

Multimodal Comprehension of Language and Graphics: Graphs with and without annotations

Cengiz Acarturk
University of Hamburg

Christopher Habel
University of Hamburg

Kursat Cagiltay
Middle East Technical University

Ozge Alacam
Middle East Technical University

An experimental investigation into interaction between language and information graphics in multimodal documents served as the basis for this study. More specifically, our purpose was to investigate the role of linguistic annotations in graph-text documents. Participants were presented with three newspaper articles in the following conditions: one text-only, one text plus non-annotated graph, and one text plus annotated graph. Results of the experiment showed that, on one hand, annotations play a bridging role for integration of information contributed by different representational modalities. On the other hand, linguistic annotations have negative effects on recall, possibly due to attention divided by the different parts of a document.

Keywords: Eye movements; Multimodal comprehension; Graph comprehension; Graph-text documents; Annotations

Introduction

Documents containing information in different representational modalities (e.g., text, tables, pictures, diagrams, graphs) are used frequently in communicative settings in daily life such as newspapers as well as in scientific articles and educational settings like textbooks and classroom presentations.ⁱ

Guidelines on design and appropriate use of statistical information graphics for communicative purposes have been published regularly.ⁱⁱ Nevertheless, well-founded research on perceptual and cognitive processes in comprehension of graph-text documents is relatively rare. The standards proposed by institutions are often limited to suggestions for how to prepare graphs for specific purposes. For example, in the American National Standards Institute (ANSI) manual on guidelines for time series charts (1979, cited in Cleveland & McGill, 1984), it is stated that “This standard...sets forth the best current usage, and offers standards ‘by general agreement’ rather than ‘by scientific test’”(p. iii).

The research literature on graph comprehension shows that the scientific studies on information graphics were mostly performed after the 1980s, except for a few in the first half of the century (e.g., Washburne, 1927; Graham, 1937). In the last three decades, research on principles of preparing, analyzing, and visualizing quantitative information by graphs and other types of visualization has attracted researchers from different disciplines. Some example studies include Bertin (1983), Tufte (1983), Wainer (1984), and Kosslyn (1994). Another research trend has mainly focused on processing perspective. Perceptual processes (i.e., pre-attentive processes as proposed by Cleveland and McGill, 1984, and Simkin & Hastie, 1987) have been the focus of research more often than further information processing (i.e., processes of graph comprehension).ⁱⁱⁱ Currently, state of the art cognitive processing models of graph comprehension are scarce (Scaife & Rogers, 1996). Some exceptions include research on problem solving in informationally equivalent but computationally different graph types (Peebles & Cheng, 2002) and cognitive architectures and processing models for graph comprehension (Pinker 1990; Mautone & Mayer, 2007).

In this study, we focus on the use of graphs for communicative purposes. Since their first use in the form of the Cartesian graph frame in the 17th Century (see Wainer & Velleman, 2001; Spence, 2005; Shah, Freedman, & Vekiri, 2005, for recent reviews), different types of graphs have been employed for various purposes. On the one hand, graphs facilitate understanding and insight. For example, as described by Tufte (1983), John Snow found the origin of the 1854 London cholera epidemic by using a map-based diagram. In this diagram, he showed that, among the 13 pumps supplying water from wells, the majority of people affected by the disease lived closer to a central pump on a specific street. On the other hand, graphs facilitate communication. For example, when researchers face difficulties in presentation of numerical results via parametric statistics and other treatment of data with mathematics, they use graphs to convey information to end-users (Arsenault, Smith, & Beauchamp, 2006). Furthermore, graphs are used in educational settings (Newcombe & Learmonth, 2005) and for the purpose of visualizing trends to laypersons (Berger, 2005).

Theoretical Aspects and the Purpose of the Study

Whereas one representational modality may have advantage over the other with respect to computational efficiency in processing (Larkin & Simon, 1987), visualizations are generally not self-explanatory. A minimum amount of verbal information is required (e.g., axis labels in a line graph identify the meaning of the lines). This information is also a prerequisite to construct internal mental representation of the graph relative to the mental representation of the text. This type of multimodality, which is beyond the traditionally researched modality between graphs and paragraphs in a multimodal document, can be called *graph-internal multimodality* (Acarturk, Habel, & Cagiltay, 2008). From this perspective, graph comprehension is inherently multimodal not only with respect to the text that accompanies a graph, but also with respect to integration of graphical and textual elements on the graph.

The minimum requirement in comprehension of multimodal documents is that comprehension is possible as long as the information contributed by different representational modalities can be integrated under common conceptual representations. More specifically, interaction between graphs and language in graph-text documents is mediated by common conceptual representations (Habel & Acarturk, 2007).^{iv} As an extension to Chabris and

Kosslyn's (2005) "Representational Correspondence Principle," which states that "effective diagrams depict information the same way that our internal mental representations do" (p. 40), we propose that information presented in different modalities contributes to a common representation as long as the internal mental representations of the constituents with different modalities are in line with one another.^v Specifically for laypersons, the construction of a common representation in a text-graphic document is possible when the internal mental representation of the graph is in line with that of the text. In particular, a set of basic spatial concepts is fundamental for the terminology of graphs as well as for the terminology of the specific domain in which an individual graph is applied.

We propose that linguistic annotations, on the one hand, provide graph-internal information, which is necessary for graph-internal multimodal integration, by foregrounding graphical entities (and/or real-world entities referred to by the graphical constituents).^{vi} Linguistic annotations, on the other hand, provide information that facilitates integration between paragraphs and graphical elements in a graph-text document. An example graph-text document with an annotated graph is shown in Figure 1. The graph region includes the graph (proper) and graph-related text information (graph title, captions, annotations, etc.). The graph region is usually separated from the rest of the document (e.g. paragraphs) by a visual frame.



Figure 1. A sample multimodal document with graph and text. The graph has the graph (proper), the graph title ("New High for the Dow") and linguistic annotations. The annotation icons (i.e., the thin straight lines) connect relevant parts of the graph to annotation text. (© The New York Times. *Dow Jones Index Hits a New High, Retracing Losses*, by Vikas Bajaj, published on October 4, 2006).

