To my guiding eyes, ASTER

Foreword By Prof. David Gries

Once in a while, something nice happens, as if by coincidence, serendipitously. It happened to me when T.V. Raman asked me to supervise his Ph.D. thesis on building a system to speak documents, especially those with technical content or a lot of structure.

The project had many interesting points, for example: the need for a programming language for writing speaking rules (a sort of postscript for the ear instead of for paper), the need for a multi-dimensional model of speech and sound (so that the speaking-rule language could be largely independent of the particular voice synthesizer being used), knowledge of mathematics and the development of ways to speak it well, the development of a new internal form for mathematics, software engineering (in his 35,000-line CLOS program AsTeR, Raman makes ingenious use of object-oriented features of CLOS), and human-interface issues in making AsTeR interactive.

Finally, there was a real *need* for AsTeR. Previously, it was almost impossible for the visually handicapped to access technical documents. RFB (Recordings For The Blind) takes up to a year to make a recording of a technical book, and the only other way for a visually handicapped person to access a technical document is to have someone read it to them. The development of AsTeR is one more instance of the usefulness of computers.

Recordings For The Blind held a workshop-conference on AsTeR three months after the thesis was completed and will soon begin producing cassette tapes using it. Not only can a tape of a book be made in a day or two (instead of a year), but the quality is far superior to that produced by humans. In addition, Recordings For The Blind would like to make AsTeR available to the blind who are computer-literate, allowing them to make full use of AsTeR's interactive capabilities.

The applications of the concepts and technology in AsTeR are many. For one, just consider being able to call your computer on the phone and have your e-mail read to you in a comprehensible fashion. The results are so noteworthy that *SIAM News* published a column on Raman and AsTeR (March 1994, page 7).

What makes this thesis even more intriguing is that Raman himself is visually handicapped. He cannot see his keyboard or monitor. He uses a guide-dog, Aster, to help him get around. And yet, Raman did *all* the programming of system AsTeR himself. My role was that of advisor. We discussed issues, I gave him advice based on my experience, I pointed him at literature, and I helped him in editing and organizing the thesis itself and a few papers that came from it (though all the writing on the computer was done by Raman himself). The real work, however, was his alone.

Raman is one of the most courageous, up-beat, and positive people I have known. Blind, he came from India to do his PhD work in Cornell's field of applied mathematics. He acquired his guide dog, Aster, after arriving here. He studied applied math and computer science and began using computers

in a big way (with a screen reader). All this without a complaint, with the most positive and optimistic attitude I have seen.

Just consider this anecdote. After his PhD, Raman accepted an offer at a major computer manufacturer. While still at Cornell, surfing the net, Raman came across a notice on a bulletin board to the following effect. Jack's (a fictitious name) job was to go to T.V.'s future department and tell them how to treat Raman (because of his handicap). Jack, however, had little real experience with such matters, and he wanted advice from whomever would give it to him —what should he tell the department? Raman came into my room laughing about it. He had replied to Jack: tell the department to treat him just the way they would treat anyone else!

ASTER is indeed a wonderful achievement. For me, however, ASTER was not the most important part of this endeavor. Instead, it was working daily with Raman over a two-to-three year period. Always positive, with never a complaint, brilliant in his own way, Raman was a pleasure to deal with. It is an experience I will value for years to come.

David Gries
William L. Lewis Professor of Engineering
Computer Science Department
Cornell University
Ithaca, NY

Preface

> The advent of electronic documents makes information available in more than its visual form —electronic information can now be display-independent. We describe a computing system, AsTeR, that *audio formats* electronic documents to produce audio documents.

The development of AsTeR was the basis of the author's dissertation, which was presented to the Faculty of the Graduate School of Cornell University in fulfillment of the Requirements for the Degree of Doctor of Philosophy in 1994. This preface, written three years later, puts the work in perspective with respect to the developments in the world of electronic information and auditory interfaces between 1994 and 1997. The main body of this work remains identical to what was presented in 1994.

Introduction

AsTeR was motivated by the insight that information presentation needs to take advantage of the specific perceptual modality in use. This typeset manuscript exploits features of visual interaction to convey information effectively; in the same vein, AsTeR introduced the notion of *audio formatting* to enable rich aural presentations of structured information.

Speech-enabling Applications

The insights gained from developing and using AsTeR have been applied to the more general problem of providing aural access to computer interfaces, starting in late 1994. Computer interfaces encapsulate man-machine dialogue, and once we realize that "The document *is* the interface", the technique of synthesizing effective aural presentations starting with the information instead of its visual presentation leads naturally to the speech-enabling approach —see [Ram96a, Ram96b, Ram97b]. The speech-enabling approach —a technique that separates computation from the user interface —is described in detail in [Ram97a]. Application designers can implement desired features in the computational component and have different user interfaces expose the resulting functionality in the manner most suited to a given user environment. This leads to the design of high-quality Auditory User Interfaces (AUI) that integrate speech as a first class citizen into the user interface.

Structured Information And The WWW

AsTeR pointed out the advantages to come in a world where documents are first created electronically before being turned into modality-specific presentations such as typeset documents for printing. The work also pointed out

the need for such electronic information to be well-structured to enable *computation* on this information.

The last few years have seen an explosive growth in electronic information on the Internet fueled by the popularity of the WWW. The initial rush to the WWW resulted in publishers putting out rich visual content with a concomitant abuse of document structure as envisioned in AsTeR. As a consequence, content providers on the WWW today face many of the challenges outlined in AsTeR when attempting to create electronic content that can be repurposed for publishing online as well as in traditional print formats. This has also led to a vast amount of Webformation that is becoming increasingly difficult to navigate and categorize —see [Hay96, Gib96].

