

## ITECH: AN INTERACTIVE TECHNICAL ASSISTANT

Except where reference is made to the work of others, the work described in this dissertation is my own or was done in collaboration with my advisory committee. This dissertation does not include proprietary or classified information.

---

Dale-Marie Wilson

### Certificate of Approval:

---

Homer Carlisle  
Associate Professor  
Computer Science and Software  
Engineering

---

Juan E. Gilbert, Chair  
Associate Professor  
Computer Science and Software  
Engineering

---

Cheryl Seals  
Assistant Professor  
Computer Science and Software  
Engineering

---

Ivan Watts  
Associate Professor  
Educational Foundations Leadership  
and Technology

---

Stephen L. McFarland  
Acting Dean  
Graduate School

ITECH: AN INTERACTIVE TECHNICAL ASSISTANT

Dale-Marie Wilson

A Dissertation

Submitted to

the Graduate Faculty of

Auburn University

in Partial Fulfillment of the

Requirements for the

Degree of

Doctor Philosophy

Auburn, Alabama

August 7, 2006

# ITECH: AN INTERACTIVE TECHNICAL ASSISTANT

Dale-Marie Wilson

Permission is granted to Auburn University to make copies of this dissertation at its discretion, upon the request of individuals or institutions and at their expense. The author reserves all publication rights.

---

Signature of Author

---

Date of Graduation

## DISSERTATION ABSTRACT

### ITECH: AN INTERACTIVE TECHNICAL ASSISTANT

Dale-Marie Wilson

Doctor of Philosophy, August 7, 2006  
(M.S., Auburn University, 2003)  
(B.S., New York University, 1995)

124 Typed Pages

Directed by Juan E. Gilbert

This dissertation concentrates on the problem of designing and developing a conversational technical assistant. The main focus is to identify and address issues related to producing a system that allows for conversational question answering that utilizes a new methodology called Answers First. Additionally the mechanics of translating traditional technical communications into a knowledge base of question-answer pairs that will allow for effective information retrieval, pose another significant challenge, especially as the technical communications will be manufacturer-independent. The resulting system should enable manufacturers to provide a new medium to their consumers for technical assistance. In this proposal, a prototype of a conversational technical assistant is developed using the technical communications from a vi manual

published by O'Reilly. The approaches used to improve upon the aforementioned issues are described in detail. Through experiments performed on the developed application the performance and potential of the selected approaches will be evaluated.

## ACKNOWLEDGMENTS

First and foremost I would like to thank Jesus Christ, my Lord and Savior, through whom all things have been made possible. Next, I would like to express my deepest gratitude to my advisor, Dr. Juan E. Gilbert, for his patient guidance, valuable advice and continued encouragement throughout my graduate studies. He has been a safe harbor through the storms, providing me with numerous opportunities. He got me started on this road to higher academia and has been at my side throughout. I would also like to thank my graduate committee members, Dr. Homer Carlisle, Dr. Cheryl Seals and Dr. Ivan Watts for their reviewing and advising efforts. In addition, I would like the following for their contributions and assistance during this process, Ernest Cross, Yolanda McMilian and Kenneth Rouse. Special thanks go out to my family for believing in me and encouraging me in my decision to pursue my goals: to my son Daniel, thank you for tolerating a cranky mother after she stayed up all night doing research; to my mother Maureen, thank you for the example you set as a strong Christian woman and mother, for your continued support and unwavering belief in me; to my friends: Jennifer DeLeon, Natasha Lamy-Ramsden and Arlette Scapino, thank you for your continued friendship, love and prayers.

Style manual or journal used Journal of SAMPE

Computer software used Microsoft Word 2002

## TABLE OF CONTENTS

LIST OF FIGURES .....	xii
LIST OF TABLES .....	xiv
1. INTRODUCTION .....	1
1.1. MOTIVATION .....	1
1.2. TECHNICAL COMMUNICATIONS .....	3
1.2.1. INTERACTIVE ASSISTANTS .....	4
1.2.2. CONVERSATIONAL QUESTION ANSWERING .....	5
1.3. PROBLEM DEFINITION .....	6
1.3.1. PROBLEM DESCRIPTION .....	6
1.4. TECHNICAL ASSISTANT .....	10
1.4.1. ITECH .....	11
1.5. GOALS, APPROACHES AND CONTRIBUTIONS .....	11
1.6. ORGANIZATION .....	12
2. LITERATURE REVIEW .....	14
2.1. TECHNICAL COMMUNICATION .....	14
2.2. PAPER COMMUNICATION .....	15
2.2.1. ONLINE MANUALS/ASSISTANCE .....	16
2.2.2. RESEARCH DEVELOPMENTS .....	17
2.3. AUTOMATIC SPEECH RECOGNITION .....	20
2.3.1. SPOKEN USER INTERFACE .....	21
2.4. INFORMATION RETRIEVAL .....	23
2.5. NATURAL LANGUAGE PROCESSING .....	23



2.5.1.	NATURAL LANGUAGE INTERFACE TO DATABASES .....	25
2.5.2.	NATURAL LANGUAGE QUESTION ANSWERING .....	26
2.5.2.1.	QUESTION ANSWERING USING STATISTICAL MODEL.....	26
2.5.2.2.	PROBABILISTIC PHRASE RERANKING.....	27
2.5.2.3.	BAYESIAN APPROACH.....	28
2.5.3.	NLQA VS. NLIDB .....	29
3.	SYSTEM DESIGN .....	30
3.1.	ITECH.....	30
3.2.	SPEECH USER INTERFACE (SUI) .....	31
3.2.1.	ANSWERS FIRST .....	34
3.3.	DESIGN PRINCIPLES .....	38
3.3.1.	SYSTEM ARCHITECTURE .....	38
3.3.2.	MULTIMODAL INTERFACE .....	40
3.3.3.	KNOWLEDGE REPOSITORY (KR) .....	42
3.3.3.1.	DATABASE TABLES .....	42
3.3.3.1.1.	ANSWERS .....	43
3.3.3.1.2.	ANSWERTYPE.....	43
3.3.3.1.3.	CATEGORIES.....	43
3.3.3.1.4.	CATEGORYTYPE.....	44
3.3.3.1.5.	QUESTIONS .....	44
3.3.3.1.6.	TERMS .....	45
3.3.3.2.	ENTITY RELATIONSHIP MODEL .....	45
3.4.	SAMPLE INTERACTIONS.....	46