Annotations highlight the relevant parts of a graph via annotation text and annotation icons. These elements attract the attention of the reader during comprehension. Furthermore, annotation text foregrounds real-world referents of graphical elements (e.g., events occurring at different times). Descriptions provided by annotations may refer to different levels of representation such as the verbal description of a graphical property or the description of a real-world event referred to by the graphical constituents. In addition, in graph-text documents, the content of annotation text usually corresponds to graphical or real-world referents of the graph mentioned in the accompanying text. For this reason, annotations may have the function of bridging the information presented in the graph and in the text. This may result in facilitatory effects in comprehension of graph-text documents.^{vii}

Multimodal comprehension research in educational psychology reveals diverging findings. Using multiple representational modalities may facilitate learning under certain conditions (e.g., Winn, 1987; Carney & Levin, 2002; Mayer, 2001; 2005). Mayer and his colleagues call this the *multimedia effect*. This effect suggests that adding graphs to text may facilitate learning. More specifically, using verbal cues may facilitate integration of the information contributed by different modalities, namely the *spatial contiguity effect*. This effect is in parallel with our proposal for the facilitatory effects of annotations on graphs. However, integrating information contributed by different modalities may require additional cognitive effort (Tabachneck-Schijf, Leonardo, & Simon, 1997; Chandler & Sweller, 1992). Sweller and his colleagues propose that such inhibitory effects, namely the *split-attention effect*, may be due to limited working memory capacity.

Our purpose in this study is to investigate the role of annotations in graph-text documents with respect to their facilitatory and inhibitory effects. Adding graphs to text may increase learning due to the multimedia effect (Mayer, 2001; 2005). Furthermore, the bridging role of annotations between the information contributed by the two representational modalities may result in facilitatory effects. If annotations are available on a graph, separately constructed representations of the text and the graph may be integrated via these constituents. If there is no annotation on the graph, integration of information contributed by the different modalities should be achieved with further cognitive effort of encoding spatially represented

information on the graph and constructing co-reference relations between the paragraphs and the graph. At the same time, the spatially separate presentation of information on the layout (i.e., the graph and the paragraphs) may result in inhibitory effects due to the split-attention effect.

The main assumptions of the study comprise assumptions about participants' knowledge of information graphics and the experimental methodology. Graph comprehension requires a minimum amount of knowledge and skill in understanding simple graph types. Graph understanding is generally included in education curriculum at early levels. Nevertheless, adults can generalize from a small set of graphs learned in school to a countless number of graphical forms used for communicative purposes (Pinker, 1990). We assume that, beyond being part of the school education curriculum, graphs reflect human cognition and, more specifically, conceptual representations of events (rather than events themselves) via graphical elements. Concerning the domain of the study, we selected the stock market because it is one domain in which annotated line graphs are used abundantly. Although multimodal documents such as stock market graphs accompanied by text appear frequently in newspapers and magazines and there are studies for knowledge-based generation of stock market reports (Kukich, 1983; Sigurd, 1995), as of our knowledge, there is no systematic experimental investigation of multimodal comprehension of stock market graphs.

Methods

Participants and Materials

Thirty-two Middle East Technical University students or graduates from different departments were paid to participate in the experiment. There were 20 females and 12 males with normal or corrected vision. The mean age was 22.0 (SD = 3.68). Each participant received three different newspaper articles. Each article included a different story in the stock market domain. Each article was presented to the participants in one of the following three conditions (see Figure 2 for examples):^{viii}

- i. *Text plus annotated graph* condition (shortly, *annotated-graph* condition): In this condition, the paragraphs were presented together with the accompanying graph. In addition, the graph included linguistic annotations.

ii. *Text plus non-annotated-graph* condition (shortly, *non-annotated-graph* condition): In this condition, the paragraphs were presented together with the accompanying graph, but the graph did not include linguistic annotations.

iii. *Text-only* condition: In this condition, the paragraphs were presented without an accompanying graph.

In conditions i and ii, the graphs were placed to the left of the paragraphs. In the text-only condition, the region occupied by the graph was left blank.

The text content of the documents (i.e., the paragraphs) was based on the original articles; some of the content was omitted for the purpose of reducing the size of the document. The resulting three articles presented to participants consisted of 152, 174 and 180 words in the paragraphs. The main protagonists in the articles (the Turkish stock market IMKB and the Dow Jones) were also changed to less known alternatives (namely, Taiwan, Hungary, Budapest, and Singapore stock markets), for the purpose of preventing interference with previous knowledge. All experimental stimuli were in Turkish, the native language of the participants. Since the graphs in the original articles were different from each other in design layout, the graphs were redrawn but maintained a similar appearance to the original graphs. The material was presented to the participants in random order on a computer screen with full screen view. Distracter scenarios were presented between the experimental scenarios. A cross sign was presented before each experimental screen at the center of the screen.

