Author: Ferres, Leo; Lindgaard, Gitte; Sumegi, Livia; Tsuji, Bruce

Title: Evaluating a tool for improving accessibility to charts and graphs

Article number: 28
Year: 2013
Journal: ACM Transactions on Computer-Human Interaction

Volume: 20
Issue: 5
URL: http://doi.org/10.1145/2533682.2533683

Evaluating a tool for improving accessibility to charts and graphs

LEO FERRES
Universidad de Concepción, Concepción, Chile
and
GITTE LINDGAARD
Swinburne University of Technology, Melbourne, Australia
& Carleton University, Ottawa, Canada
and
LIVIA SUMEGI, and BRUCE TSUJI
Carleton University, Ottawa, Canada

This paper reports a case study of the iterative design and evaluation of a natural language-driven assistive technology, **iGraph**-LITE, providing people who are blind access to line graphs. Two laboratory-based usability studies involving blind and sighted people are presented with a discussion of the ensuing implementation of changes. Blind participants were found to adopt different graph interrogation strategies than sighted participants. A small field study is then reported in which a blind user who works with graphs took part to determine the degree to which the **iGraph**-LITE commands would meet the needs of blind graph experts. The final study invited sighted graph experts and novices to visually inspect and explain a set of line graphs comparable to those used in the usability studies. It aimed to highlight the concepts and the range of words sighted people use, to ascertain the appropriateness of the **iGraph**-LITE lexicon. A set of preliminary guidelines is presented.

Categories and Subject Descriptors: H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*Auditory (speech) feedback, Evaluation/methodology, User centered design*

General Terms: Human Factors
Additional Key Words and Phrases: Accessibility (Blind and Visually Impaired), Natural language Interaction, Statistical Graphs

1. MOTIVATION AND BACKGROUND

In order to take advantage of the human ability to identify visual patterns [Bertin 1983; Larkin and Simon 1987], numerical data-sets are often represented visually in the form of graphs and charts [Tufte 2001]. These graphical aids are a common means of representing numerical data for both exploration of new hypotheses, in the

case of newly collected experimental data for instance, and mass consumption, as in the case of a line graph showing the "Increase of soybean production between 2000-2013" in a non-specialized newspaper. A large number of graphs are published every year. For instance, Statistics Canada bulletin[1], "The Daily"[2] publishes on average two graphs per day on a variety of topics (employment, GDP, etc.). This amounts to about five *hundred* graphs per year for just one of the tens of publications of this particular agency. Given the large number of newspapers, business reports, science textbooks, and Internet sites, it becomes clear that an enormous number of graphs are published every year. This overwhelming preference to visual representations gives some credibility to the saying that "a picture is worth a thousand words".

This stands in stark contrast with the fact that a majority of the almost 88,000 *reported* blind people living in Canada in 2006[3] (993,766 in the US[4] in 2004) do not yet have cost-effective, straightforward access to visually-encoded information. For people who are blind, then, the wealth of visual information encoded in graphs is likely lost unless tools using another sensory modality, such as sound, touch or a combination thereof, are developed to compensate for vision. The net effect is that the blind community is excluded from a vast repository of useful information, jobs, and their chances of succeeding in any area of education that deals with the exploration and manipulation of quantitative data.

Fortunately, excellent research into non-visual interfaces to graphs and charts has been conducted in the past decade to help alleviate this problem. These non-visual interfaces to graphs generally fall into four loosely defined categories: **haptic interfaces** (including tactile non-digital techniques such as embossed/raised paper) [Yu et al. 2001; Yu and Brewster 2002; Riedel 2001; Wall and Brewster 2006], **sonification** [Flowers et al. 2005; Brown and Brewster 2003; Cohen et al. 2005; McGookin and Brewster 2006; Ben-Tal et al. 2002], **natural language interfaces** (NLI, for short) [Kurze 1995; Ferres et al. 2007; Abu Doush et al. 2009; Demir et al. 2010; Ferres et al. 2010], and **hybrid interfaces** [Yu and Brewster 2003; Kennel 1996; Grabowski and Barner 1998]. Each of these use touch, sound (linguistic and non-linguistic), or both to convey visual information.

Raised/embossed paper graphs and pin-boards are the most common technique used in schools to allow for non-visual browsing and construction of graphs. Tactile graphs support rapid, non-sequential data browsing, and graph readers can use both hands to explore the graph. Unfortunately, cues that are visually discriminative are not always easy to discern using touch, due to limitations to tactile acuity. In addition, tangible data representations are static, cumbersome and expensive to print in large quantities. Raised paper graphs are difficult to store since embossing fades away after time and use, and a sighted person with the right skills is needed to prepare them. Certain types of graphs are difficult to produce using a pin-board approach (e.g., stacked-graphs), and again they cannot be easily stored.

According to Kramer et al. [2010], interest in augmenting purely tactile displays with sound for presenting graphical information is increasing. Studies have shown

---

[1]http://www.statcan.gc.ca
[2]http://www.statcan.gc.ca/dai-quo/index-eng.htm
[3]http://www.statcan.gc.ca/pub/89-628-x/2009013/fs-fi/fs-fi-eng.htm
[4]http://www.center4research.org/blind0204.html

that most people can understand trends, clustering, correlations, and other simple statistical features of a data set just as well by listening to a musical rendition of it as they could by reading a graph. Sonification is perhaps the most well-known and best evaluated technology for communicating information about graphs non-visually in the Human-Computer Interaction literature [Flowers et al. 1997; McGookin and Brewster 2006]. However, Kildal and Brewster [2006] claim that current accessibility tools based on the sonification of data-sets do not support mechanisms for summarizing information, highlighting salient features of the data and are difficult to work with for graphs other than simple line graphs (see also Kildal and Brewster [2007] and Ferres et al. [2007]).

Studies evaluating hybrid systems are much harder to come by in the literature. In their study using sonification, haptics and simple speech, Kildal and Brewster [2006] claim that graph users were able to rely on their tactile sense to judge the position of a stylus in relation to the borders of a tablet on which tabular data was presented (notice that tabular data omit important visual information like graph type, which can be meaningful). A click was emitted when the stylus was moved from one cell to the next, and numbers in the cells were spoken. The authors claim that the cognitive workload was lower in the sonification mode than in the speech mode, but they do not reveal how workload was assessed, nor whether it was the same for graphs.

In contrast, in another study in which participants explored bar graphs to answer certain questions, Wall and Brewster [2006] used a mixture of speech and sonification feedback together with two tactile 4×4 pin-arrays on a VTPlayer tactile mouse and a stylus. Data sonification occurred in response to stylus button clicks, with a MIDI piano note playing the proportional value of data for each bar. This allowed the user to obtain a quick auditory overview of the graph that, in turn, could be used to direct their querying of specific data values further. The authors conclude that the point-and-click metaphor is an improvement over the sequential nature of speech.

The touch-sensitive Talking Tactile Tablet (T3) of Wells and Landau [2003] provides speech output. Interestingly, despite the obvious disadvantage of speech –auditory perception is serial and hence imposes a greater cognitive load on a person's processing capability– all participants much preferred the speech feedback, using sonification only rarely. Speech was also preferred (and led to considerably shorter task-completion times) in Evreinova et al. [2008]'s within-subject study. In their report, sighted users explored simple hidden 2D line drawings with Directional-Predictive Sound (DPS) and Directional-Predictive Vibrations (DPV) devices providing the feedback. The experiential advantage of processing spoken words relative to sonification and reliance on tactile stimuli may well have overshadowed the potential benefits these alternatives have to offer. It is plausible to argue that a mixture of sonification and speech output could eventually provide a powerful platform for supporting the processing of visual information. Brown and Brewster [2003]'s binaural presentation of two-line graphs whereby data contained in each of the two series are separately presented to the left and right ears suggests that this is a possibility worth exploring.

The purpose of our work, however, was to empirically evaluate the degree to which

an NLI may facilitate the auditory processing of visual graph data. We have chosen the language modality because it has been used to greater success in summarizing and communicating information about graphs than sonification and haptics, but evaluation as an Assistive Technology with blind computer users has been lacking. There are a number of systems that are relevant for our present purposes and that have inspired **iGraph**-LITE to a large extent. Most of this work falls into one of three categories: summarization of time-series data, multimodal presentation of text and graphs, and the automatic identification of a graph's communicative purpose. The latter developed as an assistive technology, and is thus of particular interest here.

The idea of using language to summarize and communicate information about numerical data dates back at least as far as the early 1980s [Kukich 1983]. To date, data-to-text (D2T) systems have been successfully fielded in areas as diverse as turbine sensor summarization [Yu et al. 2003], weather forecasting [Sripada et al. 2003; Turner et al. 2006; Reiter et al. 2005; Sripada et al. 2004], medical technologies [Portet et al. 2007], stock exchange summarization [Dale 2002], and dive data summarization [Sripada and Gao 2007] among others. NLG systems that provide summarizations of large time-series databases such as these are perhaps one of the most thoroughly studied applications of NLG technology with generally agreed architectures based on a "pipeline"-based processing of different intermediate representations (See [Reiter and Dale 2000], [Reiter 2007] and [Ferres et al. 2007] for a good overview of current D2T systems and [Kerpedjiev 1995] for a review of earlier systems). However, one of the problems with current adoption of D2T NLG architectures as Assistive Technologies (ATs), and possibly in general, is that they are not interactive and produce a single static text for an entire description all at once. Such descriptions may be long and hard for users to parse and comprehend. Furthermore, in the case of assistive technologies just summarizing the graph is not enough [Ferres et al. 2006], and an interface is required to facilitate exploration of the information space.

Systems in the category of multimodal presentation of text and graphs take advantage of the different, and often complementary, properties of graphical and linguistic media to produce one single *multimodal* document. A few systems are noteworthy in this category: POSTGRAPHE [Fasciano and Lapalme 1996; 2000], the CAPTION GENERATION SYSTEM [Mittal et al. 1998] and the different versions of AUTOBRIEF [Kerpedjiev et al. 1997], summarized in [Green et al. 2004]. All these systems have contributed greatly to our understanding that graphs are complex representational objects [Roth et al. 1991], that authors' intentions play a major role in the graph generation process [Corio and Lapalme 1999], and the importance of the synergy between texts and graphs [Corio and Lapalme 1998; Law et al. 2005]. Genetically, the NLG side of these systems are also captured by the high-level architecture of [Reiter 2007]. These systems generate *static* reports (or captions), but part of the input to these systems consists of the author's intentions at the time of generating reports containing graphs and texts, which allows the technology to provide some interpretation of the data. As ATs, these systems are more informative than those merely summarizing numerical data mainly because the generated texts explain and describe the generated graphs, however these are still *are* sum-

marizations and the quality of the interpretation is heavily dependent on the graph author. This approach is fine for these systems because their main audience is sighted people, but they fall short as ATs providing accessibility to graphs. Both the systems described here and those described in the previous paragraph also have the drawback that the texts they produce would only make sense if the user were looking at the graph while reading the text. The objective of **iGraph**-LITE is to provide access to graphs when users *cannot* look at these graph.

Systems that automatically identify a graphs communicative purpose, which in general are designed specifically to be ATs, are of particular interest here. To the best of our knowledge, only two previous systems, seem to exist that aim to fulfill this AT goal: the system described in [Kurze 1995][5], and the SIGHT system [Chester and Elzer 2005; Elzer et al. 2007]. Two important contributions of the SIGHT system is that it takes graphs as raster images (jpg, gif) as input and tries to *automatically* extract the communicative purpose of the graph using a Bayesian network. Previous systems either suggest an appropriate graph given the author's intention explicitly, or they suggest the intention itself, given the available data. We believe the most interesting contribution of this work is that of making the system available through (JAWS[6]), a widely adopted screen reader, which adds a small element of interactivity to the system. **iGraph**-LITE itself was inspired by screen-reading interfaces. Unfortunately, SIGHT suffers from the same drawback as other systems. In spite of its potential for being interactive, it is not: SIGHT still generates a static text describing the intention of the graph's author, accessed by a keyboard command in JAWS. Our hypothesis is that the system could have been made much more interactive by accessing other messages left out of the initial summarization of purpose. We hope to provide a system that extends SIGHT in this respect, building a more interactive AT NLG system while still respecting the theoretical underpinnings discussed in [Reiter 2007].

