

An Experiment to Evaluate the Effectiveness of Cross-Media Cues in Computer Media

Nancy Green

Department of Mathematical Sciences
383 Bryan Building
University of North Carolina Greensboro
Greensboro, NC 27402
nl.green@uncg.edu

Abstract

We present the motivation for and design of an experiment to evaluate the usefulness of cross-media cues, phrases such as '**See Figure 1**'.

1 Introduction

Authors of English-language print documents containing both text and graphics traditionally have used phrases such as '**See Figure 1**'. Intuitively, these *cross-media cues* (CMCs) help the print reader to integrate information presented in different media, i.e., printed text and printed graphics. We are investigating how, if at all, these cues should be used in presentations delivered in computer media such as web pages. Our long-term goal is to develop a non-application-specific computational model for the decision of *when* to direct the reader's attention to related graphics, *what* kinds of things to say about them, and *where* to place the cross-media cues in the text.

For exploratory purposes, we previously performed an informal corpus study of the use of cross-media cues in arguments (Green 2001). However, we contend that print-media-based corpus studies may not provide sound information on which to base a model for on-screen presentations. Human-computer interaction (HCI) studies have shown that there are significant differences between reading from print and computer media, e.g., that reading from screen is slower and comprehension is worse (Dillon, 1992; Muter, 1996). Thus, as an alternative to corpus analysis we have begun controlled user studies employing "throwaway" prototypes. In this paper, we present the design and preliminary

results of an experiment on effective cross-media cue usage in computer media.

2 Related Work

2.1 Computational linguistics

Cross-media cues are similar in some respects to *discourse cue phrases*. First, some functions of cross-media cues can be classified using discourse coherence relations such as Preparation, Restatement, Summary, Evaluation, and Elaboration (Green, 2001). Second, there is not a one-to-one correspondence between form and function. For example, the same CMC can be used to indicate different coherence relations between a span of text and the named figure, e.g., Restatement and Evaluation. On the other hand, a relation of Summary can be indicated, for example, by '**From Fig. 9.5, you can see that**' or '**(see Figure 4)**'. Another similarity is that CMCs are not always provided to mark explicitly the relationship obtaining between text and graphic. Research on discourse cue placement has framed our thinking on asking *when* and *where* to generate CMCs (DiEugenio, Moore and Paolucci, 1997).

A multimedia presentation may include *multimodal referring expressions*, references to things in the world made through a combination of text and graphics (McKeown et al., 1992; André and Rist, 1994). Such cross-references are similar to cross-media cues in that they direct the user's attention to a related graphic. However, their function is different, namely, to enable the user to perform reference resolution. Another form of cross-reference, *discourse deixis* is the use of an expression that refers to part of the document containing it, e.g., '**the next chapter**' (Paraboni and van

Deemter, 1999). Although a user's interpretation of a cross-media cue may depend on discourse deixis to determine the graphic in question, the problem of selecting an appropriate description to refer to a graphic (e.g. 'Figure 4' versus 'the Figure below') is not a concern of our work at present.

In our previous corpus study of multimedia arguments, we classified text in a document as either *argument-bearing* or *commentary-bearing*, where the latter is text about a graphic included in the document (Green 2001). The topics of commentary-bearing text include the graphic's role in the argument (e.g. '**From Fig. 9.5, you can see that**'), the interpretation of graphical elements in terms of the underlying domain and data, and salient visual features of the graphic. Furthermore, we noted that commentary-bearing and argument-bearing text may be interleaved, and that the ratio of the number of sentences of commentary to their related CMC may be many to one.

Previous work in caption generation is relevant to the question of *what* kinds of things to say about accompanying graphics (Mittal et al., 1998; Fasciano and Lapalme, 1999). However, neither of those systems face the problem of integrating commentary-bearing text with text generated to achieve other presentation goals.