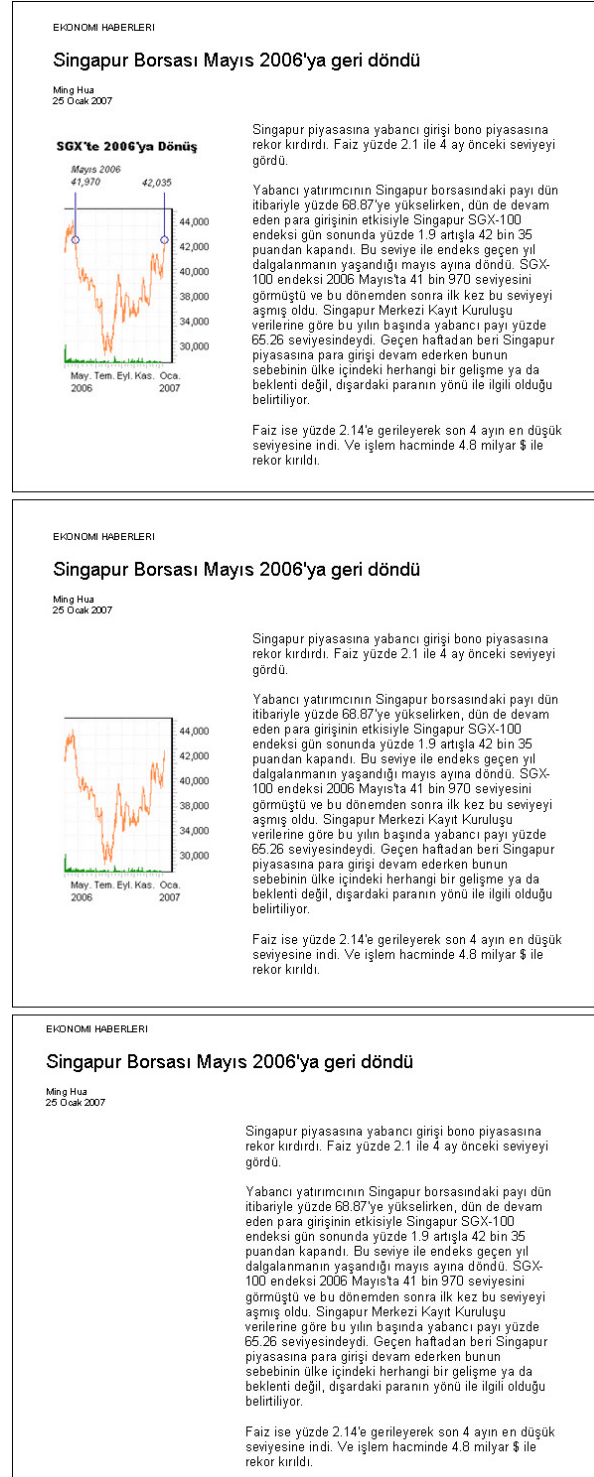


Figure 2. Sample material from the experiment. The top figure shows the annotated-graph condition; the middle figure shows the non-annotated-graph condition; and the bottom figure shows the text-only condition.

Procedure

A 50 Hz. non-intrusive eye tracker recorded eye-tracking data. The eye tracker was integrated into a 17" TFT monitor with a resolution of 1024 X 768 pixels. Subjects were seated in front of the screen at an approximate viewing distance of 60 cm. Spatial resolution and accuracy of the eye tracker was about 0.25° and 0.50° degrees respectively. Annotation labels were within 1.5° of visual angle on average. The horizontal axes of graphs were within 3.02° of the visual angle. Participants were tested in single sessions. After calibration of the eye tracker, participants read the instructions. They learned that they would see three real but modified news articles; after the presentation of the articles, they would be expected to answer a set of questions. After reading the instructions at their own pace, participants moved to the next screen by pressing a key on the keyboard. There were no time limitations in the experiment. The entire session took approximately 10-15 minutes. After the experiment session, the participants answered the post-test recall questions on paper.

Analysis

We analyzed fixation count, fixation duration, and gaze time as the three dependent variables of the experiment. The Area of Interest (AOI) specifications are shown in Figure 3. The non-annotated-graph condition and the text-only condition did not have every AOI since they did not have the corresponding stimuli regions.

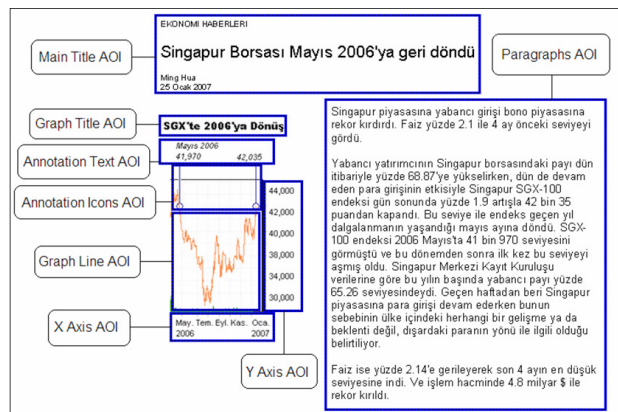


Figure 3. Sample showing Area of Interest (AOI) specifications.

Results

We investigated the results in three parts, namely analysis of the eye movement parameters, general observations for eye movement patterns, and analysis of answers to post-test recall questions.

Analysis of Eye Movement Parameters

Eye tracker calibration partially or totally failed in seven cases either due to inaccurate calibration at the beginning or loss of calibration in later stages of the experiment. The data from one participant was also eliminated after a self-assessment of experience in stock market graphs. As a result, data from 24 subjects were included into the analysis. Fixation count, fixation duration, and gaze time values were calculated on the specified AOIs for the analysis. Fixations with duration of less than 100 ms were not included into the analysis.^{ix}

Total Fixation Counts. Total fixation counts were calculated on the AOIs under the three conditions. Table 1 shows the total number of fixations of all the subjects in the experiment. Mean values can be found in Appendix A.1.

Table 1
Total fixation counts on the specified AOIs.

	Annotated-graph Condition	Non-annotated-graph Condition	Text-only Condition
Main Title	312	252	301
Paragraphs	3,462	3,560	3,327
Graph Line	29	76	-
X Axis	13	41	-
Y Axis	7	26	-
Graph Title	43	-	-
Annotation Text	71	-	-
Annotation Icons	32	-	-

A two-way analysis of variance was conducted with two within subjects factors, AOI (main title versus paragraphs) and condition (annotated-graph, non-annotated graph, and text-only). The AOI main effect, the Condition main effect, and the AOI x Condition interaction effect were tested using the multivariate criterion of Wilks's Lambda (Λ). The AOI main effect was signifi-

cant, $\Lambda = .13$, $F(1, 20) = 137.28$, $p < .01$. The Condition main effect was not significant, $\Lambda = .99$, $F(2, 19) = 0.13$, $p = .87$. Also, the interaction effect AOI x Condition was not significant, $\Lambda = .96$, $F(2, 19) = 0.37$, $p = .87$. This result suggests that the use of a graph in the stimuli, either annotated or non-annotated, did not affect mean fixation counts on Main Title AOI and Paragraphs AOI, compared to the text-only condition. In other words, there was no significant difference between the three conditions with respect to fixation counts on the title and paragraphs of the documents.

A second analysis of variance was conducted with two within subjects factors, AOI (graph line, X axis, and Y axis) and condition (annotated-graph versus non-annotated-graph). The results of the test showed that the AOI main effect was significant, $\Lambda = .62$, $F(2, 19) = 5.80$, $p < .01$; the Condition main effect was significant, $\Lambda = .70$, $F(1, 20) = 8.58$, $p < .01$; and the interaction effect AOI x Condition was not significant, $\Lambda = .85$, $F(2, 19) = 1.71$, $p = .21$. The paired-samples *t* tests, conducted to follow up the significant effects, showed that the difference in mean fixation counts on Graph Line AOI was significant between the two conditions, $t(20) = 2.66$, $p < .05$. The 95% confidence interval for the mean difference was 0.48 and 3.99. The difference between the conditions was also significant for X axis AOI, $t(20) = 2.75$, $p < .05$. The 95% confidence interval was 0.32 and 2.34. Nevertheless the difference was not significant for Y axis AOI, $t(20) = 1.96$, $p = .06$.