Finally, perhaps a word should be said about screen-readers. These technologies are the most widely adopted interface to digital information by blind and visually-impaired people. Their interactive paradigm is simple, but efficient: key combinations access different parts of the informational space, such as a website, with messages read aloud by a TTS engine. The problem with screen-readers as they exist today is that they perform little computation on what they read, instead mostly reading just the raw text from text-boxes, buttons, menus, etc.

Once again, our purpose in this paper is to evaluate a complex NLI to graphs that makes use a history of natural language description of graphs. If our results are positive, then blind and visually impaired users will be able use **iGraph**-LITE with standard office equipment and not require additional devices, such as embossing printers, or assistance from others to interpret line graphs. It could offer user independence as well as the ability to obtain any information contained in the graph, including each data point within a complex graph, which is not possible with sonification.

This paper thus reports the careful design and user-based evaluation of **iGraph**-LITE, a novel Assistive Technology (AT) designed to help blind and visually im-

---

[5]At least as reported in [Elzer et al. 2007]. We have not been able to locate a copy of this paper.
[6]`http://www.freedomscientific.com`

paired individuals interact with graphical representations of line graphs using natural language through key commands and a Text-To-Speech (TTS) engine. The remainder of the paper is organized as follows. First, we present a preliminary study designed to address the kinds of questions people who are blind may ask of graphs, and the vocabulary that trained statisticians use spontaneously to explain simple graphs, as well as the words visually impaired users will understand. That study formed the basis for **iGraph**-LITE's lexicon. Next, we describe the general architecture of **iGraph**-LITE, followed by the implementation of its navigation module. This is followed by two formative usability studies testing the original, and then the improved, version of the tool that incorporated the lessons learned from the first usability study. A small field test of the tool is then presented that aimed to determine how well **iGraph**-LITE would match the needs of blind users working with graphs in their job. The final study tested the vocabulary that sighted graph experts and novices use when describing and explaining line graphs to someone else; none of these participants were formally trained statisticians. A set of recommendations based on lessons learned in all the studies are described in the final discussion and conclusions.

To the best of our knowledge, this is the most comprehensive empirical study to date of people who are blind interacting with graphs through natural language. The stimuli and materials used, including the **iGraph**-LITE navigator versions that were current at the time of the studies and the data-sets, are freely available in the hopes that other researchers might find them useful for replication and extension[7].

2.   STUDY 1: QUERYING GRAPHICAL INFORMATION

Building on earlier work [Ferres et al. 2006] that provided initial information for **iGraph**-LITE, Study 1 was designed to obtain information about specific constraints that blind and visually impaired people face when reading graphs, as well as explore the vocabulary statisticians use when describing simple graphs. Shneiderman [1996] proposed a basic principle for presenting visual information in his "Visual Information Seeking Mantra", according to which an overview should be presented first, followed by the ability to zoom, filter, and then allow the user to request details-on-demand. Although Shneiderman's Mantra was conceptualized to allow visual search of data varying greatly in complexity, we wanted to understand if the same principle would hold for visual information presented aurally to blind and visually impaired users. The study therefore explored the kinds of questions that users who are blind or visually impaired may ask of a graph, enabling them to comprehend its meaning and describe it to others, as well as taking note of the vocabulary that statisticians may use to describe and answer questions about graphs. It also explored the sequence in which information about a graph is requested to learn more about the process by which people construct a mental representation of simple line- and bar-graphs that they cannot see. Data were gathered in two sessions involving a sighted statistician describing and/or answering questions about a graph, as described below.

---

[7]See http://www.inf.udec.cl:/~leo/assets2010.html

2.1 Method

2.1.1 *Participants.* Two sighted statisticians and three legally, non-congenitally blind participants (2 male, 1 female) with limited or no residual vision were recruited via word of mouth. All five participants were Statistics Canada employees who encountered graphs regularly at work. To read graphs, the visually impaired participants relied either on screen readers rendering the raw data tables from which the graph was generated, decomposing the numbers in the accompanying data tables, or else they asked colleagues for help. One of the sighted statisticians took part in the first session and the other in the second session in order to obtain a wide, yet precise, vocabulary as possible. Two researchers were present: one acted as a facilitator, the other observed and took notes throughout. The statisticians were editors of Statistics Canada's The Daily bulletin; the three visually impaired participants were all avid readers of The Daily. All volunteered their services, receiving no compensation for participation in the study. The two sessions took place in the Human-Oriented Technology Lab (HOTLab) at Carleton University, Ottawa, Canada. Each session took approximately 1.5 hours.

2.1.2 *Materials.* The Daily presents mostly textual information on current socioeconomic trends, but it typically also includes at least one graph embedded in the text. Graphs were selected from The Daily to secure a reasonably varied stimulus set and without consideration of the topics and dates of publication of the graphs. We made an effort to locate an equal number of graphs meeting the following criteria:

—Line- or bar graphs

—One or two data-sets

—Positive values only or both positive and negative values

—Trend line or no trend line

Thus, the ideal, completely balanced data-set should have contained a total of 16 graphs. However, it was impossible to find a line graph with both a trend line and positive as well as negative values. Thus, the final set of 13 graphs therefore contained six line graphs and seven bar graphs meeting the four criteria above.

2.1.3 *Design.* Session 1 was designed to learn more about the kinds of questions visually impaired people might ask of a graph embedded in text and the sequence in which these may be asked. The sequence of questions offers insight into the process by which people construct a mental representation. This understanding is invaluable for maximizing the efficacy of **iGraph**-Lite to support graph comprehension in situations in which the graph reader cannot see the graph. The same people participated in Session 2 with the exception of the statistician who was replaced by a colleague. Both sessions were audio- and video recorded with participants' permission.

2.1.4 *Procedure.* In Session 1, the statistician answered questions pertaining to six of the 13 stimulus graphs (3 line- and 3 bar graphs) one at a time to answer questions about each graph, alternating line- and bar graphs and starting with a line graph. Participants were called upon one at a time to ask a question. If the

person called upon had no question, the next person was called upon until there were no further questions. The facilitator then asked a visually impaired participant to describe their understanding of the graph to the group. If that person felt they could not describe it, the next participant was asked to attempt a description. The statistician then provided feedback, correcting the description if necessary. This process continued until all six graphs had been presented and described.

In Session 2, the other sighted statistician was given the seven remaining stimulus graphs (3 line- and 4 bar graphs), and asked to describe each as concisely as possible. Participants were asked not to interrupt the statistician or ask questions to allow the statistician to give an undisturbed account of the graph in her own words. None of the materials, stimuli, questions, or answers from Session 1 were made available. Upon completion of the description, each visually impaired participant was allowed to ask one question in turn to ensure that everyone was given an opportunity to do so. This continued until there were no further questions. The statistician's description revealed words that were easy to understand, as well as words that hampered recipients' understanding of the graph.

2.2   Results and Discussion

2.2.1   *Session 1.* A content analysis of the video records was performed using ANVIL [Kipp 2001]. Table I below shows the number of questions asked for each graph. Although it seems that many more questions were asked of the first two line graphs, the analysis revealed that six of the 13 questions for Line Graph 1 pertained to the X- and Y-axes (first value of each, what X and Y mean, scale on each, etc.). For the second line graph, four questions concerned the axes; thereafter, the statistician gave all axes-related information in response to a single question. Removing these questions from Line Graphs 1 and 2 then yielded a total of seven questions for each, rendering the number of questions very similar for all graphs regardless of graph type, number of data-sets, positive or positive/negative values, and the presence of a trend line.

Order	Line graph	N questions	Bar graph	N questions
1	Two lines, no trend line, positive values only	13 (7)	single data type, positive and negative values (13 bars)	5
2	Single line, trend line, positive values only	11 (7)	single data type, positive values only (17 bars)	7
3	Single line, no trend line, positive and negative values	5	two data types, positive values only (4 bars)	5

Table I. Number of questions asked for each graph and each type of graph by the order in which graphs were presented

Across the six graphs, some 13 different kinds of questions were asked, with a wider variety for line graphs (12/13) than for bar graphs (8/12). The first eight types of questions shown in Table II were common to both graph types and were about graph and display characteristics. The remaining five questions concerned

the global (questions 9-11 inclusive) and local values (12-13) of data contained in the graph.

	N	Type of question	Line	Bar
	1	Title	3	3
	2	Axes	3	2
	3	Type of graph	2	3
Specific graph characteristics	4	Purpose of graph	1	2
	5	N lines/bars	2	0
	6	Line color	1	0
	7	Portrait or landscape	1	0
	8	Scales	1	0
	9	Trend, general	1	2
	10	Trend, each line	1	0
Global graph characteristics	11	Any negative values	0	2
	12	Highest/lowest point	3	2
	13	First/last value	3	0

Table II.   The number of times each type of question was asked for each graph type

The title was requested first for five of the six graphs, and axes-related questions were invariably among the first three. Information about the graph type was requested in all but the first bar graph, which participants implicitly assumed was a line graph as was clear from a question about the number of lines in the graph. The data suggest that participants asked questions in a consistent, systematic sequence and that the graph title should be given first, followed directly by mention of the graph type.

2.2.2 *Session 2.* In Session 2, the statistician spontaneously provided the graph title first in every instance, followed by the graph type except in one case where it was given as the third item. For line graphs, she then reported the general trend of the line(s), for example: "over a 10-year time period, there was a slight decrease and then a dramatic increase". For bar graphs, the general layout of the bars was described next: "bars are stacked/paired/grouped...". Upon providing a general overview enabling participants to infer what it "looked like", the statistician offered a more detailed account of the graph such as the scales for both axes or the number and labels of the bars. Thus, as with the participants' questions in Session 1, the statistician adhered to a consistent order in which she presented information. Participants acknowledged the usefulness of this in the ensuing discussion because of the ease with which they were able to build a "skeleton" mental representation into which the graph data were then added.

In both sessions, participants struggled with the terms "X" and "Y" axes, preferring the more concrete terms "vertical" and "horizontal" instead. They also preferred using the terms "up" and "down" as opposed to "increasing" and "decreasing" to describe the trend. Interestingly, the statistician typically began the data description with the most recent entry, but it was easier for participants to insert data into their skeleton mental representation when it began from the leftmost point in the graph. In order to select the best vocabulary for the **iGraph**-LITE lexicon, the statistician's words most commonly used in descriptions and answers

to questions were compared with the kinds of questions, and the frequency with which these had been asked in Session 1. These data are presented in Table III below. Line color was not mentioned in Session 2, and neither was the trend for each of two lines in the relevant line graphs. Otherwise, the information sought in Session 1 was entirely consistent with the information provided in Session 2 either in the graph description or in response to participants' questions. The participants and statisticians' choice of vocabulary was also similar.

N	Type of question	Session 1 (n=6)	Statistician's vocabulary in Session 2
1	Title	6	Given in graph description
2	Axes	5	Given in graph description
3	Scales	1	Given in the graph description
4	Type of graph	5	Given in graph description
5	Portrait or landscape	1	Given in graph description
6	Purpose of graph	3	Implied in graph title
7	N lines/bars	2	Given in graph description
8	Negative values	2	Negative/positive; plus/minus
9	General trend	3	General trend
10	First/last value	3	Starting/ending point; bottom left corner
11	Highest/lowest point	5	Highest/lowest point, peaked, up/down
12		N/A	increase/decrease
13			Drop/dip (in data)
14		N/A	Cross (where lines actually cross)
15		N/A	Middle (of graph)

Table III. Number and frequency of question types in Session 1, and the statistician's use of vocabulary in Session 2

2.3 Discussion

Several important observations are worth discussing. First, the presentation of a graph description in a systematic, predictable format enabled the construction of a mental representation of it. Second, the sequence in which the statistician provided information in the descriptions in Session 2 closely resembled that in which participants requested information in Session 1. It was therefore adopted for **iGraph**-LITE. However, it was also apparent that participants were unable to construct a complete and accurate model of the graph from the statistician's descriptions alone. In order to refine their emerging mental models, they needed to ask follow-up questions, suggesting that the **iGraph**-LITE tool would need to allow users to interrogate graphs in addition to providing a static description of each graph. Finally, the importance of the vocabulary chosen to describe a graph was highlighted.