2.2 Human-Computer Interaction

HCI research has focused on interaction techniques and features of layout that influence effectiveness. Use of *contact points*, control buttons in text on a web page that enable readers to control related animations (Faraday and Sutcliffe, 1999), is an interaction technique that, like CMCs, explicitly marks the relationship between information presented in two media. That paper provides experimental evidence that contact points improve comprehension of integrated text and animation.

According to Moreno and Mayer's Spatial Contiguity Principle (2000), learning in multimedia presentations is improved when related text and graphics are spatially contiguous rather than separated. However, this does not imply that instead of providing CMCs a generator can rely on layout alone, for the following reasons. First, a generator may have responsibility for producing text but not

have control over layout, e.g. when a document is displayed by a web browser. Second, a graphic may be relevant to multiple non-contiguous spans of text in a document.

3 Experiment

3.1 Overview

As a first step, we must address a basic question: is it ever worthwhile to generate cross-media cues in computer presentations? Thus we designed a between-groups experiment (Lewis & Rieman, 1994) to test whether performance on tasks requiring a subject to skim for information presented in text and graphics via a web browser would benefit from the inclusion of cross-media cues in the text. *Skimming*, defined as "moving rapidly through text to locate specific information or gain the gist", is a type of reading strategy often used by readers of web pages (Dyson and Haselgrove 2001).

Each of the three groups of subjects receives a different version of a presentation consisting of four articles. Each article fills a 19 inch computer screen and consists of a short text followed by several figures with information graphics such as line graphs and bar charts. The graphics are arranged in a row near the bottom of the screen so that the cost to the user of looking up and down between text and graphics is the same for each figure. Multiple figures are provided so that the reader is required to determine which figure is relevant to the task.

In version 1, the layout of each article consists of text containing no cross-media cues followed by the figures. A short caption is given under each graphic. In version 2, the caption text has been removed from the figures and integrated into the paragraph of text above the figures, i.e., it now functions as *commentary* text. Version 3 is identical to version 2 except that for each figure a cross-media cue of the form '**See Figure n.**' has been inserted in the text; the CMC is inserted following the commentary created from the corresponding caption in Version 1.

Version 1 represents the case where it is feasible to design the layout so that text commenting upon a figure can be placed in proximity to the figure (i.e. maximizing

adherence to the Spatial Contiguity Principle). We assume that task performance will be best for version 1 and include it in the experiment to provide a baseline. The main point of the experiment, however, is to compare performance on version 2 with performance on version 3. Then, if performance on version 3 is better, we have shown that CMCs can be useful to readers performing a similar task.

3.2 Experimental Design

The independent variable is the version of the article that is presented. The three versions are constructed by varying layout and presence of cross media cue phrases as described above. The dependent variables are the time to complete the tests (*Time*) and score on the tests (*Score*). Time and Score are compared between groups.

3.3 Participants

The participants (subjects) are undergraduate college students. The participants are randomly assigned to one of three groups. Each group is tested on a different version of the same articles. Information about college major and experience using computers is collected via a short questionnaire before the experiment.

3.4 Materials

Each article was constructed by the experimenter by selecting an excerpt from a published source; the sources of the four articles represent different genre, topics, layouts, and audiences. (We chose to use excerpts rather than authoring our own articles to avoid experimenter bias.) The excerpts are approximately the same word-length and, except for the first article, which is used for practice and only includes two figures, each excerpt includes three figures. The layout was modified by the experimenter to create versions 1 through 3. Other differences in presentation (e.g., line length, color scheme, font style, and font size) between different versions of the same article and between articles were minimized as much as possible.

The multiple choice test for each article consists of one question asking the subject to identify one of the main points of the presentation, and three questions asking the

subject to identify where in the presentation certain facts were given. For the identification questions the subject is asked to select one or more of the following choices: in the text, in the graph in Figure 1, in the graph in Figure 2, in the graph in Figure 3, or none of the above.

