of users, especially if delivered through 2D graphs. Speech and non-speech sound also proved to have a good degree of popularity amongst users in the process of information delivery. Other information such as suggested keywords and related pages proved to be unpopular when delivered through different modalities. Overall, the results of the experiment were positive and showed the positive role of multimodal features in the browsing process. The experiment also offered strong encouragement in continuing the research towards a multimodality direction in the design of new search engine interfaces.

References:

- [1] Wiza, W., Walczak, W., Cellary, W., Periscope: A System for Adaptive 3D Visualization of Search Results, In *Proceedings of the Ninth International Conference on 3D Web Technology*. ACM Press, 2004.
- [2] Wiza, W., Walczak, W., Cellary, W., AVE-Method for 3D Visualization Of Search Results, In *Lecture Notes in Computer Science*. Springer-Verlag, 2003.
- [3] Cugini, J., Laskowski, S., Sebrechts, M., Design of 3D Visualization of Search Results: Evolution and Evaluation", In *Proceedings of IST/SPIE*, 2000.
- [4] ViOS, <http://computer.howstuffworks.com/vios.html>.
- [5] Bonnel, N., Cotarmanac'h, Morin, A., Meaning Metaphor for Visualizing Search Results, In *Proceedings of International Conference on Information Visualisation*, IEEE Computer Society, 2005.
- [6] Kartoo, <http://www.kartoo.com>.
- [7] Ujiko, <http://www.ujiko.com>.
- [8] Grokker, <http://www.grokker.com/grokker.html>.
- [9] Klein, P., Reiterer, H., Muller, F., Limbach, T., Metadata Visualization with VisMeb, In *Proceedings of the International Conference on Information Visualization*, IEEE Computer Society, 2003.
- [10] Reiterer, H., Tullius, G., Mann, T., INSYDER: a content-based visual-information-seeking system for the Web, in *International Journal on Digital Libraries*, Berlin-Heidelberg, p. 25-41, 2005.
- [11] Gaver, W. W., Auditory Icons: Using Sound in Computer Interfaces. *Human Computer Interaction*. 2, 1986, 167-177.
- [12] Gaver, W. W., Synthesizing Auditory Icons. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM Press, 1993.
- [13] Brewster, S. A., Wright, P. C., Edwards, A. D. N., An Evaluation Of Earcons For Use In Auditory Human-Computer Interfaces, In *Proceedings of InterCHI'93*. ACM Press, 1993.
- [14] Brewster, S. A., *Providing a Structured Method for Integrating Non-Speech Audio into Human-Computer Interfaces*, PhD Thesis, University of York, UK, 1994.
- [15] Rigas, D., *Guidelines for Auditory Interface Design: An Empirical Investigation*, PhD Thesis, University of Loughborough, UK, 1996.
- [16] Edwards, A.D.N., *Adapting User Interfaces for Visually Disabled Users*, PhD Thesis, the Open University, UK, 1987.
- [17] Rigas, D., Alty, J. L., the Use of Music in a Graphical Interface for the Visually Impaired. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM Press, 1997.
- [18] Rigas, D., Alty, J. L., the Rising Pitch Metaphor: An Empirical Study, In *International Journal of Human Computer Studies*. Elsevier, 2005.
- [19] Alty, J. L., Rigas, D., Exploring the Use of Structured Musical Stimuli to Communicate Simple Diagrams: The Role of Context, In *International Journal of Human-Computer Studies*, Elsevier, 2005.
- [20] Duggan, B., Deegan, M., Consideration In The Usage Of Text To Speech (TTS) In The creation Of Natural Sounding Voice Enabled Web Systems, In *Proceedings of the 1st International Symposium on Information and Communication Technologies ISICT '03*, Trinity College Dublin, 2003.
- [21] Rigas, D., Ciuffreda, A., Multi-Modal Web-Browsing - An Empirical Approach to Improve the Browsing Process of Internet Retrieved Results, In *Proceedings of the International Conference on Signal Processing and Multimedia Applications*, 2006.
- [22] Google Web API, <http://www.google.com/apis>.