In summary, although subjects' fixation counts on paragraphs were not different between the three conditions, fixation counts on the graph, specifically on Graph Line AOI and on X Axis AOI (but not on Y Axis AOI), were higher in the absence of annotations.

Mean Fixation Durations. The mean fixation durations on the AOIs under the three conditions are shown in Table 2.

Table 2
Mean fixation durations on the specified AOIs.

	Annotated-graph Condition	Non-annotated-graph Condition	Text-only Condition
Main Title	273.60 (81.10)	258.39 (78.86)	272.08 (67.93)
Paragraphs	287.86 (79.23)	302.10 (89.01)	292.96 (72.57)
Graph Line	220.47 (49.30)	321.85 (183.76)	-
X Axis	248.44 (73.20)	390.07 (186.52)	-
Y Axis	336.30 (230.36)	370.47 (104.10)	-
Graph Title	314.65 (202.43)	-	-
Annotation Text	326.89 (154.97)	-	-
Annotation Icons	209.45 (125.69)	-	-

Note. All numbers are in ms. The numbers in parentheses show standard deviation.

A two-way analysis of variance was conducted with two within subjects factors, AOI (main title versus paragraphs) and condition (annotated-graph, non-annotated graph, and text-only). The results showed that the AOI main effect was significant, $\Lambda = .79$, $F(1, 20) = 5.37$, $p < .05$; the Condition main effect was not significant, $\Lambda = .99$, $F(2, 19) = 0.04$, $p = .96$; and the interaction effect AOI x Condition was not significant, $\Lambda = .87$, $F(2, 19) = 1.44$, $p = .26$. The results of the test suggest that adding graph to the text, either annotated or non-annotated, did not affect mean fixation durations on Main Title AOI and Paragraphs AOI.

A second analysis of variance was conducted with two within subjects factors, AOI (graph line, X axis, and Y axis) and condition (annotated-graph versus non-annotated-graph). The test indicated a significant Condition main effect, $F(1, 54) = 5.30$, $p < .01$. The AOI main effect was not significant, $F(2, 54) = 1.54$, $p = .22$. The

Condition x AOI interaction effect was also not significant, $F(2, 54) = 0.51, p = .60$. Further pairwise comparisons showed that mean fixation duration on the Graph Line AOI was significantly different between the two conditions, $t(16) = -2.05, p < .05$. Furthermore, mean fixation duration on the X Axis AOI was significantly different between the two conditions $t(12) = -2.20, p < .05$. Nevertheless, the difference was not significant for the Y Axis AOI between the two conditions, $t(5) = 0.31, p = .38$.

In summary, although average fixation durations were different between the title and the paragraphs, more importantly, fixation durations on the titles and the paragraphs were not significantly different among the three conditions. On the other hand, fixation durations on the graph, specifically on Graph Line AOI and X Axis AOI, but not in Y AOI, were higher in the absence of annotations than in the annotated-graph condition. This reflects higher cognitive effort of subjects in the non-annotated-graph condition.

Total Gaze Time. Table 3 shows the total gaze times of all subjects on the AOIs under the three conditions. Note that the values in the table show the total time spent by all subjects. Mean values can be found in Appendix A.2.

Table 3
Total gaze times on the specified AOIs.

	Annotated-graph Condition	Non-annotated-graph Condition	Text-only Condition
Main Title	86,819	67,297	81,553
Paragraphs	1,014,568	1,095,401	985,789
Graph Line	6,916	20,640	-
X Axis	3,608	15,890	-
Y Axis	2,309	10,230	-
Graph Title	15,849	-	-
Annotation Text	24,534	-	-
Annotation Icons	11,101	-	-

Note. All numbers are in ms.

A two-way analysis of variance was conducted with two within subjects factors, AOI (main title versus para-

graphs) and condition (annotated-graph, non-annotated graph, and text-only). The results showed that AOI main effect was significant, $\Lambda = .12, F(1, 20) = 69.71, p < .01$; condition main effect was not significant, $\Lambda = .96, F(2, 19) = 0.38, p = .69$; and interaction effect AOI x Condition was also not significant, $\Lambda = .92, F(2, 19) = 0.85, p = .44$. This result suggests that addition of a non-annotated-graph or addition of an annotated-graph to the text did not change the gaze time on the title or on the paragraphs.

A two-way analysis of variance was conducted with two within subjects factors, AOI (graph line, X axis, and Y axis) and condition (annotated-graph versus non-annotated-graph). The results of the analysis showed that the AOI main effect was significant, $\Lambda = .66, F(2, 19) = 4.85, p < .05$. The Condition main effect was also significant, $\Lambda = .68, F(1, 20) = 9.26, p < .01$. The interaction effect AOI x Condition was not significant, $\Lambda = .89, F(2, 19) = 1.15, p = .34$. Three paired-samples *t* tests were conducted to follow-up the significant effects. Differences in total gaze times between the two conditions were significantly different between Graph Line AOIs, $t(20) = 2.80, p < .05$; between X axis AOIs, $t(20) = 2.54, p < .05$; and between Y axis AOIs, $t(20) = 2.06, p = .05$.

In summary, addition of the graph, either in annotated or non-annotated form, did not affect the gaze time on the paragraphs. Nevertheless, gaze time on the graph region, specifically in Graph Line AOI, X Axis AOI, and Y Axis AOI, was longer in the absence of annotations than in the annotated-graph condition.

General Characteristics of Scan Paths

In this section, we verbally describe general characteristics of scan paths, since quantitative analysis of scan paths is technically difficult. We focus on shifts of eye movements among the three regions of the presented stimuli for simplicity.

- i. *Main Title* region: This region corresponds to the Main Title AOI. It covers only the main title of the document.
- ii. *Paragraphs* region: This region corresponds to the Paragraphs AOI. It covers only the paragraphs in the document.
- iii. *Graph* region: This region corresponds to the sum of the remaining AOIs. It covers the graph (proper), axes, graph title, and annotations.