3.5 Procedure

Each participant is given a series of four tests displayed on a desktop PC with a 19 inch color monitor. The first test is used as a practice test and data collected from it will not be used. The test series is implemented by a computer program written in HTML and Javascript that is run by a web browser. Scrolling is disabled throughout the test series. The first screen of each test presents an article; the next screen contains the four test questions described above. The participant is free to move back and forth between the article and the test question screen for it by using Forward/Backward buttons, but cannot see the article and test question screens at the same time. The participant cannot go back to previous tests, and is not allowed to go on to the next test until he or she has answered all questions on the current test and has confirmed that he or she is ready to go on to the next test. The participant answers the test questions using the computer mouse. The program records the participant's answers and times automatically. Subjects are not told that their task time is being measured.

3.6 Status of Work

We have finished running the pilot version of the experiment and are currently running the main experiment. It is interesting that in the post-experiment questionnaire, some subjects who have received version 2 have commented that references to the figures (i.e. CMCs) would have been helpful.

4 Discussion

We have presented the motivation for and design of an experiment to evaluate the usefulness of cross-media cues in multimedia presentations shown on computer screens. In future work, we plan to investigate questions of cross-media cue placement, e.g., whether to insert a CMC before or after commentary

about the named figure. An interesting question is whether CMC placement should be influenced by discourse structure.

Acknowledgments

We thank Jennifer Brooks of the University of North Carolina at Greensboro for her implementation of much of the Javascript programs used in the experiment and for running an initial group of subjects through it.

References

E. André and T. Rist. 1994.. Referring to World Objects with Text and Pictures. *COLING-94*, 530-534.

A. Dillon. 1992. Reading from paper versus screens: a critical review of the empirical literature. *Ergonomics*, 35, 1297-1326.

M.C. Dyson and M. Haselgrove. 2001. The influence of reading speed and line length on the effectiveness of reading from screen. *International Journal of Human-Computer Studies*, 54, 585-612.

Barbara Di Eugenio, Johanna D. Moore, Massimo Paolucci. 1997. Learning Features that Predict Cue Usage, *Proceedings 35th Annual Meeting of the Association for Computational Linguistics*.

P. Faraday and A. Sutcliffe. 1999. Authoring Animated Web Pages Using 'Contact Points', in *Proceedings of CHI '99*, 458-465.

M. Fasciano and G. Lapalme. 1999. Intentions in the coordinated generation of graphics and text from tabular data. *Knowledge and Information Systems*, Oct 1999.

N. Green. 2001. An Empirical Study of Multimedia Argumentation. *Proceedings of the International Conference on Computational Systems, Workshop on Computational Models of Natural Language Arguments, May 2001*. Springer Lecture Notes in Computer Science 2073, pp. 1009-18.

Lewis & Rieman. 1994. Lewis, C. and Rieman, R. *Task-Centered User Interface Design: A Practical Introduction*. [ftp://ftp.cs.colorado.edu]

K. R. McKeown, S. K. Feiner, J. Robin, D.D. Seligmann, and M. Tanenblatt. 1992. Generating Cross-References for Multimedia Explanation. *Proceedings of AAAI*, 9-16.

V. Mittal, J. Moore, G. Carenini, and S. Roth. 1998. Describing Complex Charts in Natural Language: A Caption Generation System. *Computational Linguistics*, Vol. 24, issue 3, (1998), 431-467.

R. Moreno and R. Mayer. 2000. A Learner-Centered Approach to Multimedia Explanations: Deriving Instructional Design Principles from Cognitive Theory, *Interactive Multimedia Electronic Journal of Computer-Enhanced Learning*.

P. Muter. 1996. Interface design and optimization of reading of continuous text. In H. Van Oostendorp and S. DeMul (eds.) *Cognitive Aspects of Electronic Text Processing*, pp. 161-180.

I. Paraboni and K. van Deemter. 1999. Issues for the Generation of Document Deixis. In André et al. (Eds.), *Deixis, Demonstration and Deictic Belief in Multimedia Contexts, Proceedings of the Workshop associated with the 11th European Summer School in Logic, Language and Information (ESSLLI)*, Utrecht, The Netherlands, 1999, pp. 43-48.