4.	IMPLEMENTATION RESULTS .....	51
4.1.	EXPERIMENT DESIGN .....	51
4.2.	EXPERIMENTAL SETTINGS .....	52
4.2.1.	MATERIALS.....	52
4.2.2.	PARTICIPANTS AND PROCEDURE .....	53
4.3.	DATA COLLECTION METHODS .....	56
4.3.1.	PRE-EXPERIMENT QUESTIONNAIRE .....	57
4.3.2.	PERFORMANCE DATA AND USER OBSERVATIONS.....	58
4.3.3.	POST-EXPERIMENT QUESTIONNAIRE .....	58
4.4.	RESULTS AND DISCUSSION .....	59
4.4.1.	PARTICIPANT BACKGROUND .....	59
4.4.2.	PERFORMANCE DATA FINDINGS .....	61
4.4.2.1.	SPOKEN QUERY METRICS .....	68
4.4.3.	USER SATISFACTION .....	70
4.5.	EXPERIMENT SUMMARY .....	83
5.	SUMMARY .....	85
5.1.	CONTRIBUTIONS .....	85
5.2.	DIRECTIONS FOR FUTURE RESEARCH .....	86
	REFERENCES .....	88
	APPENDIX A .....	98
	APPENDIX B .....	99
	APPENDIX C .....	100
	APPENDIX D .....	101

APPENDIX E .....	103
APPENDIX F.....	106
APPENDIX G.....	108

## LIST OF FIGURES

Figure 1: Query Formulation .....	8
Figure 2: Grammar size vs WER.....	9
Figure 3: Conceptual Models.....	32
Figure 4: Excerpt from iTechGrammar.grmxl.....	34
Figure 5: Scenario #1 .....	35
Figure 6: Scenario #2.....	36
Figure 7: Scenario #3 .....	37
Figure 8: Scenario #4.....	37
Figure 9 System Architecture .....	40
Figure 10: Entity Relationship Model.....	46
Figure 11: iTech Welcome Screen.....	47
Figure 12: Example of iTech's Welcome.....	47
Figure 13: iTech is Listening .....	48
Figure 14: iTech displaying a Solution.....	49
Figure 15: iTech Dialogue .....	50
Figure 16: Side-by-Side Box Plots of Medium Search Times.....	62
Figure 17: Tukey-Kramer Test Results.....	64
Figure 18: Side-by-Side Box Plots of Medium Task Completion Time .....	65
Figure 19: Side-by-Side Box Plots of Medium Read Times .....	66

Figure 20: Excerpt of Transcribed Queries.....	69
Figure 21: Examples of keyword searches .....	69
Figure 22: Book Medium Wonderful--Terrible Bi-polar Distribution .....	71
Figure 23: Online Medium Wonderful--Terrible Bi-polar Distribution .....	72
Figure 24: iTech Medium Wonderful--Terrible Bi-polar Distribution.....	72
Figure 25: Book Medium Dull--Stimulating Bi-polar Distribution.....	73
Figure 26: Online Medium Dull--Stimulating Bi-polar Distribution .....	74
Figure 27: iTech Medium Dull--Stimulating Bi-polar Distribution .....	74
Figure 28: Book Medium Boring--Fun Bi-polar Distribution .....	75
Figure 29: Online Medium Boring--Fun Bi-polar Distribution.....	75
Figure 30: iTech Medium Boring--Fun Bi-polar Distribution.....	76
Figure 31: Book Medium Affordance Distribution .....	78
Figure 32: Online Medium Affordance Distribution .....	78
Figure 33: iTech Medium Affordance Distribution.....	79
Figure 34: Book Medium Getting Started Distribution .....	80
Figure 35: Online Medium Getting Started Distribution.....	81
Figure 36: iTech Medium Getting Started Distribution.....	81

## LIST OF TABLES

Table 1: Medium Type.....	55
Table 2 Experimental Instruments and Measures .....	57
Table 3: Participant Background Data .....	60
Table 4: Mean Search Times by Medium.....	61
Table 5: Mean Task Completion Time by Medium.....	64
Table 6: Mean Read Times by Medium .....	66
Table 7: Bi-polar Rating Scales Assessing General Usability.....	70
Table 8: Rating Scales Assessing Cognitive Modeling .....	82
Table 9: Rating Scales Assessing Perceived System Response Accuracy .....	82
Table 10: Rating Scale Assessing Cognitive Demand.....	82
Table 11: Rating Scale Assessing Likeability .....	82
Table 12: Rating Scale Assessing Habitability .....	83
Table 13: Rating Scale Assessing Speed .....	83

## **1. INTRODUCTION**

### **1.1. MOTIVATION**

Today, almost every product or device is accompanied by a manual. These manuals are included to provide assistance to the consumer. This assistance typically falls within three categories: functionality, maintenance and repair. However as Thimbleby states, “User manuals are the scapegoat of bad system design.” [Maj85]; the resulting experience of the consumer is far from desirable. The experience usually entails timely searching through a large paper manual or rigorous cognitive processing to generate the appropriate query to be applied to an online manual. As a result, the level of performance of the current mediums used for technical communications must be addressed.

There are several mediums in which technical communications are currently provided. They range from paper to interactive animation and virtual reality [Hai04], with each new medium attempting to improve upon the drawbacks of the previous one. The first medium introduced was the paper manual. The issues with this medium have been widely documented, especially by technicians in the armed forces. The problems include lack of portability, inaccuracy, and increasing content and complexity [Ven88]. Then, as the popularity of the Web grew a new medium was introduced, online manuals.

Online assistance reduces the geographical distance between the user and technical documentation and as a result reduced the portability issue. This new medium also led to a shift towards increased user satisfaction. The requirements for efficient user

assistance now included unobtrusiveness; context-sensitivity; consistency and preciseness. Development of online manuals now includes questions like ‘For which class of users is the assistance provided?’; ‘Should the user control when assistance is provided?’ and ‘From the users’ standpoint, what are the assistance requirements?’ [RP81]. A trend towards more user-centered development was occurring. These trends initiated investigations into how users access the information and also led to a collaborative approach to the field.

Technical communications is now a highly-interdisciplinary field. Contributions from several fields including Social Sciences, Technical Documentation, Human-Computer Interaction and Business Information Systems are now occurring [ZCCF01]. The progress and advancements in these individual fields are being filtered into the development of technical documentations through collaborations. This has led to the development of hypermedia and interactive applications with further investigations into the development of virtual reality and interactive animations as documentation tools. However, among these applications, the need for a natural interactive solution still exists.

During the conceptualization and development of iTech, the interactive technical assistant, the main research areas investigated include technical communications, interactive assistants and natural language question answering (NLQA).