The results of the transcription analysis showed that text was attended before the graphs. Furthermore, the paragraphs were usually read before the accompanying graph was inspected. These findings are parallel to the findings in previous research on multimodal comprehension (Hegarty, Carpenter, & Just, 1996; Rayner, Rotello, Stewart, Keir, & Duffy, 2001; Carroll, Young, & Guertin, 1992). In addition, the number of shifts between the Graph region and the Paragraphs region was higher in the annotated-graph condition than in the non-annotated-graph condition. Table 4 shows percentages of subjects who had more than one shift and more than two shifts between the Graph region and the Paragraphs region in a single session.

Table 4
Percentage of subjects who had more than one shift and more than two shifts between the Graph region and the Paragraphs region in annotated-graph condition and in non-annotated-graph condition.

	Annotated-graph Condition	Non-annotated-graph Condition
More than one shift	66.7	42.9
More than two shifts	45.8	23.8

Note. All numbers are in percentages.

It should be noted that the shifts between the Graph region and the Paragraphs region in the annotated-graph condition were mainly between annotation text and the paragraphs, whereas the shifts between the Graph region and the Paragraphs region in the non-annotated-graph condition were between graph (proper) and the paragraphs. On the other hand, the results of the fixation count analysis showed that the number of fixations on the graph (proper) was higher in non-annotated-graph condition than in the annotated-graph condition. This result reveals that the subjects investigated the documents in non-annotated-graph condition by fixating consecutively within the Graph region; whereas, the documents in the annotated-graph condition were investigated by more fixations between the paragraphs and annotations, rather than by consecutive fixations in the Graph region.

Analysis of Answers to Post-test Questions

The participants answered a post-test questionnaire after the presentation of the experimental stimuli. The questionnaire included nine multiple-choice recall questions about the stimuli scored each correct answer as 1, each incorrect answer as -1, and each missing answer as 0. We included answers from all 32 participants in the analysis. The outlier data (i.e., mean plus/minus two standard deviations) were eliminated from the analysis. The detailed scores can be found in Appendix A.3.

A one-way analysis of variance was conducted with a within subjects factor condition (annotated-graph, non-annotated graph, and text-only). The dependent variable was the Test Score; the independent variable was Condition. The results of the analysis of variance showed that the Condition main effect was significant, $\Lambda = .79$, $F(2, 25) = 3.32$, $p = .05$. Further pairwise comparisons showed that post-test scores in the non-annotated-graph condition were significantly higher than the post-test scores in the annotated-graph condition, $t(26) = -2.21$, $p < .05$, and than the post-test scores in the text-only condition, $t(26) = 2.36$, $p < .05$. The difference between post-test scores in text-only condition and annotated-graph condition was not significant, $t(26) = 0.00$, $p = 1.00$.

In summary, post-test scores showed that subjects' recall was better if the graph was included with the paragraphs, but the score was reduced if the annotations were added to the graph.

Discussion

The results of the experiment have been investigated in three parts. In the first part, the analysis of fixation counts, fixation durations, and gaze times on the specified AOIs under the three conditions showed that the presence of the graph, either in annotated or in non-annotated form, did not influence eye movement behavior on the title of the document and the paragraphs. On the other hand, the comparison between the annotated-graph condition and the non-annotated-graph condition showed that - under the assumption that fixation count, fixation duration, and gaze time reflect processing characteristics on the stimuli, the subjects' cognitive effort on the graph was higher in the non-annotated-graph condition than in

the annotated-graph condition. More specifically, we found that on the graph-related AOIs, fixation counts were higher, fixation durations were longer, and gaze times were longer in the non-annotated-graph condition than in the annotated-graph condition. Although the analysis of fixation counts and fixation durations partially supported this finding (i.e., the difference between the two conditions were significant for the Graph Line AOI and the X Axis AOI but not for the Y Axis AOI), the analysis of total gaze times resulted in significant differences between the two conditions for all three AOIs.

In the second part, the analysis of gaze patterns showed that the number of shifts between the paragraphs and the graph in the annotated-graph condition was higher than the number of shifts in the non-annotated-graph condition. On the other hand, the fixations in the non-annotated-graph condition were consecutive within the graph region. This result supports the idea that annotations form a bridge between the information contributed by the two representational modalities, namely the linguistic information contributed by the title and paragraphs and the graphical and linguistic information contributed by the elements in the graph region. A possible explanation for this is that the annotations in the experiment included date and value information, which are also represented by the x and y-axis of the graph. For this reason, the use of annotations possibly reduced the need to extract date information from the x-axis and value information from the y-axis. The situation is more evident in the analysis of eye movement data on the x-axis, which was spatially more distant from the graph line than the y-axis. In addition, since annotations highlighted the relevant parts of the graph with respect to the real-world referents of the graphical constituents and foregrounded the events mentioned in the text, it may be easier to construct the cross-modal relations between the text and the graph in the annotated-graph condition than in the non-annotated-graph condition. In summary, these findings support the idea that annotations serve the role of bridging the text and the graph.

Nevertheless, from the perspective of recall, the use of annotations introduces disadvantages. The post-test scores were presented in the third part of the analysis. The results showed that first, post-test scores in the non-annotated-graph condition were higher than the post-test scores in the text-only condition. This finding partially supports the multimedia effect (Mayer, 2001; 2005). The

finding is partial because post-test scores in the annotated-graph condition were not higher than post-test scores in the text-only condition. On the other hand, the finding that post-test scores in the non-annotated-graph condition were higher than the ones in the annotated-graph condition shows that the more frequent shifts between the paragraphs and the graph region of the document in the annotated-graph condition may have resulted in the split-attention effect (Chandler & Sweller, 1992).

In summary, the analysis of eye movement behavior supports the role of annotations as bridging the information contributed by different representational modalities in graph-text documents, but the post-test recall scores show that the shifts between the different representational modalities result in inhibitory effects.