Taken together, the above results underscore the importance of using appropriate language and provide information in a specific order when describing graphs that cannot be seen.

3. GENERAL ARCHITECTURE OF THE SOLUTION

iGraph-LITE comprises three subsystems: (1) a knowledge representation (KR) system that enriches a basic semantic representation of line, bar and combination

graphs, (2) a natural language generation (NLG) system that produces a static description of a graph, and, (3) an interface that allows users to navigate the enriched representation of (1) by means of keyboard combinations, much as Jaws[8] or Dolphin[9] do. **iGraph**-Lite can therefore be used to generate descriptions to accompany graphs through the `longdesc` tag in plain HTML, and, if a graph is published with **iGraph**-Lite's full semantics embedded[10], the **iGraph**-Lite navigator can be used to explore the graph at different representational levels.

The KR system has three associated sub-processes: extraction, curation and enrichment, in that order of application. First, **extraction** is the process whereby **iGraph**-Lite first accesses the Object Model (OM) of graphs and charts in common plotting and graphing applications such as MS Excel^{TM}, GNUmeric, GNUPlot or Lotus^{TM}. **iGraph**-Lite extracts the data from the OMs and writes an eXtended Markup Language (i.e., XML[11]) and/or the Web Ontology Language, OWL[12], (see Dumontier et al. [2010]) representation of important graph information (including geometrical properties such as width, and height, and visual properties, such as colors or line styles [Chi et al. 1981]. Notice at this point that we do not work with raster images of graphs (JPGs or GIFs files), but rather with the graphing applications' internal representation, which already has some structure. The advantage of using XML and OWL is the often recognized one of application independence. At Statistics Canada (and other agencies) officers compose graphs in Microsoft Excel (or other graphing applications with some API to access the object model). This is the input to iGraph. We did not generate synthetic graphs except for our controlled experiments. All graphs we have described, and that are available in our online repository, are real graphs from editions of Statistics Canada's The Daily. They are not synthetic graphs, and there been approximately 5000 created since the inception of The Daily bulletin in the 1990s. There are probably several orders of magnitude more graphs on the web that are in raster format, but tackling those is beyond the scope of this paper.

Since graphs are created primarily for visual consumption (i.e., published on-line as GIF or JPG files), the information present in the OMs of graph-drawing applications is incomplete. For instance, in our corpus of approximately 500 graphs taken from Statistics Canada's "The Daily" publication, the title property of the OM is never used. Instead, graphs have *text boxes* that perform the title function implicitly. This is acceptable for visual inspection of the graph, but for accessibility applications, it means that there is merely some text at some (x, y) coordinate in the graph. No function and no semantic interpretation is provided. Thus, after retrieving the graph representations from graphing applications, a process of **curation** takes place, whereby **iGraph**-Lite applies about eighteen algorithms that help curate this first representation. For a detailed account of graph representation and curation, see Ferres et al. [2008], and Dumontier et al. [2010].

The last sub-process of the KR module is **enrichment**, whereby new and im-

---

proved information is added to the XML representation coming from the **extraction** and **curation** processes. It is worth noting that **iGraph**-LITE has a plug-in infrastructure that makes it simple to add more semantics to the curated representation of previous stages. This works by simply taking the XML or OWL representations as input and applying specialized algorithms. Examples of enriching algorithms may be the presentation of visual slopes for lines and a qualification of whether these slopes are small, moderate or sharp, or naming segments using slope event predicates such as "advances", "doubles" or "plummets". The process of **enrichment** is the last in the three-stage pipeline of the KR module.

The natural language message generator (the NLG module) of **iGraph**-LITE is based on a scalable string template engine[13] that allows complex messages to be transmitted to the user (for a theoretical discussion of NLG and templates, see [Reiter 1995] and [van Deemter et al. 2005]). Depending on the type of graph and the available information in the enriched semantics, the NLG system generates either a detailed (albeit static) description of an input graph (see Figure 1 below)[14] or a set of messages generated by a user's demands (the navigator interface, discussed in the next paragraph).

From the two sessions with blind participants described earlier, it was clear that much more than a simple static description was needed to make graphs truly accessible [Ferres et al. 2006]. We therefore designed the **iGraph**-LITE navigation tool, inspired by command-driven interfaces like Jaws[TM] At its core, the **iGraph**-LITE navigator queries the XML representation in particular ways. The result is an application which we believe is a significant improvement over the current ATs. Unlike other graph accessibility tools, **iGraph**-LITE does not attempt to automatically infer and communicate the graph's intended message to the user. Instead, **iGraph**-LITE attempts to make all the information contained in graphs easily accessible to users by means of keyboard commands like the usual screen reader interface, so that the user can explore the graph and infer its intended message given his or her own needs. Although **iGraph**-LITE also generates a summary and textual description of a graph, **iGraph**-LITE's main objective is language-based interactivity with the user, rather than describing a graph, communicating its intended meaning, or the generation of a static text. In summary, **iGraph**-LITE allows users to explore and interact with graphs in the manner which they prefer, on their own terms.

4.   IMPLEMENTATION OF THE INITIAL **IGRAPH**-LITE COMMANDS

iGraph-LITE's navigator is designed to be used "in focus". While this is so, key presses drive the interaction. Thus, when a recognizable keyboard event occurs, queries are executed to locate the relevant data in the curated and enriched XML representation of the graph (see previous section). Finally, the string template engine discussed above generates the final text or speech output.

Based on previous studies [Ferres et al. 2006], nine commands were implemented in [Ferres et al. 2007], see Table IV. Five of these were used for graph interaction and querying, and the other four for general application management, such as

---

[13]See, for instance, `http://velocity.apache.org/`
[14]The curious reader may also want to inspect the daily builds of the **iGraph**-LITE static descriptions at `http://www.inf.udec.cl/~leo/igraph/test.html`

Key	Function
Graph interaction commands	
d	general description of the graph (same as Fig. 1)
w	where (am I in the graph?)
→	move right one point
←	move left one point
h	help
s + digit	Skip n points
Speech handling commands	
p	increase the speed of speech
o	decrease the speed of speec
Esc	stop the TTS engine

Table IV. Original command set for the **iGraph**-LITE navigator

increasing the speed of the speech output, stopping the speech output, asking for help, or quitting the application. In this preliminary version of **iGraph**-LITE, it was hypothesized that interaction would begin with the "description" of the graph as shown in Figure 1, and follow with a simple interaction, as shown in Figure 2, where all lines starting with [**igr**]: are output via TTS, while all lines starting with [**usr**]: are the user's key presses. The original static text description provided basic information about the title, the graph type and its purpose, as well as about both the axes and the scales. Meanwhile, the navigation commands allowed users to move back and forth just one data point at a time by using the left and right arrow keys (← or →); by pressing s followed by a number, every subsequent ← or → command would skip the specified number of points in the graph, and the w key ("where am I?") would orient the user should she or he become lost in the graph. The next two sections address the issue of whether graphs, particularly line graphs, are accessible through **iGraph**-LITE, and whether the information it conveys is comparable to what experts would say about a graph.

5. FIRST FORMATIVE USABILITY STUDY

The purpose of this Study was to isolate and identify potential strengths and weaknesses of **iGraph**-LITE as well as to investigate the degree to which information contained in line graphs with a single line would be readily accessible for all possible users through **iGraph**-LITE.

5.1 Method

5.1.1 *Participants.* Ten legally blind (4 congenital, 6 non-congenital; five males, five females; mean age = 35.5) and ten sighted people (five males, five females; mean age = 32.8), were recruited from existing contacts and through Carleton University's Paul Menton Centre for students with special needs. All participants reported having used at least two types of graphs (e.g., line, bar, scatter, pie, etc.) in the past. They were tested individually in sessions lasting approximately one hour in a laboratory or in a quiet place convenient to the participant. Participants were offered $20 for their participation. Throughout the experiment, key presses were automatically time-stamped via **iGraph**-LITE's key-logging capability. Response

Operating profits up slightly in second quarter

$ billions   Seasonally adjusted

The line graph has the title "oper-
ating profits up slightly in second
quarter".   Quarterly values from
1996 to 2006 are shown. The val-
ues are given in billions of dollars.
The lowest yearly values occurred in
the second quarter of 1996 (26.7 bil-
lion dollars), and the highest values
in the fourth quarter of 2005 (57.6
billion dollars).

Fig. 1. A sample **iGraph**-LITE static description.

[igr]: I'm now looking at a line
graph entitled "Operating profits up
slightly in second quarter"
[usr]: `d`
[igr]: Ok, I'm describing the whole
graph now. [... as per the descrip-
tion shown in Fig. 1 ...]
[usr]: `w`
[igr]: You're at the beginning of the
graph. The current point is the sec-
ond quarter for 1996 and its value
26.7 billion dollars.
[usr]: `→`
[igr]: There is an increase between
Q2 and Q3, 1996 from 26.7 to 27.9
billions of dollars.
[usr]: `s` `2` `→`
[igr]: Skipping 2 values. There is
an increase between Q3, 1996 and
Q1, 1997 from 27.9 to 29.6 billions
of dollars.

Fig. 2. A sample **iGraph**-LITE interaction with
the original set of commands.

time was calculated as the time difference between the start of a question and the
end of the answer to it. The number of commands used was analyzed from the
key logging, and performance was assessed by the number of correct answers to the
experimental questions. The number of commands used to answer each question
was also calculated.

5.1.2 *Materials.* The study used a Dell Latitude D810 laptop computer sup-
porting the **iGraph**-LITE software. Three questionnaires were prepared. The graph
familiarity questionnaire sought information about the types of graphs participants
were accustomed to work with, the frequency with which they used these, and if
they were also asked to create graphs at work. The demographic questionnaire
sought information about education level, whether the participant had taken any
mathematics or statistics courses, if they were blind, and, if so, if they were con-
genitally blind. The post-test questionnaire asked about ease of use and likeability
of **iGraph**-LITE, as well as some open questions about its best/worst aspects.

Five single-line stimulus graphs were created in Microsoft MS Excel[TM]. These
graphs resembled graphs in The Daily (similar to Figure 3)[15], but The Daily graphs
themselves were raster images and their data were not available, so artificial ones

[15]The number of points for complex graphs differed for Study 2 (six) and Study 3 (seven). Oth-
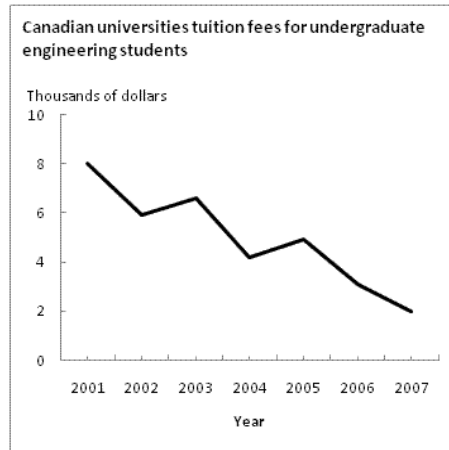erwise, graphs were similar. The graph in Fig. 3 belongs to study 3.

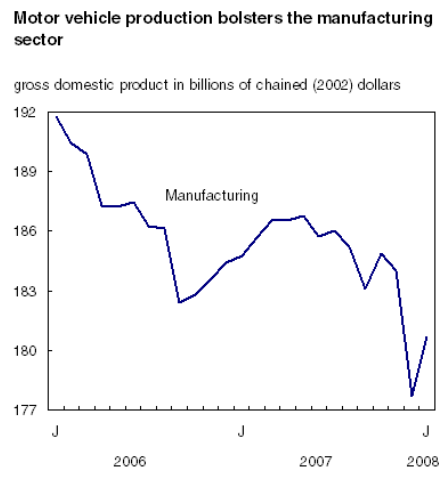Fig. 3. Actual graph stimulus from the evaluation study. Seven data points, downward trend.