## **1.2. TECHNICAL COMMUNICATIONS**

As previously mentioned, there are several mediums through which technical communications are currently presented. Currently, the most popular mediums are paper and online documentation [ZCCF01]. How the information is organized within these two mediums warrants further investigation. Information within a book manual is organized by topic or keyword. The topics are presented via the table of contents and the keywords are sorted alphabetically in the index. With respect to online assistance, this information is organized by topic. Retrieval of information from online assistance requires the presentation of a query, typically comprised of keywords. The query is executed and a ranked list of topics containing the query's keywords is presented [Bar04]. Thus for the success of either the paper or online medium, the user's information needs must be presented in either keyword or topic form. However, when people seek information, we instinctively form a mental question. If the information source is another person, we then speak this question as it was formed in our natural language. If the information source is a paper or online manual, further cognitive processing is required. This processing entails the translating of the natural question into a format that the technical communication medium understands. For a book manual, translation of the question into a topic or keyword must occur. For online assistance, translation of the question into a query occurs. Thus usage of these mediums does not afford a person's natural information-seeking process. This dissertation concentrates on this problem and proposes a medium

for technical communications that accommodates our natural information-seeking process without additional cognitive processing. This medium is an interactive technical assistant.

### 1.2.1. INTERACTIVE ASSISTANTS

Interactive assistants aim to assist users in managing their environment [KR01]. Computers are becoming more and more ubiquitous. They are permeating even more aspects of people's daily lives. Therefore the need for an efficient interface between users and their computers exists. This interface is presented in the form of an interactive assistant. To provide the most natural interaction, these assistants typically contain multimodal features including speech input and output, gesture and handwriting recognition and animated agents or avatars. These features provide users with interaction choices that can circumvent personal and/or environmental limitations. They also have great potential to promote new forms of computing and expand the accessibility of computing to a diverse group of users [OCVD00]. Although the notion, that speech and pointing is the dominant interaction style in multimodal interfaces, is listed as one of the "Ten Myths of Multimodal Interaction" [Ovi99], with respect to iTech, speech is the dominant mode of input. Studies have shown that in query formulation, the translation of a user's search expression into a query, spoken queries were found to be lengthier than their written counterparts [DC04]. This difference between spoken and textual

communications forms the basis for the methodology of conversational question answering used in iTech.

### 1.2.2. CONVERSATIONAL QUESTION ANSWERING

As the use of spoken user interfaces or speech-based user interfaces is rapidly growing, the development of spoken dialogue systems has gained popularity. Towards this research, there are ongoing investigations into not only improving the effectiveness of these applications but also into reducing the effort required to develop spoken dialogue systems. Any spoken dialogue system can be considered conversational to some degree. The extent of this conversational aspect can be categorized by the autonomy of the user's speech and the system's control over the conversation [GWCP04]. As systems become more conversational, with respect to the user, there is increased flexibility over the user's needs, how they ask for what they want and when they can ask for what they want. To allow for this flexibility, increased effort on the part of natural language processing is required.

Current techniques for natural language processing include parts-of-speech tagging, syntactic parsing, semantic interpretation and the use a statistical model. This research will present a new methodology, called Answers First (A1) that bypasses these traditional techniques and uses a bi-gram resolution. The following section will explore the problems for developing an interactive technical assistant in more detail.

### **1.3. PROBLEM DEFINITION**

Interactive technical assistants are an emerging concept that requires contributions from several research fields: technical documentation, human computer interaction, animated agents, information retrieval (IR), spoken user interface and natural language question answering (NLQA). This research will focus on the development of a conversational interactive agent that provides technical assistance. This development will build upon findings, solutions and results provided by the existing literature on technical documentation, interactive assistants, NLQA and IR. In the process, focus will be on the following issues: 1) the efficiency, effectiveness and accuracy of conversational question answering; and 2) the adaptability and scalability of existing technical documentation to natural language answers.

#### **1.3.1. PROBLEM DESCRIPTION**

The development of iTech has four major properties which contribute to the difficulty in proposing a solution. These properties are:

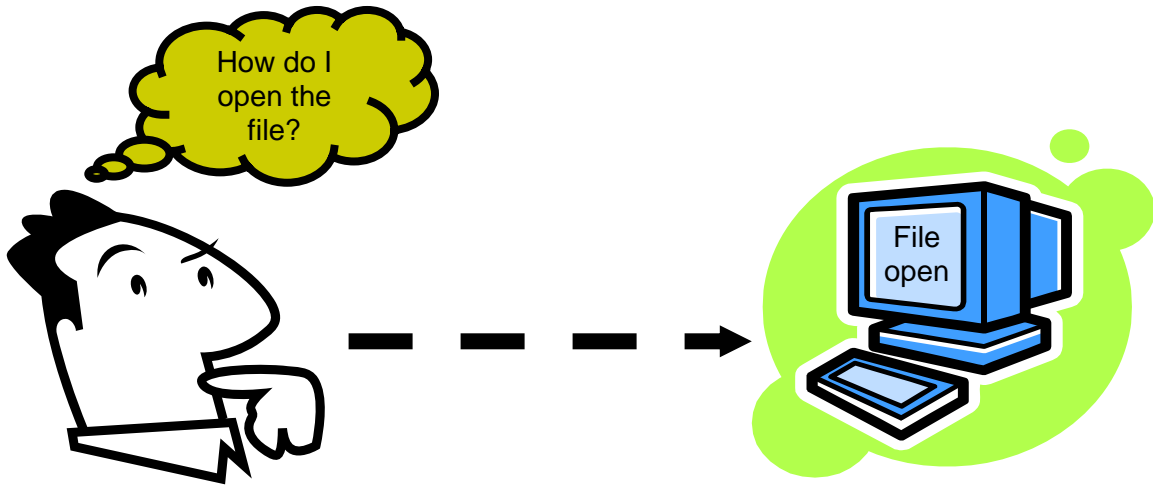
1. Limitations of current technical documentation. These include understandability, portability, accessibility, accuracy, updatability and search time. Each limitation does not apply to every medium used for technical communication, however for each current medium there is an applicable limitation.
2. The accuracy of the automatic speech recognition (ASR) engine especially in the domain of natural language queries. Speaker-independent speech recognition

engines, which this research employs, have a higher word error rate (WER) than those that are trained. This error rate is reduced by allowing for a limited grammar by the system. However, the requirements for natural language questions necessitate a large grammar. Therefore, the introduction of natural language questions will increase the standard WER.

3. Population of the database with answer-question pairs generated from the book manual. To allow for answers to the natural language questions, each solution must be matched with its relevant question and an answer is not restricted to one specific question. This matching has been previously achieved by using the results of studies. In these studies the user interacted with the system and these interactions were recorded. Thus, the questions and the responses that proved correct or most appropriate were identified. This process is timely and very expensive.
4. Success of conversational question answering. Current systems perform language processing like parts-of-speech tagging and semantic interpretations on the retrieved question, formulating a query that is executed. This processing is replaced in iTech by the Answers First (A1) approach.

The majority of shortcomings of technical communications listed in the first property require changes during the developmental period of the documentation. For the issue of search time, improvements can be accomplished outside of the developmental stage as

the predicament with search time occurs during query formulation. Query formulation is the process by which the user translates their information needs into a search expression [DC04], *Figure 1: Query Formulation* contains an illustration of the process.

