EVALUATION OF GRAPHICS ON VIDEOTEX BY INEXPERIENCED USERS

Jo W. Tombaugh, Richard F. Dillon and Nancy L. Carboni
Department of Psychology
Carleton University
Ottawa, Canada K1S 5B6

ABSTRACT

This experiment assessed the reaction of inexperienced Telidon users to various graphics currently available on a videotex system. After becoming familiar with Telidon, the users were asked to rate 32 graphics and to indicate the graphic features which affected their ratings. By examining these ratings and the characteristics of the graphic, it was concluded that user ratings can be used to identify graphics that are considered to be particularly poor and particularly good. In addition, a major determinant of user ratings was the role of the graphic relative to the text. Graphics which illustrated and explained the text were rated more highly than those which served only as decoration.

RÉSUMÉ

Cette expérience vise à étudier les réactions d'usagers du système Telidon face à diverses illustrations vidéographiques présentement disponibles dans une banque videotex. On demande à des usagers ayant peu d'expérience avec le système d'évaluer 32 illustrations vidéographiques et d'indiquer les caractéristiques qui ont influence leurs jugements. A la suite de l'analyse des évaluations et des caractéristiques graphiques des illustrations, nous concluons que les évaluations par les usagers s'averent très utiles pour distinguer les bonnes illustrations des illustrations médiocres. De plus, les évaluations sont fortement influencees par la relation entre le texte présenté et l'illustration. Les illustrations qui favorisent la compréhension du texte sont évaluées plus favorablement que celles qui ne jouent qu'un rôle décoratif.

KEY WORDS: Telidon, videotex, graphic, ratings, users

INTRODUCTION

Videotex systems, especially the Canadian Telidon system, include the capability for high resolution colour graphics. However, it is not clear what functions the graphics will play, or how they can best be used in the videotex environment. It is inappropriate to extrapolate what is known about the use of graphics in a static medium to videotex. In fact, there are a number of differences between a graphic displayed on videotex and on the printed page as listed below.

Timing
: A printed picture is immediately available to the observer in full detail, while a videotex picture develops over time.

Sequencing
: The order in which the printed picture is drawn is not represented in the picture, while the order of drawing a videotex graphic is reproduced each time the page is shown.

Attentional value
: Printed pictures are often used as attention getters, for example to direct the eye toward an advertisement in a newspaper or magazine. On videotex, attention is normally already directed toward the television screen, so that the limited information on any retrieved page will be noticed.

This research was supported in part by a grant from the National Science and Engineering Council of Canada (Strategic Grant G0365).

Amount of information

The amount of text and graphics which can be displayed simultaneously on a videotex page is very restricted compared to the possibilities for a page in a book or magazine.

Thus, with these differences in mind, it seems likely that graphic artists will need expanded guidelines for the development of graphics on videotex -- guidelines which are sensitive to the timing and sequencing of the graphic presentation and sensitive to the expectations of viewers.

In a first attempt to deal with these issues, inexperienced users were asked to rate actual pages already created for videotex. The objective was to determine which graphics are liked by users and which are not, and to relate user preferences to the characteristics of the graphics. Ultimately, the goal is to furnish information providers -- the people responsible for putting pages on videotex -- with data on what graphic characteristics people like, and to provide a practical method of pretesting acceptance of specific graphic pages.

PROCEDURE

Subjects. Fifty subjects were volunteers from two sources. Twenty one responded to notices inviting them to see the new Telidon technology while 29 were recruited from a course on mass communications offered by the School of Journalism at Carleton University. Seventeen of the subjects were tested individually, while the other 33 viewed the graphics in 14 small groups of two to four persons.

Materials. Thirty two pages were selected from the information available on the VISTA field trials database in Toronto during the period from October to December 1981. The pages were chosen to represent a wide range of different approaches to the use of graphics on videotex.

Experimental Task. Use of the VISTA system was described and then subjects spent 15 minutes browsing through the database to provide them with the context in which graphics occur. The order of showing the 32 graphic pages was randomized for each subject to avoid order effects. Then, to provide a general idea of the quality of graphics in the set, the first 12 graphic pages were shown without subjects rating them. Subjects then evaluated the last 20 graphic pages. In answer to the question, "What is your general impression of this graphic?", subjects checked a seven point rating scale. Subjects were then asked to check which, if any, of 11 characteristics labelled "bad features", and 10 characteristics labelled "good features" influenced their opinion of the graphic. These features are described in the next section.

RESULTS

The user ratings were scored from 1 (terrible) to 7 (excellent) for each graphic. Because subjects differ in their use of any rating scale, the ratings for each subject were converted to standardized z scores having a mean of 0 and standard deviation of 1. Thus, each subject's mean rating was used as a criterion for determining what he judged as better than his own average (positive z score) or worse than his own average (negative z score). The z-score ratings were then used to determine the extent to which agreement existed among the subjects in identifying the better and poorer graphics. If the process of rating were essentially random, mean z-score ratings should be near zero when averaged over all subjects since some would give positive and some would give negative ratings. On the other hand, positive or negative means which were clearly different from zero would indicate general agreement among the subjects with positive means identifying those graphics which are better liked than most and negative means identifying those which are rated worse than most. For each graphic, 90% confidence intervals were placed around the z-score means to identify those where the means were clearly different from zero. Out of 32 graphics, there were 10 with ratings consistently below zero and 12 with ratings consistently above zero, showing that general agreement does exist on these graphics.