Fig. 4. A graph published in "The Daily" on March 31, 2008, for comparison of ecological validity.

had to be constructed. Using the aesthetics of graphs in The Daily was a pragmatic decision driven by our commitment to bring accessibility to Statistics Canada's publications. Of the four test graphs, two graphs contained three data points and two graphs contained six data points. A fifth six data point line graph was used for pre-experimental practice to familiarize participants with the commands in **iGraph**-LITE. Five questions were asked of each of the four graphs[16] covering three main areas: global trend issues, patterns within the graphs, requests for specific items of data for sample questions, and descriptive issues such as title and units of the axes. Thus, the selected questions represent a range of information content an individual may find interesting when interrogating a graph (see [Ferres et al. 2006]). Response times and executed keyboard commands were automatically recorded. Participants' verbal answers were recorded manually.

5.1.3 *Procedure.* All participants were given the same materials. Questions were read aloud. The participant was seated at the computer and presented with information about the practice graph upon pressing h (help), which broadcasts all nine available commands and the function of each (see Table IV). Participants were encouraged to familiarize themselves with all the commands, spending as much time and repeating the commands as many times as they liked, until they felt comfortable with the system. Once they were ready, the four test graphs were presented aurally, one at a time, and the five questions were asked for each graph, also one at a time. The graphs were presented in a Latin Square design, but questions were always given in the same sequence. Upon hearing a question, the participant attempted to answer it by navigating the graph using the keyboard commands generated by

---

[16]Note, again, that all material is available for download, see Footnote 7

the **iGraph**-LITE engine[17]. Finally, the post-test questions were administered.

5.2   Results and discussion

Results from the pre-experimental questions are presented first, followed by an analysis of the performance data as well as of the response-time data. Reflections on command usage are then presented. Finally, we introduce a new list of commands, extended and improved as a consequence of the results of this study.

5.2.1   *Pre-experimental questions.* The graph familiarity questions revealed that all 20 participants had used line- and bar graphs before. Seven of the 10 blind participants encountered graphs less than once a week, and the three others indicated they encounter graphs several times a week. Only two of the 10 sighted participants encountered graphs less than once a week; the remaining eight participants said they encountered graphs almost every day. Only five of the sighted participants created graphs in their work, generally line- or bar charts, and usually in MS Excel^{TM} or SPSS^{TM}. The graph familiarity data confirmed that all participants had sufficient understanding of line graphs enabling them to use **iGraph**-LITE for its intended purpose. Eight of the 10 participants in each group had taken at least one course in mathematics or statistics at college or university. Five sighted and two visually impaired participants had taken more than five such courses.

5.2.2   *Performance- and response-time data analysis.* Performance data was analyzed by response time (in seconds), accuracy, and number of commands used to answer each question. As expected, all participants took significantly longer to answer questions for the complex graphs ($M = 400.20s$, $SD = 160.90$) than for the simple graphs ($M = 288.05s$, $SD = 134.33$), $t(19) = 2.98$, $p < .01$, and they used significantly more commands to answer those questions ($M = 82.70$, $SD = 64.86$) than questions for simple graphs ($M = 90.00$, $SD = 28.56$), $t(19) = 3.04$, $p < .01$. However, graph complexity did not affect performance accuracy, $p > .05$.

Visually impaired participants used nearly twice as many commands ($M = 163.70$, $SD = 101.12$) as sighted participants ($M = 90.00$, $SD = 40.32$), $t(18) = 2.14$, $p > .05$, regardless of graph complexity. Performance of the two participant-groups did not differ on the complex or on the simple graphs ($p > .05$). The visually impaired participants used the [Esc] command to stop the speech much more frequently than the sighted participants, probably as a result of their experience and proficiency with screen readers. Highly experienced screen reader users who are adept at listening to high-speed speech decide quickly if what they hear is relevant to the information they are seeking. Despite the difference in the number of commands the two groups used, response times did not differ ($p > .05$) and only one visually impaired participant actually used the speed-up command. Visually impaired participants were roughly as accurate in answering graph questions as sighted participants, $p > .05$, and, on average, they rated the system as being slightly easier to use ($M = 7.5/10$, $SD = 1.18$) than sighted participants ($M = 6/10$, $SD = 1.76$), $t(18) = 2.24$, $p < .05$. However, they did not like using **iGraph**-LITE more than the sighted participants, $p > .05$. Evidently, both participant-groups were able to

---

[17]FreeTTS was used as the text-to-speech system for output, see `http://freetts.sourceforge.net/`

discern information accurately. The visually impaired participants reported that **iGraph**-Lite was superior to the means they normally use for graph interpretation (tactile graphs, other people interpreting graphs for them). All but one of these participants reported their only means of gaining access to graphs was to rely on other people to provide them with descriptions. Participants reported appreciating the fact that **iGraph**-Lite eliminates this reliance on others. Several participants mentioned that they saw this reliance as a burden they impose on the person assisting them. One visually impaired participant reported **iGraph**-Lite as being superior to her other methods of gaining access to graphical information, which had been almost exclusively by tactile means. All the visually impaired participants much preferred to interpret the graphs themselves. The main attraction of **iGraph**-Lite was thus its ability to satisfy their desire for autonomy.

5.2.3 *Reflections on command-usage.* None of the participants used the minimum number of commands needed to answer the questions, which had been calculated *a priori*. It was also noteworthy that, whereas participants tended to listen to the graph description to answer the first graph question, the description was not repeated when a later question could also have been answered by it. It is possible that participants forgot about the contents of the description once they had answered some of the questions. If so, then it could be advantageous to provide a separate command for each component in the description while keeping the ⎡d⎤ (description) command.

For reasons that are unclear, the ⎡s⎤ ("skip") command was used rarely. Instead, participants were observed moving forwards and backwards through the graph several times using the less optimal arrow keys to find a particular answer. It is possible that they may have felt more comfortable progressing step-by-step through the graph because they feared getting lost. Alternatively, it may be unclear exactly how the command works. For instance, ⎡s⎤ ⎡1⎤ ("skipping one") actually moves the cursor two data points (i.e., it skips one point). Renaming the "skip" command to "move" may eliminate any confusion about the distance moved. Thus, "move 2" would achieve the same as "skip 1". Furthermore, since the number of data points was not reported in the description, participants would not have known the appropriate number of points to skip until they had moved through the graph at least once. This might also have discouraged them from using the "skip" command. Participants often began a search for an answer with a backward skip, even when they already were at the beginning of the graph. If they were unsure of their current location in the graph, they could have used the ⎡w⎤ ("where am I?") command yielding precisely that information. Yet, the "where am I?" command was used rarely, suggesting that it was either forgotten, or not useful.

Because of the tendency to use the description only when seeking the answer to the first question in a graph trial, the number of commands was increased to allow future participants to issue the precise command that fits their current information need. The description was therefore broken into separate commands in addition to its current operation. The directional commands of left and right arrows were broken down to allow the user to move in either direction point by point, but with the ability to choose the level of granularity at which they would like the information to be provided (i.e., full point description such as "increases from 3.2 percent in

Key	Function
Graph interaction commands	
x	help
d	general description of the graph
r	title of graph and axes titles with
	respective units of measurement and scales
a	go to the first point in the graph
f	go to the last point in the graph
h	move left one point (full point information)
l	move right one point (full point information)
j	move left one point (only point value)
k	move right one point (only point value)
n	Move left one point (only slope direction)
m	move right one point (only slope direction)
t	go to the lowest point in the graph
u	go to the highest point in the graph
w	where (am I in the graph?)
s + digit	Skip *n* points
Speech handling commands	
Ctrl	stop the TTS engine
p	increase the speed of speech
o	decrease the speed of speech

Table V. Improved command set for the **iGraph**-LITE navigator.

2002 to 5.4 percent in 2003" versus just the point value "5.4 percent" versus just the slope direction "up"). We created additional commands to increase efficiency in graph navigation and accommodate expressed desires from participants. These included commands to bring the user to the beginning/end of the graph as well as commands that provided the highest/lowest points in the graph. The Esc command to stop speech was replaced with Ctrl to be consistent with typical screen reader commands, as indicated by the blind users in the study. Attempts were made to select commands based on the spatial organization of the keyboard relative to finger placement and proximity, rather than basing the commands on mnemonic association alone (like those of VIM[18]). However, commands that did not have an inherent spatial mapping were assigned a mnemonic (i.e., w remained "where am I"). Therefore, commands to move left were assigned keys to the left of the index finger on the right hand and those to move right were assigned to letters spatially located to the right of the index finger of the right hand. The extended command set shown in Table V was therefore implemented in preparation for the second usability test.

6. SECOND FORMATIVE USABILITY STUDY

The purpose of this Study was to test the next iteration of the **iGraph**-LITE design, taking into consideration the lessons learned from the above Study. Thus, once the extended command set had been implemented, the second formative usability test

---

[18]http://www.vim.org/

was conducted.

6.1 Method

6.1.1 *Participants.* A new sample of 10 legally blind (6 congenital, 4 non-congenital; five males, five females; mean age = 42.9) and 10 sighted people (five males, five females; mean age = 43.5) was recruited from the same sources as before. They were tested individually in two sessions, the first lasting approximately 10 minutes and the second taking approximately 30 minutes. Participants received $20 in compensation for their participation, paid at the end of the second session.

6.1.2 *Materials.* Four new single-line test graphs were prepared in MS Excel. As before, the number of data points was used to indicate graph complexity: the two simple graphs contained four data points, and the two "complex" graphs contained seven data points. For each graph, six questions were generated covering three main areas: (1) global trend-related questions, (2) local questions requiring information about individual data points, and (3) background questions related to the topic of the graph. A fifth graph was prepared for pre-experimental practice as before. Throughout the experiment, key presses were automatically recorded and time-stamped via **iGraph**-Lite's key logging component on the experimenter's computer. Response time was calculated as the time difference between the start of a question and the end of the answer. The number of commands used was analyzed from the key logging component, and performance was assessed by the number of correct answers to the experimental questions. The number of commands used to answer each question was also calculated.

6.1.3 *Procedure.* Cued recall was tested in Session 1, and Session 2 was a free recall test as well as offering an opportunity to use the commands to answer questions about the new set of line graphs. In preparation for session 1, participants had learned the functions and commands of the entire command set on their own time. This was done to reduce the likelihood that less-than-optimal command usage in Session 2 could be attributed to participants forgetting some of the commands.

Upon agreeing to participate, the list of the 18 commands shown in Table V was emailed to participants together with explanations of the functions. These 18 commands included the breakdown of the description command [d] : [r] offered the axes and graph titles, the [u] and [t] commands would take the user to and communicate the highest and lowest point values, respectively, and finally the [a] and [f] would take the user to and communicate the first and last data point values of the graph, respectively, providing the range. In addition to breaking down the description, we also created, as per users expressed requests, the option of **iGraph**-Lite communicating only the value of a given point while moving left and right ([j] and [k]), instead of the full point information (i.e. [h] and [l]) and the option of moving left and right ([n] and [m]) communicating only the slope direction (up, down, or stays the same). Once participants felt confident that they would be able to recall all commands and functions, they contacted the experimenter to arrange a phone session. For one half of the participants, the experimenter cited the commands in random order, asking the participant to explain the function of each. This was reversed for the other half, for whom the experimenter provided the

explanation, requiring the participant to cite the relevant command. This process continued until all commands had been matched correctly with the explanations once. The purpose of this session was to ensure that participants knew the commands and functions of the entire **iGraph**-Lite command set before using them to answer questions. Participants were not introduced to **iGraph**-Lite in Session 1. Session 2, designed to assess the degree to which **iGraph**-Lite would be used optimally, was conducted approximately one week after Session 1.