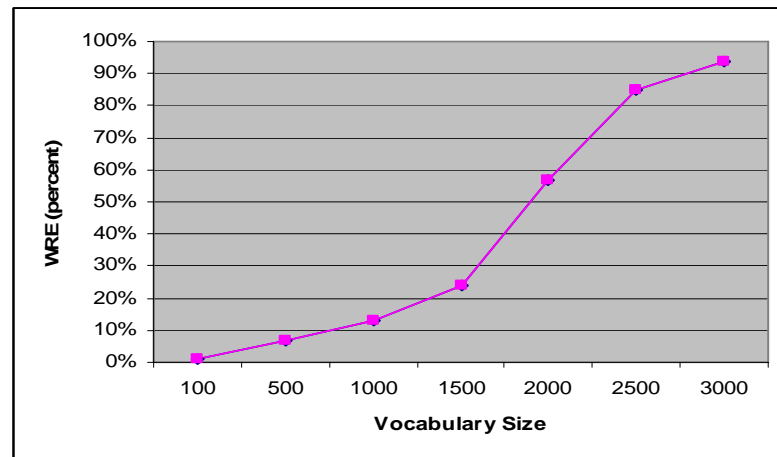


**Figure 1: Query Formulation**

This research eliminates the need for query formulation by accepting the user's information needs in its natural form. The elimination of this cognitive processing step thereby not only improves the user's perception of the application, but also reduces the user's search time. As a result, iTech allows the user to speak their information needs exactly as they are mentally generated and recognizes them. The recognition of this natural language produces the second property.

There are two types of speech recognition engines: speaker dependent and speaker independent. Speaker dependent engines require training by the user and will only return high recognition rates for the trained user. While speaker independent engines

require no unique training and allow anyone to use the engine and be recognized; the drawback is reduced accuracy [ZLBDH03]. The added feature of natural language query acceptance increases the size of the grammars. Consequently, as grammar sizes increase, so to do the word recognition errors increase [GZ03]. As shown in *Figure 2: Grammar size vs WER*, when the grammar size reaches three thousand words then the WER surpass 90% [GZ03]. To circumvent the potential exponential increases in grammar size with each new question introduced, unique word pairs were used to generate the grammar. This grammar is customized with respect to the question database.



**Figure 2: Grammar size vs WER<sup>1</sup>**

iTech's knowledge is stored in question-answer pairs in a database. This repository is generated from the existing technical documentation, which for this research is a vi manual. Translation of this manual into question-answer pairs presents a new issue.

The final property mentioned is critical for any conversational interactive system. Once the question is recognized, the goal of the system is to answer that question. Therefore, iTech must understand the question asked and retrieve the correct answer. Currently systems that perform NLQA use statistical methods and language processing to generate a query that is executed.

To address the properties mentioned previously, iTech was developed and evaluated. However before development could begin, the justification for iTech was determined via interviews with the projected subject pool.

#### **1.4. TECHNICAL ASSISTANT**

Interviews with 10 subjects were conducted to determine the current opinion of manuals. This interview was recorded and included questions that not only determined the shortcomings during the users' previous experiences with manuals, but also the potential interest in an interactive technical assistant, see Appendix D. The results ranked the main problems as follows:

1. Understandability
2. Presentation
3. Poor quality diagrams
4. Color

---

<sup>1</sup> Figure from reference [GZ03].



Also, 95% of participants pooled showed interest in an interactive technical assistant. Consequently, iTech was developed.

#### 1.4.1. ITECH

iTech is domain-specific and was developed from the vi manual entitled ‘Learning the vi Editor’, published by O’Reilly. The application contains the following features:

- ASR that recognizes natural language
- Knowledge repository of paired natural language questions to natural language answers housed in a database
- Answers First as conversational question answering methodology
- Multimodal interface

### 1.5. GOALS, APPROACHES AND CONTRIBUTIONS

The major goal of this research was to design and develop an interactive technical assistant which would accept natural language questions and provide answers to the users’ questions. Such a system should be suitable for any personal computer i.e. desktops, laptops and tablets. There are several limitations that make the design and development of such an interactive agent difficult especially for the natural language processing. This research first identified and addressed these issues and limitations which affect the design and development of a conversational technical assistant. For example,

are spoken queries as effective as written queries and are therefore feasible modes of interaction? Rules governing the generation of natural language grammars to automatic speech recognition engine were investigated and applied, followed by the investigation and application of a new natural language question answering approach to improve the search time of a solution. Finally, a prototype conversational technical assistant was developed and evaluated.

This research made the following contributions to the fields of Technical Documentation, Natural Language Processing and Human-Computer Interaction:

- Identified and addressed the limitations of technical documentation.
- Introduced a new medium for technical communications that improves on existing mediums.
- Introduced a novel methodology for conversational question answering called Answers First.
- Provided more evidence for the multi-disciplinary approach towards technical communications.
- Demonstrated that the research goals were successfully achieved by conducting a formal usability study.

## **1.6. ORGANIZATION**

In the chapters that follow, a research agenda will be examined. Chapter 2 gives an overview of the areas of research that pertain to the development of iTech. These areas

are Technical Documentation, Natural Language Processing and Understanding, Automatic Speech Recognition, Interactive Agents and Spoken Information Retrieval. Chapter 3 will discuss the system design and implementation. Following the design and implementation, an illustration of how the research goals were achieved will be given by discussing the details of the experimental results in Chapter 4. Chapter 5 discusses the conclusions of the study, including limitations; summarizes main contributions that this work made and points to some issues for future work.

## **2. LITERATURE REVIEW**

This research is wide ranging, with one of its main contributions being to provide more evidence towards the multi-disciplinary approach to technical communications. The research areas of Technical Communications, Automatic Speech Recognition, Natural Language Processing, Information Retrieval and Interactive assistant will be discussed. This highly focused literature review will concentrate on the aspects of these areas that directly impact this research.

### **2.1. TECHNICAL COMMUNICATION**

Technical communication refers to the process of delivering technical information to the user. Albing defines it as, “..the creation, control, delivery, and maintenance of a distributed information across the enterprise.....” [Alb96]. Technological advances have led to a restructuring within this field. Steps towards a more interdisciplinary approach are being made. Collaborations with differing fields are now required to develop more effective communications. These fields include, but are not limited to, psychology, ergonomics, human-computer interaction and instructional technology. An effective technical document is determined by the following factors [ZCFC01]:

1. How complete is the analysis of the communication problem?
2. How clearly identified is the goal/task to be explained?

3. How comprehensive, while following the conventional guidelines of technical writing, is the vocabulary used to explain the goal?

These factors are used for evaluating all mediums of technical communications.