Conclusions

Compared to research on eye movement control in reading, there are few studies investigating eye movement characteristics in multimodal documents. Most of these studies investigate multimodal documents such as text and pictorial or diagrammatic illustrations presented in static or dynamic displays. Furthermore, due to the abundance of different types of illustrations, the illustration-type-specific differences in multimodal processing have been seldom considered in detail in educational psychology research.^x

Understanding the nature of the interaction between parts of a document (or hypertext interface) with different representational modalities is a complex task with respect to internal characteristics of the information content and to the type of representations used other than text (i.e., pictorial illustrations, diagrams, information graphics, etc.). Information graphics and pictorial illustrations have different characteristics that are relevant to multimodal comprehension. Pictorial illustrations—also called “representational pictures” by Alesandrini (1984)—can informally be characterized by their visual resemblance to the objects they stand for. Pictorial illustrations and their referents have spatially similar layouts (i.e., iconic similarity), which, for example, in the case of photographs guarantees an optically veridical mapping from the visual-world object to the external representation.^{xi} Even line drawings must have a high degree of systematic resemblance to the depicted entity to be able to function successfully in a multimodal document. Furthermore, pic-

torial illustrations do not possess an internal syntax in the sense of representational formats as discussed by Kosslyn (1980, p. 31). On the other hand, information graphics are representational artifacts that possess internal syntactic structures. The syntactic analysis of a graph is fundamental for succeeding processes of semantic and pragmatic analyses in graph comprehension (Kosslyn, 1989; Pinker, 1990). From the perspective of applicability to cognitive and computational architectures developed in Artificial Intelligence, the formal characteristics of information graphics can be seen as an advantage over pictorial illustrations.

Other than the assumptions behind the methodology used in this study—the pre-requisite knowledge of participants about graph interpretation and the complexity of inhibitory as well as facilitatory effects of multimodal comprehension—more specific limitations of this study can be summarized as follows. Although using modified real documents with several paragraphs published in electronic or printed media has the advantage of the ecological validity of the study, it also introduced reduced experimental control and effect size, as well as high standard deviations in the results. A similar situation occurs with the self-paced reading of the participants. Although reading under time pressure increases control, it reduces the ecological validity of the experiment since people rarely read such articles in newspapers under time pressure in natural settings. In this study, we did not use a scalable metric for participants' prior knowledge about graph comprehension or prior knowledge about the domain referred to by graphical constituents (namely, stock market graphs) except for an explicit yes/no statement asked in the demographic questionnaire. A more systematic analysis requires better investigation of these parameters. In addition, even though researchers have proposed classifications for functions of pictorial illustrations in multimodal documents, mostly from instructional design perspective (Levin, Anglin, & Carney, 1987; Carney & Levin, 2002, Ainsworth, 2006), as of our knowledge, there is no up-to-date systematic analysis on the role of information graphics in multimodal documents. A more systematic analysis should take into account these specifications. Furthermore, a broader set of post-test questions (e.g., transfer questions) from the perspective of instructional design would reveal more detailed information about facilitatory and inhibitory effects of multimodal representations in learning, as well as in comprehension.

This study has an exploratory character, and it is limited in generalizability, as are most of the studies in the area. For generalizability of the results, research in other domains that employ information graphics frequently (e.g., meteorology) is needed. Despite its limitations, we see our study as an initial step for further investigation of multimodal comprehension of text and information graphics. As of our knowledge, contra to the models for eye movement control in reading, there is no model for eye movement control in text-graphics documents, despite their potential for formal description. The formal characteristics of information graphics articulates a path to the development of cognitive architectures for multimodal comprehension as well as computational models for eye movement control in multimodal documents. These models can be used for further design and development of multimodal output generation systems of graph comprehension, especially for people with visually impairments.

Acknowledgements

This paper is based on the study presented in the poster session of the 14th European Conference on Eye Movements, August 19-23, 2007, in Potsdam, Germany. The research reported in this paper has been partially supported by DFG (German Science Foundation) in ITRG 1247 “Cross-modal Interaction in Natural and Artificial Cognitive Systems” (CINACS). The experiments in this study were conducted at the Human-Computer Interaction Research and Application Laboratory at Middle East Technical University; thanks for their kind support. We also thank two anonymous reviewers for their helpful comments.

Appendix

A.1. Mean Fixation Counts. The mean fixation counts on the AOIs under the three conditions are shown in Table A.1.

Table A.1.
Mean fixation counts on the specified AOIs.

	Annotated-graph Condition	Non-annotated-graph Condition	Text-only Condition
Main Title	14.86 (8.01)	12.00 (5.47)	14.33 (8.10)
Paragraphs	164.86 (90.49)	169.52 (74.37)	158.43 (62.27)
Graph Line	1.38 (1.53)	3.62 (4.15)	-
X Axis	0.62 (1.02)	1.95 (2.50)	-
Y Axis	0.33 (0.66)	1.24 (1.97)	-
Graph Title	2.05 (2.22)	-	-
Annotation Text	3.38 (5.20)	-	-
Annotation Icons	1.52 (3.06)	-	-

Note. The numbers in parentheses show standard deviation.

A.2. Mean Gaze Times. The mean gaze times on the AOIs under the three conditions are shown in Table A.2.

Table A.2.
Mean gaze times on the specified AOIs.

	Annotated-graph Condition	Non-annotated-graph Condition	Text-only Condition
Main Title	4,125 (2,263)	3,205 (1,749)	3,883 (2,164)
Paragraphs	48,313 (30,102)	52,162 (30,054)	46,943 (26,212)
Graph Line	329 (396)	983 (1,145)	-
X Axis	172 (337)	757 (1,149)	-

Y Axis	110 (244)	487 (848)	-
Graph Title	755 (1,139)	-	-
Annotation Text	1,168 (1,983)	-	-
Annotation Icons	529 (1,709)	-	-

Note. The numbers in parentheses show standard deviation. All numbers are in ms.

A.3. Post-test Scores. The post-test scores for the nine post-test questions are shown in Table A.3.

Table A.3.
Post-test scores.

	Annotated-graph Condition	Non-annotated-graph Condition	Text-only Condition
Question 1	-0.75 (0.72)	0.50 (0.93)	-0.40 (0.97)
Question 2	1.00 (0.00)	0.64 (0.81)	0.00 (1.05)
Question 3	0.78 (0.67)	0.00 (1.10)	0.43 (0.98)
Question 4	-0.60 (0.84)	-0.75 (0.71)	-1.00 (0.00)
Question 5	0.40 (0.97)	0.50 (0.93)	0.33 (1.00)
Question 6	0.11 (1.05)	0.20 (1.03)	0.50 (0.93)
Question 7	0.33 (1.03)	1.00 (0.00)	-0.50 (1.00)
Question 8	-0.56 (0.77)	1.00 (0.00)	0.25 (1.04)
Question 9	0.11 (1.05)	0.09 (1.04)	0.33 (1.00)

Note. The numbers in parentheses show standard deviation.