It was possible to determine what types of characteristics were noticed, and how those characteristics related to the overall ratings by examining the positive and negative features checked by subjects. These features, which each subject either checked or left blank for each graphic, are listed in Table 1. The percent of subjects who checked a given feature was calculated for each graphic. The mean percent checked for each feature, averaged

Table 1. Mean percent of features checked, and correlation of percent with rating scale.

Negative Features	Mean Percent Checked	Correlation With Rating	
Takes too long to present	27	-.29	
Poor use of colour	17	-.61	p<.05
Boring	12	-.58	p<.05
Confusing	12	-.57	p<.05
Non-informative	13	-.60	p<.05
Too detailed	7	-.42	p<.05
Annoying	8	-.70	p<.05
Unrelated to context	6	-.64	p<.05
Not enough detail	13	-.33	
Poor order of appearance of elements of graphic	7	-.53	p<.05
Distracting	12	-.56	p<.05
Mean	12		
Positive Features			
Adds interest	36	.65	p<.05
Helps to explain text	38	.49	p<.05
Entertaining	20	.44	p<.05
Very detailed	18	.23	
Good use of colour	19	.67	p<.05
Imaginative	24	.35	p<.05
Well designed	38	.78	p<.05
Pleasing	17	.56	p<.05
Colourful	21	.26	
Good use of movement	27	.31	
Mean	26		

over the 32 graphics is shown in the first column of Table 1. These values indicate the frequency of use of each feature in identifying the characteristics influencing the ratings. For example, the feature "Takes too long to present" was selected on 27% of the rating sheets while "Poor use of colour" was selected on 17% of the sheets. There was a greater tendency to check positive features, (26%) than negative features (12%). This difference was statistically significant ($t=4.29$, 19 df).

Also shown, in the last column of Table 1, are the Pearson Product Moment correlations between scores on the seven point rating scales and the frequency of use measures described above. The 32 graphics served as the units of analysis. Thus, for each correlation, 32 pairs of scores, one pair for each graphic, were analysed. The percent of the subjects who selected the feature for a given graphic was one of the scores in a pair, and mean rating on the scale from 1 to 7 for that graphic by the subjects was the other. Positive correlations indicate that the percent who checked a feature increased as the rating of the graph increased. Positive correlations were obtained for all 10 positive features. Negative correlations indicate that the percent who checked a feature increased as the mean rating decreased. As would be expected, negative correlations were obtained for all 11 "bad" features. In general, positive and negative correlations were fairly high indicating that the features people checked were important in their overall rating of each graphic. It is interesting, however, that the most frequently checked negative characteristic ("Takes too long to present") is poorly correlated with the rating ($r = -.29$,

Table 2. Subscales suggested by Factor Analysis, and correlations of subscales with Ratings.

Subscale Name	Features Included	Correlation of Subscale With rating Scale	
Bad Design	Distracting, Confusing, Unrelated to context, Annoying	-.72	p<.05
Slow and Detailed	Takes too long, Too detailed	-.35	p<.05
Amusing	Entertaining, Imaginative, Good use of movement	.45	p<.05
Good Colour	Colourful, Good use of colour	.55	p<.05
Functional	Helps to explain text, Adds interest	.66	p<.05

p>.05). This suggests that subjects rated some graphics positively, even when they felt that they took too long to present.

Trying to predict whether people will like graphics from 21 features which are correlated with each other to varying degrees is undesirable. Consequently, with the aid of factor analysis and regression analysis, five subscales were identified to summarize the information about features. These scales are shown in Table 2. For example, the frequencies for the "Distracting", "Confusing", "Unrelated to context", and "Annoying" features tend to group together and measure the same thing. The "bad design" scale is simply the sum of these frequencies. Simple correlations of the subscales with the ratings were statistically significant and a multiple regression showed that the independent contributions of subscales to the predicted rating scale were all statistically significant. In addition, the multiple correlation coefficient was .95 indicating a strong relationship between a subject's overall rating of a graphic and the types of strengths and weaknesses which he identifies for it. In fact, the square of the multiple correlation coefficient (.90) is the coefficient of determination. This measure indicates that 90% of the differences in overall ratings among the graphics can be explained in terms of the five subscales or dimensions.