## **2.2. PAPER COMMUNICATION**

As the need for manuals grew exponentially, few companies responded with proportional investments into the development of these manuals. This led to paper manuals that are outdated, hard to understand, inaccessible, incomplete, inefficient, and inaccurate, with poor maintenance quality and no portability. These characteristics are due to the following problems:

1. Portability – the increasing volume and weight of paper manuals make them harder to transport, harder to store and harder for technicians to carry.
2. Accuracy – the lag between development and documentation has led to outdated, incomplete and inaccurate manuals. This erroneous information is then used to make decisions.
3. Complexity – the increasing complexity of computer systems and everyday devices require more complex manuals. This increased complexity affects both the developer and the consumer. The developer must now spend more time formatting the increased information, while the consumer/user has a more difficult time understanding the manual.

4. Search Time – as manuals increase with respect to volume and complexity, the retrieval time of solutions also increases.
5. Schematic Viewing – the housing of diagrams should occupy one page per diagram. However, the increasing complexity of systems directly influenced the complexity and size of diagrams. As a result, a single diagram can reside on multiple pages making it difficult for the user to decipher.

Several attempts were made to improve upon paper manuals. These attempts will be discussed in Section 2.1.2. Then as the popularity of the internet grew, a new medium for technical communication also grew.

#### 2.2.1. ONLINE MANUALS/ASSISTANCE

Internet usage has been steadily increasing in the past few years. This trend has been occurring both in the United States and worldwide. As users saw the possibilities for Web-based applications, their expectations for Web-based documentations grew. The potential benefits of online assistance included: reduction in the geographical distance between users and conventional documentation and consultation; and an increased desire to make the systems usable regardless of the user's experience (novice to expert). This led to trends that evolved according to the expectations of the user. As online documentation developed, the factors that influenced its development and determined its effectiveness became:

1. Available technological capabilities

2. End user expectations
3. User centered design [ZCFC01].

These developments introduced the question of how to develop the more effective medium.

The process of generating a technical communication now had an additional step, the medium choice. This choice was determined by the answers to the following questions:

- What are the needs of the audience
- How can the functionalities of the new information technology be utilized
- Will the application of the new medium prove the concepts for introduction of this new medium [SK92].

As technological advances continued, additional mediums were made available. These included animations and virtual reality. These new media require modifications in the actual content and content presentation. This is beyond the scope of this research. However, with the increasing options for communication's mediums came increased attempts to improve the individual mediums.

### 2.2.2. RESEARCH DEVELOPMENTS

Suggested improvements for technical communications range from user involvement to generating a personalized manual to the development of a system to enrich assistance. The user edit is a process that evaluates a current technical

communication by having a novice perform a task using the communication [Atl81]. It is proposed that this process results in manuals that are easier to use. Another solution includes the development of a well-designed manual by generating a user manual that contains pointers to cookbooks, tutorials and other information sources. These pointers are chosen by the user and result in a personalized user manual [Maj85]. Other improvements involve the development a system.

The Electronic Document System (EDS) creates electronic hypertext documents from a book manual [KM92]. A similar system called Doorway was utilized by the Air Force for technicians that maintained automatic test equipment. There was a high turnover rate among the technicians, and it took approximately six to nine months for the newcomer to develop the desired expertise [Col91]. Genie, an interface that answers user questions via natural language, attempts to enrich the user's experience of online assistance [Wol92]. This system is an intelligent interface that accepts text input of natural language queries and provides an answer to the user. Coincidentally, its main drawback is its main attribute. Genie takes initiative in providing enrichment and does not rely on user control. However, when the user's answer requires instruction, a set dialogue related to the curriculum is required.

More recent improvements include the idea of open-source documentation. Software development processes like Extreme Programming (XP) and Rational Unified Process (RUP) present additional difficulties in the upkeep of a technical document. As



these processes go through the iteration process adding new features and correcting malfunctions, the documentation can not keep up with the changes [BP01]. Therefore, open-source documentation is suggested as it will blur the line between the writer and reader and allow readers access to implemented solutions. The combination of technical support and documentation has also been suggested. This collaboration will reduce the redundancies and inconsistencies in the information repositories and improve the performance of both the technical support and documentation. This process works using a Solutions Database that indexes users' question to solutions. A keyword search is performed against the database [Pie03]. The A1 approach used by iTech eliminates this cognitive step that must occur to translate the user's question into keywords. Another improvement method that removes additional cognitive processing during the information search is the use of spatial maps for cellular phone manuals. The menu within cellular phones follows a tree structure. However, most manuals follow a step-by-step presentation. Conversion of the cellular phone manual to a tree structure eliminates the translation of the step-by-step instruction to a menu-tree. Studies show that this technique is influenced by age and middle-age subjects have the most success [Bay03].

As developers generate and investigate solutions to the previously mentioned issues, the problems with understandability and search time prevail. Improvements towards providing a more natural interaction between the user and the technical documentation are required.

### **2.3. AUTOMATIC SPEECH RECOGNITION**

Automatic speech recognition (ASR) refers to the ability of a system to receive and interpret human speech. The system is equipped with a sound source for input, typically a microphone. Once the speech is accepted, a statistical method is employed to allow for recognition. The most commonly used statistical methods are the Hidden Markov Models (HMM). Using Bayes' Rule, the speech recognition engine determines the most likely word sequence by calculating the probability of an observed sequence of acoustic data given another word or sequence of words. There are several classifications for ASR systems that range from speaker dependence vs independence to large (thousands of words) vs small (hundreds of words) grammars to noisy vs quiet environments for use to continuous vs discrete speech. These classifications determine the system's accuracy as well as the pace of user speech allowed. Within automatic speech recognition, there are two common approaches used to achieve the best results: grammar constrained recognition and natural language recognition [AC06].

Grammar constrained recognition utilizes a constrained set of possible phrases that the system will recognize. This set is considered to be a formal grammar of small or medium size. It is typically defined using grammar specification languages. Due to the restricted size of the grammar, this approach works well for systems that require limited, specific responses e.g. choosing from a menu or well-defined list and responding to yes/no questions. Phrases that are not included in the specified grammar will not be

recognized. With natural language recognition, the user is allowed natural, sentence-like responses to specific questions. Here, the grammar is constructed using a corpus of typical responses, words or phrases. Each response is linked to a specific token or concept. Another methodology that falls under the umbrella of natural language recognition is key-word spotting (KWS). KWS as the name implies, identifies key-words specified in a grammar [Alo05]. This is done by the ASR during natural day-to-day conversations. This capability is currently employed by Call Center (CC) markets and Homeland Security. KWS is not used for continuous recognition as it indicates whether a key-word is spoken and does not recognize the entire speech dialog.