References

- Acarturk, C., Habel, C., & Cagiltay, K. (2008). Multimodal comprehension of graphics with textual annotations: The role of graphical means relating annotations and graph lines. In J. Howse, J. Lee, & G. Stapleton (Eds.), *Diagrammatic representation and inference: Lecture notes in computer science* (Vol. 5223, pp. 335-343). Berlin: Springer.
- Ainsworth, S. E. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction, 16*, 183-198.
- Alesandrini, K. L. (1984). Pictures and adult learning. *Instructional Science, 13*, 63-77.
- Arsenault, D. J., Smith, L. D., & Beauchamp, E. A. (2006). Visual inscriptions in the scientific hierarchy. *Science Communication, 27*(3), 376-428.
- Berger, C. R. (2005). Slippery slopes to apprehension: Rationality and graphical depictions of increasingly threatening trends. *Communication Research, 32*, 3-28.
- Bernard, R. M. (1990). Using extended captions to improve learning from instructional illustrations. *British Journal of Educational Technology, 21*, 212-225.
- Bernsen, N. O. (1994). Foundations of multimodal representations: A taxonomy of representational modalities. *Interacting with Computers, 6*(4), 347-371.
- Bertin, J. (1983). *Semiology of graphics: Diagrams, networks, maps* (W. J. Berg, Trans.). Madison: University of Wisconsin Press.
- Carney, R., & Levin, J. (2002). Pictorial illustrations still improve students' learning from text. *Educational Psychology Review, 14*(1), 5-26.
- Carroll, P. J., Young, R. J., & Guertin, M. S. (1992). Visual analysis of cartoons: A view from the far side. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 444-461). New York: Springer.
- Chabris, C. F., & Kosslyn, S. M. (2005). Representational correspondence as a basic principle of diagram design. In S.-O. Tergan & T. Keller (Eds.), *Knowledge and information visualization: Searching for synergies: Lecture notes in computer science* (Vol. 3426, pp. 36-57). Berlin: Springer.
- Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology, 62*, 233-246.
- Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association, 77*, 541-547.
- Glenberg, A. M., Meyer, M., & Lindem, K. (1987). Mental models contribute to foreground during text comprehension. *Journal of Memory and Language, 26*, 69-83.
- Graham, J. L. (1937). Illusory trends in the observation of bar graphs. *Journal of Experimental Psychology, 6*(20), 597-608.
- Habel, C., & Acarturk, C. (2007). On reciprocal improvement in multimodal generation: Co-reference by text and information graphics. In I. van der Sluis, M. Theune, E. Reiter, & E. Krahmer (Eds.), *Proceedings of the Workshop on Multimodal Output Generation: MOG 2007* (pp. 69-80). United Kingdom: University of Aberdeen.
- Hegarty, M., & Just, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language, 32*, 717-742.
- Hegarty, M., Carpenter, P. A., & Just, M. A. (1996). Diagrams in the comprehension of scientific texts. In R. Barr, M. L. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of Reading Research* (Vol. 2, pp. 641-688). Mahwah, NJ: Erlbaum.
- Henderson, J. M., & Ferreira, F. (Eds.). (2004). *The interface of language, vision, and action: Eye movements and the visual world*. New York: Psychology Press.
- Kosslyn, S. M. (1980). *Image and mind*. Cambridge, MA: Harvard University Press.
- Kosslyn, S. (1989). Understanding charts and graphs. *Applied Cognitive Psychology, 3*, 185-225.
- Kosslyn, S. (1994). *Elements of graph design*. New York: W. H. Freeman.
- Kukich, K. (1983). *Design of a knowledge-based report generator*. Paper presented at the 21st Annual Meeting of the Association for Computational Linguistics, Cambridge, MA.
- Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11*, 65-99.

- Levin, J. R., Anglin, G. J., & Carney, R. N. (1987). On empirically validating functions of pictures in prose. In D. M. Willows & H. A. Houghton (Eds.), *The Psychology of Illustration: I. Basic Research* (pp. 51–85). New York: Springer.
- Mautone, P. D., & Mayer, R. E. (2007). Cognitive aids for guiding graph comprehension. *Journal of Educational Psychology*, 99(3), 640-652.
- Mayer, R. E. (2001). *Multimedia learning*. Cambridge, MA: Cambridge University Press.
- Mayer, R. E. (Ed.). (2005). *The Cambridge handbook of multimedia learning*. Cambridge, MA: Cambridge University Press.
- Newcombe, N. S., & Learmonth, A. (2005). Development of spatial competence. In P. Shah & A. Miyake (Eds.), *Handbook of visuospatial thinking* (pp. 213-256). Cambridge, MA: Cambridge University Press.
- Peebles, D. J., & Cheng, P. C.-H. (2002). Extending task analytic models of graph-based reasoning: A cognitive model of problem solving with Cartesian graphs in ACT-R/PM. *Cognitive Systems Research*, 3, 77-86.
- Peeck, J. (1987). The role of illustrations in processing and remembering illustrated text. In D. M. Willows & H. A. Houghton (Eds.), *The psychology of illustration* (Vol. 1, pp. 115-151). New York: Springer-Verlag.
- Pinker, S. (1990). A theory of graph comprehension. In R. Freedle (Ed.), *Artificial intelligence and the future of testing* (pp. 73-126). Hillsdale, NJ: Erlbaum.
- Preim, B., Michel, R., Hartmann, K., & Strothotte, T. (1998). *Figure captions in visual interfaces*. Proceedings of the Working Conference on Advanced Visual Interfaces, L'Aquila, Italy.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372-422.
- Rayner, K., Rotello, C. M., Stewart, A. J., Keir, J., & Duffy, S. A. (2001). Integrating text and pictorial information: Eye movements when looking at print advertisements. *Journal of Experimental Psychology: Applied*, 7, 219-226.
- Scaife, M., & Rogers, Y. (1996). External cognition: How do graphical representations work? *International Journal of Human Computer Studies*, 45, 185-213.
- Shah, P., Freedman, E., & Vekiri, I. (2005). The comprehension of quantitative information in graphical displays. In *The Cambridge Handbook of Visuospatial Thinking* (pp. 426-476). Cambridge, MA: Cambridge University Press.
- Sigurd, B. (1995). *STOCKTEXT – Automatic generation of stockmarket reports*. Unpublished manuscript, Department of Linguistics, Lund University, Lund, Sweden.
- Simkin, D., & Hastie, R. (1987). An information processing analysis of graph perception. *Journal of the American Statistical Association*, 82(398), 454-465.
- Spence, I. (2005). No humble pie: The origins and usage of a statistical chart. *Journal of Educational and Behavioral Statistics*, 30, 353-368.
- Tabachneck-Schijf, H. J. M., Leonardo, A. M., & Simon, H. A. (1997). CaMeRa: A computational model of multiple representations. *Cognitive Science*, 21, 305-350.
- Tufte, E. R. (1983). *The visual display of quantitative information*. Cheshire CT: Graphic Press.
- Underwood, G., Jebbett, L., & Roberts, K. (2004). Inspecting pictures for information to verify a sentence: Eye movements in general encoding and in focused search. *The Quarterly Journal of Experimental Psychology*, 57A, 165-182.
- Wainer, H. (1984). How to display data badly. *American Statistician* 38(2), 137-147.
- Wainer, H., & Velleman, P. F. (2001). Statistical graphics: Mapping the pathways of science. *Annual Review of Psychology*, 52, 305-335.
- Washburne, J. N. (1927). An experimental study of various graphic, tabular and textual methods of presenting quantitative material. *Journal of Educational Psychology*, 18, 361-376, 465-476.
- Winn, B. (1987). Charts, graphs, and diagrams in educational materials. In D. M. Willows & H. A. Houghton (Eds.), *The psychology of illustration* (Vol. 1, pp. 152-198). New York: Springer-Verlag.
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123(2), 162-185.