Using a free recall paradigm, participants first cited the commands and functions. If recall was less than 100%, the experimenter provided the explanation for the missing commands. These were repeated until the complete set had been recalled correctly. Participants were then given the practice graph, spending as much time as they wanted to familiarize themselves with using the commands and listening to the synthetic voice in **iGraph**-Lite. They were encouraged to try all the commands. Although participants had memorized the functions of each command, it was important for them to experience practical application of these before commencing the experiment. When they were ready, they were given the first test graph which, as before, could not be seen. They used the keyboard for command entry and listened to the voice response. Commands were logged automatically to the same computer as in Section 5.1.2.

A Latin Square design was adopted again for presentation of the four test graphs, and questions were always given in the same sequence to each participant. When all questions had been attempted, the same post-experimental questions were administered as before, and participants were invited to make any comments before being debriefed.

6.2   Results and discussion

As before, results of the pre-experimental questionnaires are presented first, followed by the performance- and response-time analyses. Participants' navigation strategies and reflection on their comments are then presented and discussed.

6.2.1   *Pre-experimental questions.* Results from the graph familiarity questionnaire revealed that all 10 sighted participants and seven of the 10 blind participants had worked with line graphs before. Although six of the 10 blind and six of the 10 sighted participants indicated that they encountered graphs less than once per week, the graphs they did encounter were most often line- and bar graphs. This was also true for the remaining participants who reported encountering graphs more frequently during a typical week. Only three of the sighted participants said they were required to create graphs on a monthly basis.

The demographic questionnaire data revealed that eight of the blind and three of the sighted participants had taken at least one course in mathematics or statistics at college or university. Blind participants said they relied on screen readers or on others to describe the graphs they encountered, or else they used tactile graphs. However, participants reported that they knew of no tool that could assist in graph accessibility.

6.2.2   *Performance- and response-time data analysis.* As graph complexity did not affect response times, the number of commands used, or accuracy (all $p >$ .05), data from all graphs were pooled in the following analyses using independent

samples t-tests. As before, response time did not differ between the two participant-groups ($p > .05$), nor did the accuracy scores ($p > .05$). Both groups answered approximately 85% ($n = 17/20$) of the questions correctly (blind participants: $M = 84.17$, $SD = 11.08$; sighted participants: $M = 85.42$, $SD = .35$), thereby clearly demonstrating their ability to explore and extract information on a global, local, and background level in **iGraph**-LITE.

As before, the visually impaired participants used significantly more commands to answer the questions ($M = 171.70$, $SD = 60.28$) than the sighted participants ($M = 113.80$, $SD = 34.67$), $t(18) = 2.63$, $p < .05$. This difference was mainly due to them issuing a redundant direction command much more frequently when they were at either end of a graph (total 189 times) than the sighted participants (total 94 times). The data are insufficient to reveal if they lacked awareness of their current location in the graph, or if they lacked "confidence" in the same. We are inclined to interpret it as evidence of their need to confirm their location to reassure themselves because their performance was as accurate, and as fast as that of sighted participants. This observation is consistent with studies showing that visually impaired people need reference points for spatial orientation when exploring objects in space [Sribunruangrit et al. 2003]. As discussed earlier, some software applications designed for visually impaired users provide haptic or sonification cues to convey landmarks (e.g. major gridlines, or start and end lines, see MultiVis, or Phantom). Future iterations of **iGraph**-LITE should therefore provide cues on current location at both endpoints of a graph.

The post-experimental questions showed that both groups found **iGraph**-LITE easy to use (blind $M = 7.8/10$, $SD = 1.32$; sighted $M = 7.7/10$, $SD = 1.16$) and that they liked using **iGraph**-LITE (blind $M = 7.5/10$, $SD = 1.72$; sighted $M = 7.4/10$, $SD = 1.84$). Both these sets of ratings represented a slight increase compared with the first formative usability study, suggesting that the amendments to the **iGraph**-LITE commands improved the system. However, recall that participants here were already familiar with the commands, in contrast to participants in the earlier study. Blind users all agreed that **iGraph**-LITE was "much better" than the tools they normally use. All but one participant reported that their only means of gaining access to graphs was to rely on other people to provide descriptions. Participants said they appreciated that this tool does not require them having to rely on others to interpret the graph and then relay their interpretation to them. Several participants mentioned that they see this as a burden they are putting on the person assisting them. One blind participant reported **iGraph**-LITE as being superior to her other means of gaining access to graphs which had been through the use of tactile graphs. She reported that tactile graphs are expensive and time consuming to create. The general consensus among the participants was that they would prefer the option to interpret the graphs they are interested in for themselves, by themselves.

6.2.3 *Navigation strategies.* Analysis of navigation paths revealed that blind participants used commands moving to the left ($\boxed{\text{h}}$ / $\boxed{\text{j}}$) nearly three times more frequently than sighted people who, in turn, tended to go to the beginning of the graph ($\boxed{\text{a}}$ command) and move right through the graph, and then return to the beginning ($\boxed{\text{a}}$ command) again if they needed more information. Sighted

participants returned to the beginning of the graphs twice as often as blind people. The finding suggests that blind users do not use the same visualization strategies as sighted users. Going "backwards" through a graph is apparently not as confusing or problematic for their mental representation of that graph as it is for sighted users. Their strategy of navigating forwards and backwards through the graph, rather than using the sighted participants' common strategy of skipping to the beginning, likely contributed to the higher number of commands they used.

Study 1, discussed earlier, showed that visually impaired participants found it difficult to construct a mental model of a graph when given the most recent data point first. The reason for visually impaired participants adopting a step-by-step bi-directional strategy here instead of moving straight to the leftmost, oldest data point, suggests that they may visualize a graph differently when they control the commands than when they request information from another person. The option of moving to the beginning of a graph was not available in the earlier study, making a comparison of the use of this command impossible. Finally, the observation that visually impaired participants used the directional commands yielding the value of individual data points much more often than sighted participants, confirms the benefit of breaking down the graph description into single elements. Barring their similarity to canonical graphs published by The Daily, the test graphs and questions tested thus far were produced in the lab. They may not be indicative of the kind of information that visually impaired people who work with graphs may actually need. In an attempt to verify the usefulness of **iGraph**-LITE to visually impaired people working with graphs, a small field study was therefore conducted next.

7. FIELD STUDY: QUESTIONS A VISUALLY IMPAIRED USER MIGHT ASK

This study was designed to assess the degree to which the current **iGraph**-LITE capabilities met the requirements of visually impaired users in a field setting. Because it proved almost impossible to locate visually impaired people who routinely work with graphs, we were able to recruit only one user meeting that requirement. Over several weeks he selected and submitted a sample of four articles containing graphs from which he wanted more detailed information than could be obtained from his screen reader. The articles were read using his screen reader, but as he did not have access to **iGraph**-LITE on his computer as yet, he relied on the text in the article for identifying and comprehending the contents of the graph. One of the researchers read the submitted articles before setting a time for discussing the graphs with the participant by phone. The conversation was audio recorded and transcribed *ad verbatim* with the participant's permission.

7.1 Results and discussion

The kinds of questions the participant wished to ask were identified for each graph to determine how well **iGraph**-LITE would meet these informational requirements if it had been used when reading the accompanying articles. As Table VI shows, all information sought could have been obtained for graphs 2 and 4. Fewer questions were asked for graphs 1 and 3, although the software could have provided the additional information that was not sought. The example below shows that eight of the nine questions (shown in brackets below) would, in fact, be answered by the description alone:

N	Information sought	Graph 1	Graph 2	Graph 3	Graph 4
1	Title of graph	✓	✓	✓	✓
2	Type of graph	✓	✓	✓	✓
3	Axes	✓	✓	✓	✓
4	Highest/lowest point	✓	✓	✗	✓
5	In-between points*	✓	✓	✓	✓
6	Purpose of graph	✓	✓	✗	✓
7	Time frame shown	✓	✓	✗	✓
8	N bars (data points)	✗	✓	✓	✓
9	Title of each bar	✓	✓	✓	✓

Table VI. Information sought that is handled by iGraph-Lite for each submitted graph (*The participant called all points falling between the highest and lowest on the graph "in-between points")

> The line graph (1) has the title "operating profits up slightly in second quarter" (1;6). Quarterly values (8) from 1996 to 2006 (7) are shown. The values are given in billions of dollars (3). The lowest yearly values occurred in the second quarter of 1996 (26.7 billion dollars), and the highest values in the fourth quarter of 2005 (57.6 billion dollars). (4).

The information that this participant wanted from a graph appears to be highly consistent with the findings from Study 1. This was also confirmed by the participant's comments when he listed and rank ordered the information items that were most important to him:

(1) Graph title
(2) Type of graph
(3) Title of x and y axis (or bar's units of measurement)
(4) General shape of the graph ("visual description at a glance" or "what does the graph look like at a first glance?"), use familiar shapes (such as letters) to liken the graph's general shape to them, for instance, "looks like an 'M'"
(5) Highest and lowest points of the graph and any context to support those points (such as unit of measurement, year, weight, percentage, etc.)
(6) Color of the axes, bar, etc.

Although graph shape (e.g. resembling the letter "W") had been discussed in study 1, this caused confusion for congenitally blind participants who relied on non-visual cues. Line color had also been requested earlier. However, as color is likely to be more relevant for generating than for reading graphs, color was left out of the **iGraph**-LITE lexicon.

The participant also emphasized the importance of consistency in the sequence in which information is given in the speech output, in order to render the content predictable. Finally, he identified several issues in the four submitted graphs that made it difficult or impossible for him to visualize or comprehend the information:

(1) Type of graph was unclear
(2) Purpose of graph was unclear
(3) Trend line in bar graph (type of graph missing in accompanying text)

(4) Very little explanation of the graph in the text (questionable if a graph was needed in that instance at all)

(5) Text describing the contents of the graph spatially too far removed from the graph, and graph labels were missing

(6) Graph talked about "the previous month" without saying which month it refers to

(7) The fact that one graph (number 3) was a stacked bar graph with three types of information was not obvious at all (type of graph, n data points) and required quite some discussion as well as several questions to resolve

(8) Unit of measurement was unclear

(9) Year when data were collected was not mentioned

Most of these issues are addressed in **iGraph**-Lite. However, during the pre- and post-test questionnaire, blind participants in our studies reported that they usually asked the help of *other people* to explain graphs to them. The natural question to follow was how **iGraph**-Lite's descriptions and access to information fared against those provided by confederates. It was therefore desirable to also explore graph reading/interpretation strategies that sighted people may adopt. Moreover, we assume that most of the people recruited by our participants to explain the graphs would not necessarily be expert graph readers, that is, people who work with these representations on a daily basis but family, friends and colleagues. Thus, another natural test for **iGraph**-Lite was to find out how it fared against graph descriptions by experts. The next study addressed these questions.

8. EXPERT AND NOVICE GRAPH-INTERPRETATION STRATEGIES

The purpose of this Study was to assess the degree to which the **iGraph**-Lite lexicon corresponded to vocabulary that sighted graph experts and novices use spontaneously when explaining the content and meaning of a graph. Contrary to the above studies in which participants were unable to see the graph being interrogated, participants in this study were presented with the graphs, complete with titles and axis labels. They were not introduced to, nor did they interact with, **iGraph**-Lite. Previous research has shown that expertise tends to be domain-specific [Abernethy et al. 2005; Ericsson 2005; Sims and Mayer 2002], but it is unclear if this specificity also applies to graph interpretation.

8.1 Method

8.1.1 *Participants.* Twenty six participants were recruited. Experts were faculty or senior graduate students in Business ($n = 8$) or in Psychology ($n = 8$), and novices were undergraduate Business or Psychology students ($n = 10$). Each individual was asked to explain the semantic characteristics of 24 monochrome black/white line graphs, presented one at a time on a computer screen in a standard size, and each comprising two lines and three data points, a graph title and axes labels. Graphs were presented in a different random order each time they were shown, with orders generated by a software application, DirectRT$^{\text{TM}}$. The graph content referred to psychology-relevant material in one half, and to business-relevant material in the other half of the graphs.

iGraph-Lite function	iGraph-Lite command
Graph title	d / r
Scale/X-axis	d / r
Unit/Y-axis	d / r
Lowest point	u
Highest point	t
Comparison between specific points with full point description	h / l
Point-specific information	w
Beginning of graph	a
End of graph	f

Table VII. **iGraph**-Lite functions noted in the transcripts.