The two approaches to speech recognition have their own individual advantages and disadvantages. While natural language recognition allows for a more natural interaction style, the recognition accuracy is decreased and the creation of a suitable corpus can prove expensive and time-consuming. The latter problem could be improved by the use of an established corpus e.g. SUSANNE corpus or the CHRISTINE corpus which are freely available [Sam05]. This research utilizes the natural language recognition approach within a spoken user interface.

### 2.3.1. SPOKEN USER INTERFACE

The use of spoken user interfaces or speech-based user interfaces is rapidly growing. Speech has gained popularity as both an input and output modality, as it represents the way users naturally communicate with others. The desire for more natural

interactions between humans and computers is typically cited as the foundation for speech interfaces [THTS05]. Speech also provides, the visually- and literacy impaired and ‘hands busy’, a medium to interact with computers and computing devices [KSCL00].

Generally speech interfaces fall into three categories:

1. Command-and-control
2. Directed dialog
3. Natural language

These categories are delineated by their grammar constraints. Command-and-control systems have a very restricted grammar. Due to these restrictions, the grammar size tends to be small and yield low WER. With directed dialog systems, the application guides the user using machine-prompted dialogs and accepts one piece of information at a time. These systems also have restricted grammars that are specific to their occurrence within the entire system dialog. The natural language systems, as the name suggests, allows users to pose questions to the system as if they were addressing another person. The complexity of the required grammar is controlled by the complexity of the task to be accomplished [AMRS96]. Extensive grammars and typically complex statistical language processing of the recognized speech accompanies this system [RFQW05, RQZB01]. A high-quality interface is most effectively generated using an iterative approach. These studies employed the natural language system but introduced a new methodology for its information retrieval.

## **2.4. INFORMATION RETRIEVAL**

Information Retrieval (IR) is a multidisciplinary branch of computer science that deals with the automatic storage and retrieval of information. The goal of an IR system is to return the information that specifically satisfies the user's needs [KLM97]. The user presents their needs to the system using a query. Statistical cues including word frequencies, document length and word importance are used to assign potential relevance of a document [SK01]. A list of ranked documents is then presented to the user. The user queries are expected to follow specific syntax rules that do not coincide with natural language. While, a natural language question is accepted by the search engine, stop words are stripped and the remaining question is treated as a query. However, these stop words provide deeper insight into the user's needs [RZBZ01] i.e. the type of answer that is sought. The question "Where is the carburetor?", would be reduced to 'carburetor' and the results of an IR system would include all mentions of the word 'carburetor'. However, inclusion of the stop word 'where', indicates that the positioning of a carburetor is desired. As trends towards more natural interactions between humans and computers continue, a new approach to information retrieval with respect to natural language questions evolved.

## **2.5. NATURAL LANGUAGE PROCESSING**

Natural language processing (NLP) is a subfield of artificial intelligence and linguistics. It studies the problems inherent in the processing and manipulation of natural

language, and, natural language understanding devoted to making computers "understand" statements written in human languages [Wik05].

Natural language understanding research can be categorized several ways. There is open domain question answering and natural language interfaces to databases, as well as text-based natural language research and dialogue-based natural language research. Text-based research is performed with respect to text-based applications e.g. newspapers, magazines, books, manual and e-mail messages and includes topics such as:

1. Information extraction and comprehension
2. Document Retrieval
3. Translations
4. Summarization.

Dialogue-based natural language research is performed with respect to dialogue-based applications. These applications involve human-computer interaction and the interaction techniques include spoken dialogue, keyboard and stylus. Examples of applications that require dialogue-based natural language understanding include: question-answering systems, tutoring systems and automated response and help centers. While some of the problems of natural language understanding pertain to both research groups, each has its own unique issues. This research will focus on improving the issues related to dialogue-based applications.

Dialogue-based applications are involved in a co-operative relationship with the user. These applications must manage a natural flowing dialogue between the user and the interface. This dialogue should acknowledge that things are understood and if not, create sub-dialogues to provide further clarification. Additional categorizations of IR include open domain question answering and natural language question answering.

#### 2.5.1. NATURAL LANGUAGE INTERFACE TO DATABASES

Natural language interfaces to database (NLIDB) allows users to access information from a database using natural language queries [AR00]. They remove the user's need to know the structure of the database and details about the data and also provides a more natural interaction [SKD03]. Examples of existing NLIDBs include RENDEVOUS (early seventies), ASK (mid-eighties) and currently IBM's LANGUAGEACCESS. The general architecture for NLIDBs is as follows.

The architecture consists of a linguistic front-end and database back-end [AR00]. The front-end is where the natural language question is input and translated into a meaning representation language (MRL). MRL is then passed to the back-end. At the backend, MRL is translated into a supported database language and executed. Preprocessing is performed on the query input, followed by parsing and semantic analysis of the question and finally semantic post-processing. This approach is similar to those utilized in NLQA.

## 2.5.2. NATURAL LANGUAGE QUESTION ANSWERING

Natural language question answering (NLQA) is the process of retrieving answers for questions [RFQW02]. Questions are posed in natural human language and a precise answer is given following the same form. The main goal of NLQA is to return an answer rather than a list of documents to the proposed question. There are several approaches to NLQA, many of which utilize approaches used in document retrieval [RZBZ01]. This literature review will briefly look at three popular approaches.

### 2.5.2.1. QUESTION ANSWERING USING STATISTICAL MODEL

Question answering using Statistical Models (QASM) is used to convert natural language questions into a search engine specific query. It is based on the premise that the selection of the best operator to apply on a natural language question is possible. QASM is combined with another algorithm (AnSel) to produce precise answers to natural language questions [RQZB01]. The operator produces a new query that improves upon its original. A classifier decides on the best operator to apply to a question, N. The operator is then matched to its question-answer pair.

It is very expensive to provide a large corpus of question-answer pairs, thus an algorithm is used that is stable to missing data. This algorithm is the expectation maximization (EM) algorithm. The EM algorithm iteratively maximizes likelihood estimation. The missing data mentioned refers to paraphrases of the natural language questions [RZBZ01].



#### 2.5.2.2. PROBABILISTIC PHRASE RERANKING

Probabilistic Phrase Reranking (PPR) is fully implemented at the University of Michigan [RFQW02]. This process goes through a set of subtasks to retrieve the most relevant answer to the proposed question. The tasks are as follows:

- Query modulation - The question is converted to an appropriate query at this stage.
- Question Type Recognition – Queries are organized according to the question type i.e. location, definition, person, etc.
- Document Retrieval – the most relevant unit of information e.g. documents are returned in this stage i.e. the units with the highest probability of containing the answer.
- Passage/Sentence Retrieval – The sentences, phrases or textual units that contain the answers are identified from within the information unit returned in the previous stage.
- Answer Extraction – The chosen textual units are split into phrases, each of which is a potential answer.
- Phrase/Answer Reranking – The phrases generated in the previous stage are ranked. At the top of the list should reside the phrase with the greatest possibility of containing the correct answer.