A final interest of the study was to attempt to identify properties of the graphics themselves which correlated with user ratings. The question addressed is whether graphic characteristics can be used by the information provider in the absence of user ratings to lead to well accepted graphics. These characteristics will be called graphic properties to distinguish them from the user-checked features. Mean ratings for some of the properties measured are shown in Table 3. Again, the z-scores for the rating scales are shown, so that negative scores indicate graphics which are less well liked than average, while positive scores represent those which are better liked. Many of the properties measured tended to be categorical in nature, so that analysis of variance rather than correlation was used to examine the relationship of the property to the rating scale results. For example, the "type of fill" categories -- predominately unfilled line drawings, filled with solid colours, filled with a mixture of lines, checks, dots, etc. — are qualitative categories. Time to display the graphic, a continuous variable, was also categorized, because the relationship between time and ratings was not linear. In Table 3, N refers to the number of graphics which fell within each category. For example, ten graphics took less than ten seconds to display on the screen.

As can be seen from Table 3, the only graphic properties which led to statistically significant differences on the rating scales were (1) whether or not the drawing was illustrative of the text and (2) whether the function of the graphic was to decorate, entertain, or explain the text. Drawings which illustrated the text, and those which explained the text received better ratings. Those unrelated to the text or used as decoration were poorly rated.

Table 3. Relationships between properties of graphics and users ratings.

Graphic Properties	N	Mean Rating (z scores)	Signif.
Time to display graphic			n.s.
0-10 sec.	10	-.06	
11-20 sec.	12	.11	
21-30 sec.	4	.10	
> 30 sec.	6	-.26	
Type of fill			n.s.
Predominately unfilled line drawings	7	.16	
Filled with solid colours	19	-.07	
Filled with mixture of lines, checks, dots, etc.	6	-.04	
Changes in graphic over time (Other than development of a drawing)			n.s.
No	24	-.07	
Yes	8	.15	
When text appears			n.s.
Before graphic	11	-.15	
With graphic	9	.18	
After graphic	12	-.04	
Screen dominance			n.s.
Less text then graphic	3	-.21	
Equal text and graphic	14	-.15	
More text than graphic	15	.15	
Illustration of text with graphic			$p<.05$
No	9	-.37	
Yes	23	.12	
Function of graphic			$p<.05$
Decoration	3	-.45	
Entertainment	18	-.14	
Explanation	11	.30	
Number of colours used in graphic (excluding grays)			n.s.
1	5	-.39	
2	4	.04	
3	5	-.36	
4	10	.25	
5	6	.07	
6	2	.06	

DISCUSSION

It is clear that inexperienced users of videotex have preferences for certain types of graphics, and that they can relate these preferences to graphic features. Thus, information providers can evaluate the acceptability of their graphics by having them rated by potential users. Using the method outlined in the procedure section, a large number of graphics can be rated in a short period of time, and the results can be quickly summarized to show which graphics are preferred, as well as the specific features which relate to their strengths and weaknesses.

In terms of graphic properties shown in Table 3 -- characteristics which the information providers can examine without collecting user ratings -- the present study shows that users are sensitive to the utility of the graphic as a supplement to the text. Thus, users prefer graphics which illustrate the text, as well as those which are more than just decoration. It is clear that information providers should concentrate their efforts on the development of graphics which are related to the text as opposed to decorative.

Other than the function served by the graphic, the kinds of graphic properties that the information provider could use in the absence of the user ratings were not clearly identified by the present study. However, the potential for detecting such effects was not high, due to the large amount of uncontrolled variation among graphics. That is, two graphics differing in whether they were a line drawing, or filled drawing would also vary on many other dimensions such as content, number of colours used, etc. Thus, the failure to find significance in this exploratory study should not lead to the conclusion that these graphic properties have no effect on the ratings. In fact, with the exception of the classification in terms of number of colours used, the orderings of the mean z-score ratings shown in Table 3 seem quite reasonable. For example, line drawings had a higher rating (+.16) than those which contained fill (-.06). A more sensitive investigation of type of fill would involve creating several sets of graphic pages differing in terms of fill, but having identical content and composition otherwise. A similar approach could be taken to investigate a factor such as the sequencing of text and graphic content.

Because of its special relevance to page display on videotex, the findings related to the time required for page transmission is of special interest. Time, whether measured as a graphic property, or checked as a user feature, did not significantly relate to the ratings, even though "Takes too long to present" was the most frequent user complaint. The rating of slower graphics depends on what is happening to the graphic during that time. Slow presentation was associated with negative ratings only for graphics which were also judged to be too detailed, as indicated by the second subscale in Table 2. This low rating of slow and detailed graphics was predicted by Mills, 1981, who noted that detailed drawings may not be as suited for videotex as simple line drawings, precisely because of their slower rate of transmission.

A limitation of the present study is that only inexperienced videotex users were asked to rate the graphics. In addition, the graphics were presented out of context, and only once. To establish more general guidelines, the ratings of experienced, as well as first time users of videotex should be examined, and the role of the graphic (e.g., as a menu page, a page repeatedly encountered in a game, a drawing of a house for sale) should be considered. It might be hypothesized that experienced users would be more likely to check negative features, and would be less tolerant of the slower graphics.

REFERENCE

Mills, M. I. A study of the human response to pictorial representations on Telidon. Telidon Behavioural Research3. Department of Communications, Ottawa, Canada.