Faced by these challenges, content providers on the WWW are now looking to create richly tagged information using markup systems like XML. As the first such example, mathematical Markup Language (MathML) is an XML application for describing mathematical notation and capturing both its structure and content. The goal of MathML is to enable mathematics to be served, received, and processed on the Web, just as HTML has enabled this functionality for text —*URL* http://www.w3.org/TR/WD-math/.

Conclusion

As humans, we see, hear, feel and smell. Human interaction is enriched by the concomitant redundancy introduced by multimodal communication. In contrast, computer interfaces until now have relied primarily on visual interaction —today's interfaces are like the *silent* movies of the past! As we approach the turn of the century, computers now have the ability to talk, listen and perhaps, even *understand*. Integrating new modalities like speech into human-computer interaction requires rethinking how information systems are designed in today's world of *visual* computing.

Visually rich computing introduced the notion of What You See Is What You Get (WYSIWYG) documents; but by carrying it too far, we risk ending up in a world of "What You See Is All You Have" documents. On the positive side, the exponential growth in electronic information combined with a desire to be able to intelligently process this content and access it whenever, wherever and however the user chooses provides adequately strong reasons to suggest that the world will move away from the present situation of *see-only* documents.

That a blind person can navigate the Internet just as efficiently and effectively as any sighted person attests to the profound potential of digital documents to improve human communication. Printed documents are fixed snapshots of changing ideas; they limit the means of communication to the paper on which they are stored. But in electronic form, documents can become raw material for computers that can extract, catalogue and rearrange the ideas in them. Used properly, technology can separate the message from

the medium so that we can access information wherever, whenever and in whatever form we want.

Archiving information in a structurally rich form will ensure that this vast repository of knowledge can be reused, searched and displayed in ways that best suit individuals' needs and abilities, using software not yet invented or even imagined.

The coming millenium is likely to prove an exciting one in the world of electronic information.

T. V. Raman

December 6, 1997 Mountain View, CA. *URL* http://cs.cornell.edu/home/raman

Acknowledgements

I thank my adviser, David Gries, for his help and guidance in turning a collection of useful ideas into a practicable thesis. His insight into defining a language for audio formatting proved crucial in realizing my ultimate goal of producing a system that does for audio documents what systems like (LA)TEX have achieved in the world of printed documents. I also acknowledge the help and support of the other members of my committee, John Hopcroft, Dan Huttenlocher, Dexter Kozen, Keith Dennis and John Guckenheimer.

My former office-mate, M. S. Krishnamoorthy (RPI), was the first to spot the potential presented by my prototype, TEXTALK. He, along with John Hopcroft, Keith Dennis and Brian Kernighan (ATT), encouraged me to take up the problem of producing audio renderings from electronic markup source for my dissertation. Tim Teitelbaum and Anne Neirynck helped in the initial phase when I was defining the problem. Bruce Donald was my adviser during the first phase of the project. We had many useful discussions, and I am grateful to him for convincing me to implement my system in Lisp-CLOS. Bruce Donald and CSRVL (Computer Science Robotics and Vision Lab) supported my work with a research assistantship and equipment.

My summer experience at the Xerox Palo Alto Research Center (PARC) helped me crystalize many ideas. Dennis Arnon of Xerox PARC pointed out the importance of working with document logical structure. Xerox Corp. also supported my work with an equipment grant in spring 1992. Jim Davis (Cornell DRI) advised me on lexical choice when producing spoken mathematics, helped improve my Lisp programming skills, and also contributed some Lisp code used to communicate with the speech synthesizer.

Intel Corp. supported my work with a one-year fellowship for the academic year 1992-93 and a research grant for fall 93. I acknowledge the help and support of my Intel mentor, Murali Veeramoney, and the other members of his group. Jim Larson (Intel Architecture Labs) helped me crystalize some of my ideas on user-interface design during the many stimulating discussions we had over the summer of 1993.

I implemented ASTER and wrote this thesis using an Intel-486 PC from CSRVL running IBM Screen Reader. I thank the Center for Applied Mathematics (CAM) for opening up the world of computing to me by acquiring an Accent speech synthesizer and IBM Screen Reader —Screen Reader, de-

signed by Jim Thatcher (IBM Watson Research Center), is one of the most robust screen-reading programs available today. I acknowledge the support of our systems administrators for their untiring help in my efforts to adapt my setup to use the software available.

I also thank the USENET community for their support in helping configure the various pieces of software that I use. The *Emacs* editor and Screen, a public-domain window manager for ASCII terminals, have together provided a powerful computing environment that has enabled me to be fully productive. Lack of online documentation for Lisp was overcome with help from the USENET (comp.lang.lisp) community. I also thank Nelson Beebee for his invaluable help on (LA)TEX throughout the writing of this thesis. I thank the authors and publishers of the texts listed in Table B for providing me online access to the electronic sources —these proved invaluable both as online references as well as test material for ASTER.

Taking classes at Cornell was an enjoyable experience, and I thank all the faculty for their help. Every effort was made to provide online lecture notes —ASTER was motivated by the availability of online notes for CS681 taught by Dexter Kozen. Talking books from Recording for the Blind (RFB) proved invaluable. I was also ably assisted by a dedicated group of readers. Anindya Basu, Bill Barry (ORST), Jim Davis, Harsh Kaul and Matthai Phillipose proof-read this thesis and suggested many useful improvements. I also thank Holly Mingins, Dolores Pendell (CAM) and the rest of the administrative staff of the CS department for their help and support. I thank Bert Adams of the Cornell Physical Education program for helping me stay fit during the last four years and Mike Dillon (NYSCBVH) for orienting me around the Cornell campus.

Finally, I thank my family for their love and support throughout.