### 2.5.2.3. BAYESIAN APPROACH

The Bayesian approach to IR uses a probabilistic IR model and applies Bayes' Law. The goal of the probabilistic model is to estimate the probability that a document,  $d_k$ , is relevant (R) to a query,  $q$  i.e.  $P_q(R | d_k)$  [KLM97]. With respect to IR, each document is represented by a set of words. These are the words that remain after the stop words have been purged from a document. These words are then stemmed by removing the suffixes and prefixes, after which they are known as index terms. Each document is thus represented by a vector,  $t = (t_1, t_2, \dots, t_p)$  where  $p$  is the number of index terms. Bayes' Rule is then applied to this model to express the probability that a document is relevant to a specific query,  $q$ .

$$P_q(R | t) \propto P_q(t | R) P_q(R) .$$

The assumption that the terms are independent given the relevance and non-relevance of a document, results in an expression for the log odds of relevance. Documents are ranked according to this expression.

$$\log \frac{P_q(R | t)}{P_q(\overline{R} | t)} = \log \frac{P_q(R)}{P_q(\overline{R})} + \sum_{i=1}^p \log \frac{P_q(t_i | R)}{P_q(t_i | \overline{R})}$$

A document is considered relevant if it satisfies the user's needs and non-relevant if it does not. To apply this expression, the frequency of terms in the relevant and non-relevant documents is needed. However, initially the status of documents i.e. whether relevant or non-relevant, is not known. To facilitate this, an ad hoc estimation of the

probabilistic model parameters is used to determine an initially-ranked list of documents. The addition of the Bayesian approach to probabilistic models overcomes some of the weaknesses of the existing probabilistic models. These strengths include: producing an initial document ranking not based on ad hoc considerations, providing an automatic mechanism for learning and incorporating relevance information from other queries.

### 2.5.3. NLQA VS. NLIDB

There are fundamental differences in the scientific goals and technical constraints of NLQA verses NLIDB. The advantages of NLIDBs include a search domain of restricted documents, semantic interpretations and pre-identified relations and entities [AR00]. With respect to the familiarities, the two areas accept a natural language query, interpret it and finally process it. The approach used in this research will accept natural language questions and retrieve their solutions without additional language processing.

### **3. SYSTEM DESIGN**

This chapter will discuss the logical and physical aspects of iTech's design. It also describes the approaches used to address the problems described in Section 1.2. The client-server architecture, database design and population, component, user interface and other features will be discussed.

#### **3.1. ITECH**

The main issues with online assistance are described as: requirement for an Internet connection, knowledge of appropriate information resources and translation of information needs to keywords [CBWB01]. The latter two issues also relate to paper manuals. Though, the second issue is resolved with the inclusion of the manual in the product package. Thus the most prevalent issue remains the additional cognitive processing required to translate the user's information needs into keywords or topics. iTech is an interactive technical assistant that improves upon the current usage of technical communications today by removing the additional mental arithmetic. iTech allows the user to input their information needs exactly as they are generated, i.e. in natural language form, via speech.

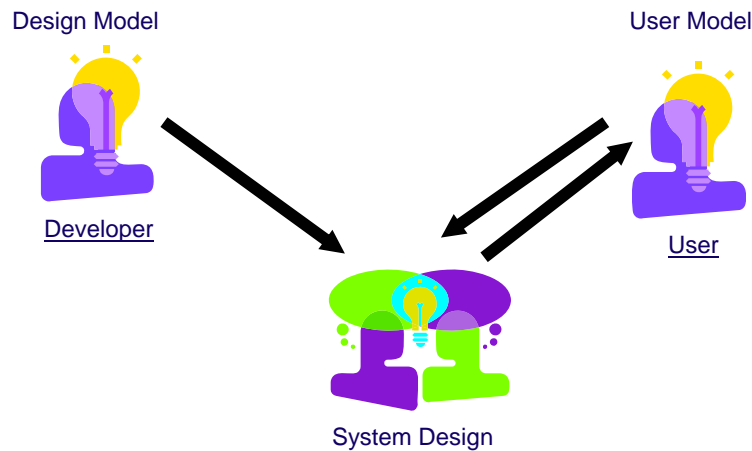
### **3.2. SPEECH USER INTERFACE (SUI)**

While accessibility is typically mentioned as the main motivator for a speech based application, other features of speech were instrumental in its inclusion in this research [THRS05]. Some studies have shown that it is natural for users to communicate via speech with computers using short imperative commands, succinct responses with restricted vocabularies [THTS05]. Others indicated that when spoken queries were compared with written queries, the spoken queries proved longer and just as effective in retrieving results [DC04]. The increased length of the spoken query is attributed to increased semantic content. This increased content offsets the effects of speech recognition errors. The Answers First approach used in this research exploits the increased length of spoken queries to facilitate increased resolution success between the recognized question and knowledge repository. As speech was chosen as the iTech's input modality, guidelines to developing an effective speech interface were investigated.

Guidelines for the development of an effective, usable speech interface fall into two basic categories. There are those guidelines that are specific to a speech interface and those that are universal and can be applied to any user interface. However, there are three main functionalities that must exist for an effective speech interface [SRZT01]. An effective speech interface should:

- Be a proper participant in the conversation dialogue
- Handle problems from word recognition errors
- Provide understandable interpretation and facilitate the user completing their task.

In short, the main focus is to have developer's conceptual model match the user's mental model of the interface as illustrated in *Figure 3: Conceptual Models*.



**Figure 3: Conceptual Models**

Additionally, the following guidelines taken from Schniederman's golden eight were followed [Sch98, Kra01]:

1. Consistency – consistency complements the matching of the user's mental model with the developer's conceptual model. It is imperative for the usability of an interface. However, when an alert or grasping of the user's attention is required, an obvious inconsistency is very effective.
2. Enable user shortcuts – this allows the interface more flexibility, enabling it usable by both novice and expert users. iTech allows the user to interrupt him when speaking. Thus an expert user need not sit through entire prompts and dialog that they are familiar with.
3. Informative feedback – feedback is very important especially during an error. iTech's responses not only inform the user as to the status of the system, but also suggest further action when necessary.

4. Internal locus of control – users should feel that they are the initiators and not responders during system interactions. Conversations in iTech are initiated by the user only.
5. Reduced short term memory load – with respect to the limits of human information processing of short term memory, the solutions in iTech are presented visually. This output mode was also chosen as it is consistent across the technical communication mediums used in the evaluation of iTech.

Another major contributor to the success of a SUI, especially a natural language SUI, is the grammar.