8.1.2 *Procedure.* Upon reading and signing the informed consent form, participants were given a set of four graphs to familiarize them with the method and the requirement to separately describe the syntactic characteristics and then explain the content and meaning of each graph. Data from these were not included in the analysis. Participants described and explained a total of 20 line graphs, half of which were cast in a business context and half in a psychological context. The order of graphs was randomized for each participant. At the end of the study, they were debriefed and dismissed. There were no time constraints for the explanations, and no feedback was provided during the experiment.

8.2 Results

The results were analyzed in several ways. First, the graph characteristics participants mentioned that were also contained in the **iGraph**-Lite description and command set were determined. Second, the frequency with which these were noted was ascertained for each of the three participant-groups to uncover potential and systematic differences between novices and experts and between people representing different disciplines. Finally, the vocabulary used to describe various aspects of the test graphs was noted.

Graph explanations were transcribed ad verbatim to identify items of information that corresponded with **iGraph**-Lite's lexicon. Due to occasional technical glitches, recordings could not be retrieved for 10 of the 624 explanations, leaving a sample of 614 for the analysis. Since participants were able to see the to-be-explained graphs, some of the commands offered by **iGraph**-Lite were not relevant here. For example, commands used to "increase the speed of speech" and to "move one point right with only point value given" could not be assessed. Table VII below outlines the nine characteristics identified in the analysis that are also contained in **iGraph**-Lite.

The frequency of occurrence of these was noted as shown in Figure 5. Each item was only counted once per participant per graph even if a given utterance was repeated in an explanation. Due to the missing data, the figure shows percentages rather than raw numbers.

—*Graph Title*: Business and psychology experts provided the graph title for almost every graph ($n = 144/158 = 91\%$ and $n = 158/176 = 90\%$, respectively), whereas novices only did so in 36% ($n = 82/226$) of the graphs. When the graph title was
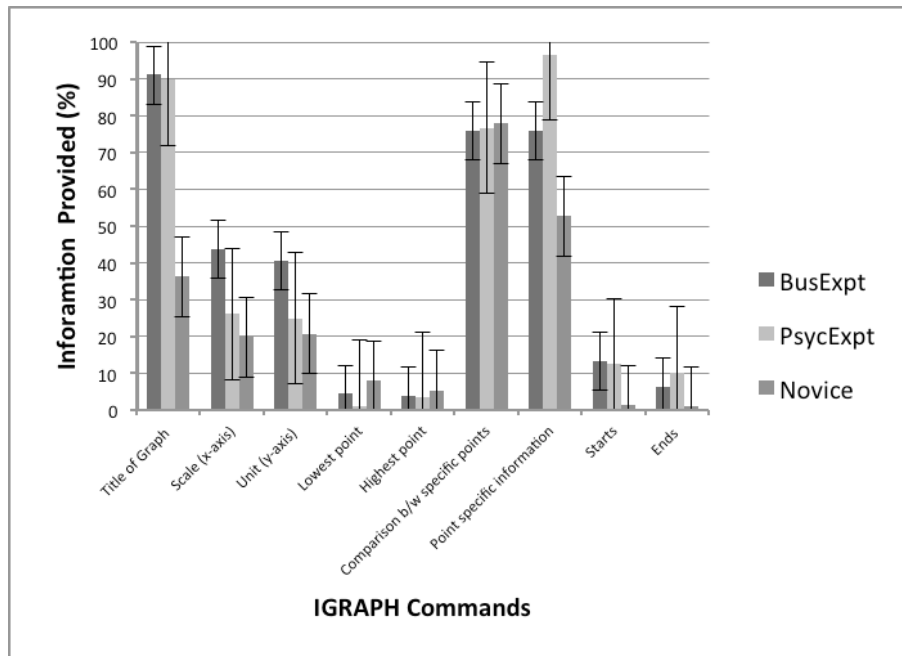
Fig. 5.   Graph illustrating percentage of information provided

paraphrased, it was coded as a summary. This was true for 38% ($n = 54/144$) of business experts' explanations, for 34% ($n = 53/158$) of psychology experts, and for 63% ($n = 52/82$) of novices' explanations. However, novices' summaries were incorrect in 23% ($n = 12/52$) of the cases. Novices' summaries omitted pertinent information included in the title, and their attempts to paraphrase changed the meaning of the graph in some way. For business experts, the title was the first item of information in all but two instances ($n = 142/144 = 99\%$), for psychology experts it was first in 95% ($n = 150/158$) of all instances, and for novices, in 34% ($n = 28/82$) of all instances.

—*Axis labels*: Business experts gave axis labels (scale and Unit in Figure 5) in 42% ($n = 67/158$) of the cases; psychology experts in 26% ($n = 45/176$) of cases, and novices in 20% ($n = 46/226$). Generally, these were mentioned immediately after the graph title. Business experts referred to the axes as "X-Axis" and "Y-Axis" three times more often than calling them "horizontal axis" and "vertical axis", whereas psychology experts always referred to the axes as X- and Y-axes.

—*Comparisons and Point Specific Information*: Comparisons of data points showed that participants occasionally discussed a particular data point in relation to another, for example, "... increased from 2005 to 2006", and at other times they noted point-specific data, for example, "low percentage in 2005". Business experts discussed both comparisons and point-specific explanations in 76% ($n = 120/158$) of the cases. Interestingly, psychology experts dwelled on point specific information far more often than providing comparisons of the data points. Psychology experts provided point-specific information for 97% ($n = 170/176$)

of all graphs and data comparisons in 77% ($n = 135/176$) of the cases. Novices provided point-specific information only in 53% ($n = 119/226$) of the time, relying more on comparisons, which they provided 78% ($n = 176/226$) of the time, slightly more than the experts.

—*Highest/Lowest Points*: Lowest and highest points were rarely mentioned at all.

—*Beginning/End Information*: Beginning and end points of the graph were coded when they were specifically referred to, for example, as in "initially India had a low percentage in the first year, but ended up high in the last year". They were not coded if these points were mentioned more generally, for example, "India started low, increased in the next year, and increased again in the final year". The former shows how **iGraph**-Lite's Beginning `a` and End `f` commands could be implemented. Generic descriptions would not benefit from using the `a` / `f` commands but would be classified as a comparison between data points. As Figure 5 shows, these were also mentioned very rarely.

All the nine types of data that corresponded to **iGraph**-Lite commands were thus mentioned by all three participant-groups, although the frequency of occurrence varied substantially both between groups and for the various data types. Apparently, sighted and visually impaired experts seem to agree on the importance of the graph title, information about the graph axes, and the data points, as well at as on the sequence in which to report these.

8.3 Concepts vis à vis vocabulary

The next analysis was performed to ascertain how closely the vocabulary in the present version of **iGraph**-Lite resembles the vocabulary spontaneously chosen by the experts and novices to explain the sample of graphs included in this Study. The above analysis showed that experts referred to the abscissa as the "X-axis", and the ordinate as the "Y-axis" invariably, whereas both business experts and novices used these terms as well as the "horizontal" and "vertical" axes.

The sighted participants' vocabulary was sorted into four main concepts: Axis, Slope Events, Slope Quality, and Colour. Axis vocabulary included any terms referring to the axes, specifically x/y, or horizontal/vertical. Slope events are defined by words that depict changes in the slope of the graph (e.g., dropped, spiked, plateaus, etc.), whereas Slope Quality refers to terms that describe the slope event (e.g., rapidly, gradually, sharply). Colour simply refers to whether colour of the lines was mentioned. Despite the use of monochrome graphs, some participants still described colour (e.g., the light gray line versus the black line). Table VIII below indicates the frequency of each concept by each expertise group. Again, each occurrence was only counted once per participant per graph.

All three participant groups referred to the graph axes, but the experts (both business and psychology) relied heavily on the use of x/y axis ($n=167/193=87\%$) compared to the terms "horizontal/vertical" 13% of the time ($n=26/200=13\%$), whereas novices used the terms "horizontal/vertical" ($n=61/89=69\%$) much more often to describe the axes.

Although all three participant-groups described slope events, novices mentioned these most frequently. Close scrutiny of the vocabulary used to describe these events showed clearly that individuals used a wide variety of terms, resulting in

Concept	Bus. Expert	Psych. Expert	Novice
Axis	104	89	89
Slope Events	114	112	147
Slope Quality	100	129	119
Colour	8	34	18

Table VIII.   Frequency of concept utterances by expertise group.

little commonality between groups. The top ten slope event terms used by each group showed that only two identical terms were used by all three groups, namely "dropped" and "equals". Far more agreement was found for the terms reflecting slope quality, i.e., the extent to which the lines changed. Terms shared among all groups included "steadily", "dramatically", "gradually", "slowly", and "steeply". Terms that were only uttered once throughout the entire study are not included in the list in Table IX. It shows the actual frequency, from highest to lowest, with which the variety of terms was used to describe slope events and slope quality across all participants.

Slope event	dropped (68), positive/negative slope (43), goes up/down (41), stayed the same (30), equaled (26), remains stable (18), Remains constant (16), continual increase (13), rose (11), overtook (9), surpasses (9), intersection (9), flattens out (8), lines meet (8), spiked (7), levels off (7), plateaus (6), no change (6), pattern flips (6), fell (6), lines cross (6), peaked (4), goes horizontally (3), recovered (3), rebound (3), plummeted (3), forms diagonal line (2), dip (2)
Slope quality	steadily (50), dramatically (41), gradually (39), sharply (35), slowly (32), steeply, (28), slightly (24), fast (22), rapidly (18), significantly (12), huge (10), strong (9), substantially (6), tremendously (6), moderate (5), large (3), sudden (2), drastic (2), severe (2), noticeably (2)

Table IX. Frequency of slope event and slope quality descriptor utterances collapsed across expertise group.

Although **iGraph**-LITE does define slope descriptors based on the angle of the line in the graph, the program does not yet generate the rich language noted here to support this fine-tuned level of graph description. More research is needed to determine if it will be necessary to add synonyms to **iGraph**-LITE, and to investigate if the standard vocabulary may give rise to misunderstandings.

9. PRELIMINARY RECOMMENDATIONS FOR INCLUDING GRAPHS IN TEXTUAL ARTICLES

Taken together, all the above findings obtained from sighted as well as from visually impaired participants allow us to propose a set of preliminary recommendations intended for authors of written articles that include graphical material in an effort to help them improve the accessibility of such material. The first recommendation applies only to blind and visually impaired audiences; the remainder would apply to all audiences.

(1) Remember that screen readers cannot decode graphical material

(2) Place a description of the graph close to the graph itself to facilitate construction of a mental representation of it

(3) Report items of information about graphs in a consistent sequence:
  (a) Graph title
  (b) Type of graph (line-, bar-, pie- etc.)
  (c) Purpose of graph
  (d) Title of X and Y axis
  (e) Units of measurement (for bar graphs: n bars, mention if stacked bar graph)
  (f) General shape of the graph ("visual description at a glance": sloping from bottom left to top right; sloping m shape, almost horizontal, etc.)
  (g) General trend
  (h) Highest and lowest points of the graph and any context to support those points (such as unit of measurement, year, weight, percentage, etc.)
  (i) Color of the axes, bar A, B, etc.

10.   GENERAL DISCUSSION, CONCLUSIONS AND FUTURE WORK

In this paper we have discussed **iGraph**-LITE, an advanced graph navigation system. In addition to enabling graph navigation, it also includes the potential for analyzing data through commands at several different levels of granularity and the possibility of easily extending what can be said at each step of an analysis. Most importantly, and unlike many Artificial Intelligence applications, it does not "think" or aim to replace the blind or visually impaired user. Rather, it was designed to help users to understand graphs by themselves, in their own time, and according to their unique needs. Hence, **iGraph**-LITE does not "understand" graphs or provide outputs showing what it "thinks" the graph is about. Some authors comment that it would be desirable to highlight the most salient features of a graph [Kildal and Brewster 2006], however we would argue that the relative salience of information depends partly on the question(s) the user has in mind when interrogating graphs as well as on the message the graph creator intended to convey. This is why **iGraph**-LITE offers users active interaction with graphs by making everything about these accessible in different ways. This helps the user to form their own internal model of the graph rather than communicating one particular model inferred by the tool from the graph data.