ⁱ In this paper, we use the term “modality” as shorthand for “representational modality” (see Bernsen, 1994). Accordingly, the term “multimodal document” is used for “a document that includes more than one representational modality,” such as a document with text and illustrations. “Multimodality” concerning various “sensory modalities” is beyond the scope of this study, since the processing of multimodal documents containing text and diagrams—and presented on paper or a computer screen—is based on the visual sensory modality only.

ⁱⁱ “Information graphics,” as used in the present paper, subsumes graphs, diagrams, and charts as characterized by Kosslyn (1989). Furthermore, information graphics seems to correspond largely to “diagrams” as focused on in Chabris and Kosslyn (2005) and “arbitrary pictures” discussed by Alesandrini (1984). In the following, we use the term “graph” as shorthand for “information graphics,” or statistical graphs representing relations between abstract variables.

ⁱⁱⁱ There is no clear-cut border between perception and comprehension of graphs. What we mean by “comprehension” is task-dependent processes beyond perception such as problem solving, which are also discussed extensively in Artificial Intelligence.

^{iv} The integration task partially corresponds to comprehension of language during perception of the visual world (Henderson & Ferreira, 2004). However, it differs from the “visual world” case since graphs are conventionalized representations, specified by syntactic and semantic principles.

^v Ainsworth (2006) has a similar line of argumentation independent of Chabris and Kosslyn (2005) with respect to learning with multiple representations.

^{vi} We use the term “annotation” to refer to verbal elements (i.e., annotation labels or annotation text) that are connected to specific parts of a graph (e.g., the top or the maximum point in a line graph) via connecting symbols (e.g., a symbol such as a circle attached to a thin straight line). We use the term “foregrounding”, within the framework of mental model approaches, to mean activation of certain elements (tokens) and retainment of this information foregrounded during comprehension (see Zwaan & Radvansky, 1998; Glenberg, Meyer, & Lindem 1987).

^{vii} It should be noted that the term “annotation” has different meanings in literature. For example, an author’s personal marks on a working paper or marks for communicative purposes in co-authoring activities are also called annotations. Although both are used for communicative purposes, our use of the term “annotation” is different than those used in co-authoring activities. Annotations are also used in a different way than “captions” (i.e. verbal descriptions of the figure, located generally below the figure and starting with a phrase like Figure 1.). Although both enrich an illustration in a way that labels and legends cannot achieve (Preim, Michel, Hartmann, & Strothotte, 1998), figure captions refer to a whole illustration, while annotations refer to the parts of a graph. In addition, we use static annotations in static documents and hypertext rather than dynamic annotations in animations and video files. Last of all, our use of

the term “annotation” has a more descriptive than instructive aspect in Bernard’s (1990) terminology.

^{viii} The original articles were: ©The New York Times. Dow Jones Index Hits a New High, Retracing Losses, by Vikas Bajaj, published on October 4, 2006 (translated to Turkish by the experimenter); ©Radikal. Endeks 30 bin sınırını aştı [The Index exceeded the 30 thousand limit (Translation by the experimenter)], published on August 5, 2005; ©Sabah. Borsa, Mayıs 2006’ya geri döndü [The stock market retraced to May 2006 (Translation by the experimenter)], published on January, 25, 2007.

^{ix} Hegarty and Just (1993) have included gazes with a duration of more than 250 ms on the text as well as gazes with a duration of more than 100 ms on the diagram components in their analysis of text-pulley diagram documents. On the other hand, other researchers include gazes with fixation durations less than these values for other types of stimuli. For example, Underwood, Jebbett, and Roberts (2004) include gazes with a fixation duration above 60 ms in the analysis of real world photographs. The time to encode sufficient information for object identification is about 50-75 ms, the value being found by eye-contingent masking technique (Rayner, 1998). In this study, for the sake of being conservative with respect to text-graph co-reference constructions specifically, we included gazes above 100 ms in the analysis. The outlier data was eliminated by excluding data above or below mean plus/minus two standard deviations.

^x This is astonishing, since thorough discussions about types of illustrations started in the 1980s (Alesandrini, 1984; Peeck, 1987; Winn, 1987).

^{xi} The veridicality commitments of drawings are less strict, but—as Chabris and Kosslyn (2005) argue for the sub-case of caricatures—this property of drawings can facilitate processing.