The studies presented here suggest that people with visual impairments can interact with natural language interfaces accurately to answer relatively complex questions about simple line graphs. They also suggest that visually impaired people are keen to use natural language technologies. Other researchers have shown that people seem to prefer speech output to other possible modalities such as interfaces providing a blend of tactile feedback and non-speech sound [Wall and Brewster 2006], or interfaces based on sound and vibration [Evreinova et al. 2008]. This is true even though some of these studies show that modalities other than speech, for example, sonification [Flowers et al. 2005] enable an effective overview of graphs that would surpass the speed with which words can be spoken and comprehended. One would think that the motivation to reduce task-performance time to an absolute minimum would be most important to users in a pressured work situation such as that of the Statistics Canada employees who took part in most of the above

studies. However, our results suggest that the value of autonomy may rank even higher than efficiency in the visually impaired population for whom **iGraph**-Lite is intended. The fact that a synthetic voice output as in the current version of **iGraph**-Lite did not appear to detract from its likeability may, at first glance, seem surprising. However, note that experienced screen reader users are so accustomed to synthetic voices that they are able to increase the speed of these such that others lacking that skill would be unable to discern individual words. Another possibility for the positive feedback our participants provided is that our stimulus graphs were very simple, even those we called "complex". The degree to which tools like **iGraph**-Lite may be as easy to use and as suitable for highly complex graphs with many more lines and more data points remains to be determined.

To understand why we chose a Natural Language Interface (NLI) for **iGraph**-Lite, a more detailed comparison with the alternatives is called for: embossed paper and Braille, sonification and haptic interfaces. The embossed paper and Braille solution is possibly the most straightforward for communicating graph information to users who are blind. Combined sonification and haptic solutions could work well, as those two modalities together can supplement one another without negative interference. However, the mere cost of embossing printers, together with the enormous number of graphs produced by agencies such as Statistics Canada, makes this technology impractical. Also, the resolution of the embossed graphs is often too low, for instance, to determine the exact value of the maximum point of the graph, unless there is some note in Braille specifying it. Not all points will be accessible in the embossed representation and, just as importantly, relatively few blind and visually impaired users are Braille-literate.

The issue of embossed representations ties in with some of the disadvantages of sonification and haptics. The auditory and haptic sensory systems are best for communicating trends rather than for navigation and the communication of specific results, and we cannot communicate the title of a graph in sonification and haptics. We can signal that we are now at the maximum point of the graph, but which point is it and where exactly are we in the graph? How do we communicate the precise value of the maximum point? Both sonification and haptics suffer from this drawback; they cannot, by themselves, communicate specific information in a way the NLI can. By the same token, the NLI interface cannot really communicate overall trends as well as the sonification and haptic interfaces. Deterioration in task performance is to be expected in situations in which people's auditory working memory is exhausted. It would therefore be reasonable to assume that the interfaces using mixed modalities could be very attractive, especially as it has already been shown that people are able very quickly to get an overview sense of even very complex data via sound [Wall and Brewster 2006]. Since people with severe visual impairments rely heavily on both the auditory and the tactile sensory modalities, one would assume that a well-designed interface using a mixture of those modalities as well as speech would be quite easy for that population to learn to use quickly and skillfully. This points again to a hybrid system as possibly the best solution, or "silver bullet", to graph accessibility. It is undeniable that languages (understood here as a system of symbols that has a combinatorial syntax and a compositional semantics, thus including, for instance, sign language) are the most complex, and richest systems

of communication known to humankind, and probably one of the easiest to understand without additional training or special training or special talents. The power of language lies in the possibility of infinite levels of abstraction, going from the highest possible level ("the graph has an upwards trend") to the lowest ("point two is July, and it has a value of 23.2"). This richness of vocabulary and syntax makes the NLI to graphs desirable. The other systems are, in all practical situations, more constrained in the achievable specificity of a potential message. However, natural language also has its shortcomings. Its richness comes at the cost of much work to be done on the semantic representation of the graph information to be spoken or reasoned about. Every time an NLI must say something, that information must be encoded in the representation with enough detail to say exactly that and not something else. Besides, language generation systems (systems that take some semantic representation as input and output a natural language sentence or text) nowadays are not extremely flexible; quite some work must also be undertaken to adapt them to particular applications such as **iGraph**-LITE. **iGraph**-LITE uses a simple, yet powerful, template-based system that can be extended with little work, but at the same time, the complexity of the produced sentences is only just good enough (no embedded clauses, for instance). **iGraph**-LITE produces only regular language, not context-free language. This does not impair the efficacy of the solution, but at times, it makes the messages sound artificial.

The reader may wonder at the sequence in which we have presented the above studies. Sections 7 and 8, in particular, could be conceptualized as requirements capture studies. With the benefit of hindsight, these could more logically have preceded the order in which they were conducted, revealing the way our thinking evolved with the project even if that also reveals a weakness in our research planning. In our defense, it is often the case that research develops in a non-linear fashion that, retrospectively, may seem back-to-front. Our goal here was to explore what happens when we rely exclusively on spoken words to describe graphical representations of data. We believe that goal has been accomplished even if the order in which some parts of the research are reported may seem suboptimal.

The design of **iGraph**-LITE was informed by studies of graph comprehension strategies involving blind and visually impaired participants. One may say that the target community effectively designed **iGraph**-LITE. Again, with the benefit of 20:20 hindsight, there are several things we would have liked to include, or done better. For instance, there is a rich vocabulary of ways to describe the quality of a line (e.g. "steep" in the phrase "a steep increase") and its direction in the graph (e.g. "skyrocketed" as in "automotive sales skyrocketed in October"). We have identified and characterized a set of these words (24 for quality and 70 for direction) but have not yet included them in the **iGraph**-LITE interface. This is important because statistical communicators often use these words in text accompanying online graphs. However, given the inherent visual nature of these words, their meaning is often lost, especially to congenitally blind screen reader users. The fact that we have worked mainly on line graphs and bar graphs is another limitation. There are other kinds of graphs that should be included, such as stacked bar graphs, area graphs, and several combinations of these. One difficulty of working this into **iGraph**-LITE is that each kind of graph was designed to convey a particular meaning: lines represent time,

bars represent comparisons, pies represent percentages of a whole, but combination graphs represent complex interactions of these messages that must be studied more carefully. This also includes more complex messages when two lines in a time series graph interact in complex ways: they cross, they converge, they diverge, etc. A picture may well be worth a thousand words, but sometimes a hundred words is all we have to get the graph message across.

Also in hindsight, we should have implemented **iGraph**-LITE as a web application rather than as a standalone application. This could have yielded portability and facilitated mass adoption of the tool. It would also have avoided the need to install more software on the client machine, indeed on every machine with which the user works. In terms of future work, several of the visually impaired participants commented that they would prefer commands that are similar to other common speech programs. However, the tasks people are able to perform using conventional screen readers differ so much from those carried out here that a perfect match would be impossible to achieve. Still, there is room for improvement to the mapping of **iGraph**-LITE commands to functions.

A final thought is that we believe that screen readers are an under-utilized and under-researched technology in the sense that they may be able to work with much richer semantics and, hence, linguistic messages. These deep-knowledge interactive language technologies could provide knowledge of the navigational and, in general, cognitive strategies people use to access those items of visually presented information that cause problems.

REFERENCES

ABERNETHY, B., BAKER, J., AND COTE, J. 2005. Transfer of pattern recall skills may contribute to the development of sport expertise. *Applied Cognitive Psychology 19*, 705–718.

ABU DOUSH, I., PONTELLI, E., SIMON, D., SON, T. C., AND MA, O. 2009. Making Microsoft Excel accessible: multimodal presentation of charts. In *Proceedings of the 11th International ACM SIGACCESS Conference on Computers and Accessibility*. Assets '09. ACM, New York, NY, USA, 147–154.

BEN-TAL, O., BERGER, J., COOK, B., DANIELS, M., SCAVONE, G., AND COOK, P. 2002. SONART: The sonification application research toolbox. In *Proc. ICAD, Kyoto, Japan*.

BERTIN, J. 1983. *Semiology of Graphics*. The University of Wisconsin Press, Madison, WI.

BROWN, L. AND BREWSTER, S. 2003. Drawing by ear: Interpreting sonified line graphs. In *Proceedings of ICAD 2003*. ICAD, 152–156.

CHESTER, D. AND ELZER, S. 2005. Getting computers to see information graphics so users do not have to. In *ISMIS*. 660–668.

CHI, M. T. H., FELTOVICH, P. J., AND GLASER, R. 1981. Categorization and representation of physics problems by experts and novices. *Cognitive Science 5,* 2 (June), 121–152.

COHEN, R. F., YU, R., MEACHAM, A., AND SKAFF, J. 2005. Plumb: displaying graphs to the blind using an active auditory interface. In *Assets '05: Proceedings of the 7th international ACM SIGACCESS conference on Computers and accessibility*. ACM Press, New York, NY, USA, 182–183.

CORIO, M. AND LAPALME, G. 1998. Integrated generation of graphics and text: a corpus study. In *Proceedings of the COLING-ACL Workshop on Content Visualization and Intermedia Representations (CVIR'98)*, M. T. Maybury and J. Pustejovsky, Eds. Montréal, 63–68.

CORIO, M. AND LAPALME, G. 1999. Generation of texts for information graphics. In *Proceedings of the 7th. European Workshop on Natural Language Generation (EWNLG'99)*, P. Saint-Dizier, Ed. Toulouse, 49–58.

DALE, R. 2002. StockReporter. `http://www.ics.mq.edu.au/~ltgdemo/StockReporter/`, Web resource.

DEMIR, S., OLIVER, D., SCHWARTZ, E., ELZER, S., CARBERRY, S., AND McCOY, K. F. 2010. Interactive SIGHT demo: textual summaries of simple bar charts. In *Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility*. ASSETS '10. ACM, New York, NY, USA, 267–268.

DUMONTIER, M., FERRES, L., AND VILLANUEVA-ROSALES, N. 2010. Modeling and querying graphical representations of statistical data. *Web Semant. 8*, 241–254.

ELZER, S., SCHWARTZ, E., , CARBERRY, S., CHESTER, D., DEMIR, S., AND WU, P. 2007. A browser extension for providing visually impaired users access to the content of bar charts on the web. In *In Proceedings of International Conference on Web Information Systems (WEBIST)*.

ERICSSON, K. A. 2005. Recent advances in expertise research: A commentary on the contributions to the special issue. *Applied Cognitive Psychology 19*, 231–241.

EVREINOVA, T. G., EVREINOV, G., RAISAMO, R., AND VESTERINEN, L. K. 2008. Non-visual interaction with graphs assisted with directional-predictive sounds and vibrations: a comparative study. *Univers. Access Inf. Soc. 7*, 93–102.

FASCIANO, M. AND LAPALME, G. 1996. Postgraphe: a system for the generation of statistical graphics and text. In *Proceedings of the 8th International Workshop on Natural Language Generation (INLG-96)*. Herstmonceux, Sussex, UK, 51–60.

FASCIANO, M. AND LAPALME, G. 2000. Intentions in the coordinated generation of graphics and text from tabular data. *Knowledge and Information Systems 2,* 3, 310–339.

FERRES, L., DUMONTIER, M., AND VILLANUEVA-ROSALES, N. 2007. Semantic query answering with time-series graphs. In *Proceedings of the 3rd International Workshop on Vocabularies, Ontologies and Rules for The Enterprise (VORTE 2007)*.

FERRES, L., LINDGAARD, G., AND SUMEGI, L. 2010. Evaluating a tool for improving accessibility to charts and graphs. In *Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility*. ASSETS '10. ACM, New York, NY, USA, 83–90.

FERRES, L., PARUSH, A., ROBERTS, S., AND LINDGAARD, G. 2006. Helping People with Visual Impairments Gain Access to Graphical Information Through Natural Language: The iGraph System. In *Proceedings of the 10th International Conference on Computers for Handicapped Persons*. Lecture Notes in Computer Science. Springer–Verlag.

FERRES, L., VERKHOGLIAD, P., LINDGAARD, G., BOUCHER, L., CHRÉTIEN, A., AND LACHANCE, M. 2007. Improving accessibility to statistical graphs: the igraph-lite system. In *Assets '07: Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility*. ACM Press, New York, NY, USA, 67–74.

FERRES, L., VERKHOGLIAD, P., SUMEGI, L., BOUCHER, L., LACHANCE, M., AND LINDGAARD, G. 2008. A syntactic analysis of accessibility to a corpus of statistical graphs. In *Proceedings of the 2008 international cross-disciplinary conference on Web accessibility (W4A)*. W4A '08. ACM, New York, NY, USA, 37–44.

FLOWERS, J. H., BUHMAN, D. C., AND TURNAGE, K. D. 1997. Cross-Modal Equivalence of Visual and Auditory Scatterplots for Exploring Bivariate Data Samples. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 341–351.

FLOWERS, J. H., BUHMAN, D. C., AND TURNAGE, K. D. 2005. Data sonification from the desktop: Should sound be part of standard data analysis software? *ACM Trans. Appl. Percept. 2, 4*, 467–472.

GRABOWSKI, N. A. AND BARNER, K. E. 1998. Data visualization methods for the blind using force feedback and sonification. In *Proc. SPIE Vol. 3524, p. 131-139, Telemanipulator and Telepresence Technologies V, Matthew R. Stein; Ed.*, M. R. Stein, Ed. Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference, vol. 3524. 131–139.

GREEN, N., CARENINI, G., KERPEDJIEV, S., MATTIS, J., MOORE, J., AND ROTH, S. 2004. Autobrief: an experimental system for the automatic generation of briefings in integrated text and information graphics. *International Journal of Human-Computer Studies 61, 1*, 32–70.

KENNEL, A. R. 1996. Audiograf: a diagram-reader for the blind. In *Assets '96: Proceedings of the second annual ACM conference on Assistive technologies*. ACM Press, New York, NY, USA, 51–56.

KERPEDJIEV, S. 1995. Model-driven assertion-based generation of multimedia weather information. *Bulletin of the American Meteorological Society 76*, 1791–1800.

KERPEDJIEV, S., CARENINI, G., ROTH, S., AND MOORE, J. 1997. Autobrief: a multimedia presentation system for assisting data analysis. *Computer Standards & Interfaces 18*, 6-7, 583–593.

KILDAL, J. AND BREWSTER, S. A. 2006. Non-visual overviews of complex data sets. In *CHI '06 extended abstracts on Human factors in computing systems*. CHI EA '06. ACM, New York, NY, USA, 947–952.

KILDAL, J. AND BREWSTER, S. A. 2007. Interactive generation of overview information using speech. In *CHI '07 extended abstracts on Human factors in computing systems*. CHI EA '07. ACM, New York, NY, USA, 2489–2494.

KIPP, M. 2001. Anvil - a generic annotation tool for multimodal dialogue. In *Proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech)*. 1367–1370.

KRAMER, G., BONEBRIGHT, T., COOK, P., FLOWERS, J., MINER, N., AND NEUHOFF, J. 2010. Sonification report: Status of the field and research agenda, report prepared for the national science foundation,. Tech. rep., `http://www.icad.org/websiteV2.0/References/nsf.html`, Washington D.C., US.

KUKICH, K. 1983. Design and implementation of a knowledge-based report generator. In *Proceedings of ACL-1983*. 145–150.

KURZE, M. 1995. Giving blind people access to graphics (example: Business graphics). Darmstadt, Germany.

LARKIN, J. AND SIMON, H. 1987. Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science 11, 1*, 65–100.

LAW, A., FREER, Y., HUNTER, J., LOGIE, R., MCINTOSH, N., AND QUINN, J. 2005. A comparison of graphical and textual presentations of time series data to support medical decision making in the neonatal intensive care unit. *Journal of Clinical Monitoring and Computing 19*, 183–194.

MCGOOKIN, D. K. AND BREWSTER, S. A. 2006. Soundbar: exploiting multiple views in multimodal graph browsing. In *NordiCHI '06: Proceedings of the 4th Nordic conference on Human-computer interaction*. ACM Press, New York, NY, USA, 145–154.

MITTAL, V. O., CARENINI, G., MOORE, J. D., AND ROTH, S. 1998. Describing complex charts in natural language: a caption generation system. *Computational Linguistics 24, 3*, 431–467.

PORTET, F., REITER, E., HUNTER, J., AND SRIPADA, S. 2007. Automatic generation of textual summaries from neonatal intensive care data. In *Proceedings of AIME 2007*.

REITER, E. 1995. NLG vs. templates. In *Proceedings of the 5th European Workshop in Natural Language Generation*. Leiden, NL, 95–105.

REITER, E. 2007. An architecture for data-to-text systems. In *Proceedings of ENLG-2007*.

REITER, E. AND DALE, R. 2000. *Building Natural Language Generation Systems*. Cambridge University Press.

REITER, E., SRIPADA, S., HUNTER, J., YU, J., AND DAVY, I. 2005. Choosing words in computer-generated weather forecasts. *Artificial Intelligence 167,* 1-2, 137–169.

RIEDEL, B. 2001. Perception of gradient in haptic graphs: a comparison of virtual and hysical stimuli. In *Conference Proceedings to Eurohaptics.*

ROTH, S., MATTIS, J., AND MESNARD, X. 1991. Graphics and natural language as components of automatic explanation. In *Intelligent User Interfaces*, J. Sullivan and S. Tyler, Eds. Addison-Wesley, Reading, MA.

SHNEIDERMAN, B. 1996. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the 1996 IEEE Symposium on Visual Languages.* IEEE Computer Society, Washington, DC, USA, 336–.

SIMS, V. K. AND MAYER, R. E. 2002. Domain specificity of spatial expertise: The case of video game players. *Applied Cognitive Psychology 16*, 97–115.

SRIBUNRUANGRIT, N., MARQUE, C., LENAY, C., , AND GAPENNE, O. 2003. Improving blind people'sspatial ability by bimodal-perception assistive device for accessing graphic information. In *Proceedings of the 7th biannual Conference for the Advancement of Assistive Technology in Europe.* 476–480.

SRIPADA, S. AND GAO, F. 2007. Summarizing dive computer data: A case study in integrating textual and graphical presentations of numerical data. In *Proceedings of Workshop on Multimodal Output Generation.* Vol. CTIT Proceedings of the Workshop on Multimodal Output Generation. 149–157.

SRIPADA, S., REITER, E., AND DAVY, I. 2003. Sumtime-mousam: Configurable marine weather forecast generator. *Expert Update 6*, 4–10.

SRIPADA, S., REITER, E., DAVY, I., AND NILSSEN, K. 2004. Lessons from deploying nlg technology for marine weather forecast text generation. In *Proceedings of PAIS-2004.*

TUFTE, E. 2001. *The visual display of quantitative information*, Second Edition ed. Graphics Press, Cheshire, CT, USA.

TURNER, R., SRIPADA, S., REITER, E., AND DAVY, I. 2006. Generating spatio-temporal descriptions in pollen forecasts. In *Proceedings of EACL06 Companion Volume.*

VAN DEEMTER, K., KRAHMER, E., AND MARIET, T. 2005. Real versus template-based natural language generation: A false opposition? *Computational Linguistics 31,* 1, 15–24.

WALL, S. AND BREWSTER, S. 2006. Feeling what you hear: tactile feedback for navigation of audio graphs. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems.* ACM, New York, NY, USA, 1123–1132.

WELLS, L. R. AND LANDAU, S. 2003. Merging of tactile sensory input and audio data by means of the talking tactile tablet. In *In Eurohaptics 2003.* Media Lab Europe, Dublin, Ireland, 414–418.

YU, J., REITER, E., HUNTER, J., AND SRIPADA, S. 2003. Sumtime-turbine: A knowledge-based system to communicate time series data in the gas turbine domain. In *Proceedings of IEA/AIE-2003.* 379–384.

YU, W. AND BREWSTER, S. 2002. Comparing two haptic interfaces for multimodal graph rendering. In *HAPTICS '02: Proceedings of the 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems.* IEEE Computer Society, Washington, DC, USA, 3–9.

YU, W. AND BREWSTER, S. 2003. Evaluation of multimodal graphs for blind people. *Universal Access in the Information Society 2,* 2, 105–124.

YU, W., RAMLOLL, R., AND BREWSTER, S. A. 2001. Haptic graphs for blind computer users. In *Proceedings of the First International Workshop on Haptic Human-Computer Interaction.* Springer-Verlag, London, UK, 41–51.

APPENDIX 1: GRAPH QUESTIONS FOR THE FIRST FORMATIVE USABILITY STUDY

(1) (a) What's the unit of the values in vertical axis? (percentage)
   (b) How many years are reported? (3)
   (c) What year is the highest percentage reported? (2004)
   (d) Is the 6% mark surpassed in 2002? (no)
   (e) Is the graph generally going up or down? (Up)

(2) (a) What city is the graph reporting on? (Ottawa)
   (b) What year is the highest percentage of profits? (2006)
   (c) does the graph look like an upside-down V? (yes)
   (d) Is Ottawa better or worse in 2007 compared to 2005? (better)
   (e) How many years are reported? (3)

(3) (a) Is the graph generally going up or down? (down)
   (b) How many years are reported? (6)
   (c) E: what's the unit of the values in vertical axis? (Thousands of dollars)
   (d) In what year is tuition closest to the $4000 mark? (2004)
   (e) How much money did students pay in 2003? (6.5)

(4) (a) Is there a dip? (yes)
   (b) When does the dip occur? (2000)
   (c) When does it first increase above the 5 mark? (1999)
   (d) Does it increase or decrease from 2001 to 2002? (Decrease)
   (e) What's the last year reported in the graph? (2002)

APPENDIX 2: GRAPH QUESTIONS FOR THE SECOND FORMATIVE USABIOITY STUDY

(1) (a) How many years are reported? (15)
   (b) Is the graph generally going up or down? (up)
   (c) What country is the reporting on? (Canada)
   (d) Which census year saw a drastic increase in the number of heads of bison? (2001)
   (e) What is the unit of values in the vertical axis? (Thousands of bison)
   (f) In what census year were there the fewest number of bison in Canada? (1991)

(2) (a) What are the categories on the horizontal axis? (four seasons)
   (b) Which season saw the lowest percentage of government employee vacations? (summer)
   (c) Was the 10% mark ever surpassed? (no)
   (d) What year is this graph reporting on? (2007)
   (e) Is the graph generally going up or down? (down)
   (f) How many years are reported in the graph?
   (g) In which seasons do most people tend to take vacation? (winter)
   (h) What is the last point in the graph? (summer)

(3) (a) What is the title of the graph? (number of prescriptions filled)
   (b) In what year were the most prescriptions filled? (2006)
   (c) What is the first value in the graph? (100 million)

   (d) What is the unit of values in the vertical axis? (millions) 5. How many
       years are reported in the graph? (6)
   (e) What do you think is the main point of this graph? (it shows the trend in
       number of prescriptions filled)
(4) (a) What is the unit of values in the vertical axis? (thousands of dollars)
   (b) In which year was tuition closest to $6,000? (2002)
   (c) How many years are reported in the graph? (6)
   (d) What were the tuition fees in 2004? ($4,000)
   (e) What type of university program is this graph showing? (undergraduate
       engineering)
   (f) Based on the trend in the graph, would you expect that 2008's tuition fees
       would be higher or lower than it was in 2007? (lower)