The grammar consists of all the words and phrases that a SUI will understand. Typically for a natural language SUI, a commercial CORPUS is used. However, iTech can only answer questions to which he already knows the answers. These questions are stored in the knowledge repository; consequently the grammar for iTech was generated from the question knowledge repository. In the first iteration of iTech, the grammar resembled a bag of words. These words originated from the question knowledge repository. However, the WER proved unacceptable. iTech was unable to provide reliant recognition due to the large grammar size and common phonemes of several words. For the second iteration of iTech, the grammar was regenerated. The individual words in the bag have been replaced by word pairs or bigrams, see *Figure 4: Excerpt from iTechGrammar.grmxl*. These pairs were created as they occur in the question repository. The bigram grammar shows a marked improvement in recognition accuracy. Once the speech is input and recognized, the process of information retrieval begins.

```

<rule id="words">
  <tag> $=new Object(); </tag>
  <one-of>
    <item>do i <tag> $_.value="do i"; </tag> </item>
    <item>do you <tag> $_.value="do i"; </tag> </item>
    <item>how do <tag> $_.value="how do"; </tag> </item>
    <item>what does <tag> $_.value="what does"; </tag> </item>
    <item>do what <tag> $_.value="do what"; </tag> </item>
    <item>can i <tag> $_.value="can i"; </tag> </item>
    <item>i move <tag> $_.value="i move"; </tag> </item>
    <item>what is <tag> $_.value="what is"; </tag> </item>
    <item>what do <tag> $_.value="what do"; </tag> </item>
    <item>i get <tag> $_.value="i get"; </tag> </item>
    <item>of the <tag> $_.value="of the"; </tag> </item>
  </one-of>
</rule>

```

**Figure 4: Excerpt from iTechGrammar.grmxl**

### 3.2.1. ANSWERS FIRST

Answers First (A1) is the approach utilized by iTech for conversational question answering. A1 does not follow the typical IR techniques. No language processing is performed on the recognized question before a query is executed. The recognized speech is sent to the server and decomposed into word pairs or terms. These terms are then matched against the knowledge repository of questions. The question with the highest concentration of matched terms is identified. Answers First proposes that relevant information is lost during query formulation, the process by which a user's questions is translated into a query. This information leak occurs when stop words are removed, as these stop words could provide further insight into the user's information needs [RQZB01]. Answers First also proposes that the order of words within a statement or question does not influence the interpretation of that statement.

The prequels to the Star Wars trilogy highlighted an old and often forgotten hero, Yoda. Yoda, in addition to being a peculiar creature to look at, also has a unique way of



speaking. Yoda's sentence constructs do not follow the guidelines for correct usage of the English language. His speech reverses the order of words. However, he is always understood. Here are two actual excerpts from a speaking Yoda doll, followed by the same statements in traditional American English:

*Yoda: Happy I am to see you.*

American English: I am happy to see you

.....

*Yoda: Tired I am, to sleep I must go.*

American English: I am tired, I must go to sleep.

Thus, A1 performs a straight match of recognized terms to the terms in the KR until a unique match is found. A unique answer indexed to the matched question is retrieved and presented to the user. This scenario represents the 'best-case' experience. However, there are other possible scenarios.

### Scenario #1

The question resolution algorithm yields two or more unique questions from the recognized terms. These matched questions are indexed by one unique answer. Since all of the matched questions resolve to a unique solution, iTech presents that answer to the user.

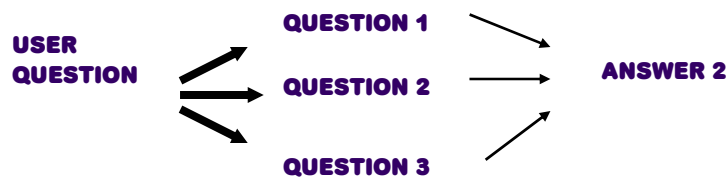


Figure 5: Scenario #1

## Scenario #2

The question resolution algorithm yields two unique questions from the recognized terms. These matched questions are indexed by two unique answers. Since the choice is restricted to two, in his response to the user, iTech will present the two unique questions to the user. The user is then given the autonomy to select either question or create a new question to request of iTech.

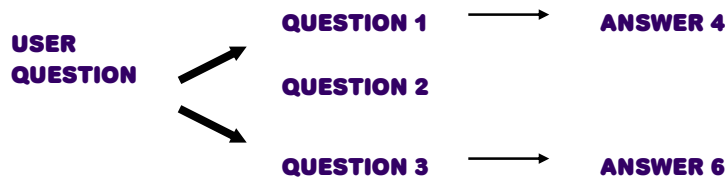


Figure 6: Scenario #2

## Scenario #3

The question resolution algorithm yields one question. However, this question indexes two unique answers. The QRA will identify two unique questions that index the retrieved answers. iTech will present these two questions to the user, and they will be given the opportunity to select on the presented questions or rephrase the original question.

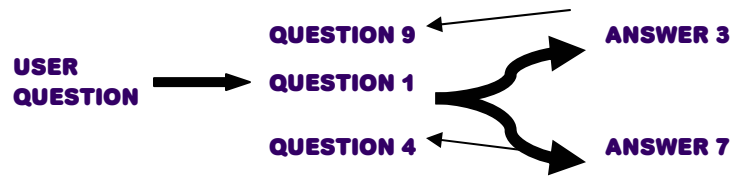


Figure 7: Scenario #3

#### Scenario #4

The question resolution algorithm yields more than two questions. Each question is indexed by a unique answer. Therefore there are more than two questions that could be presented to the user. The QRA then applies case-based reasoning. It looks for the answer, in the retrieved group of answers that has the highest resolution score. This resolution score is indicated by the NumOfOccurrences field in the Answers table which gets incremented every time that specific solution is presented to the user. A discussion of the design principles will follow.

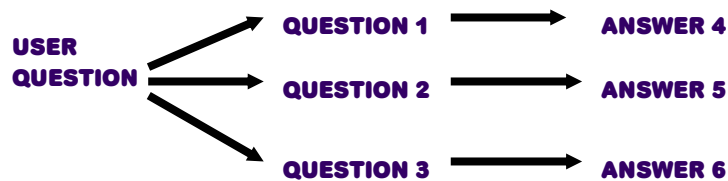


Figure 8: Scenario #4

### **3.3. DESIGN PRINCIPLES**

The main objective for the prototype assistant developed in the course of this research was to enable users to seamlessly retrieve the necessary information from a technical communication to accomplish a task. Thus, the major principles of the assistant's framework were designed to include:

- Conversational question answering methodology that accepts and answers natural language questions without performing typical natural language processing.
- Knowledge repository that enables expedient retrieval of answers.
- Continuous speech recognition engine.
- Effective and efficient speech users interface (SUI).

Guided by these principles, the framework of the new interactive technical assistant is introduced in the following section.

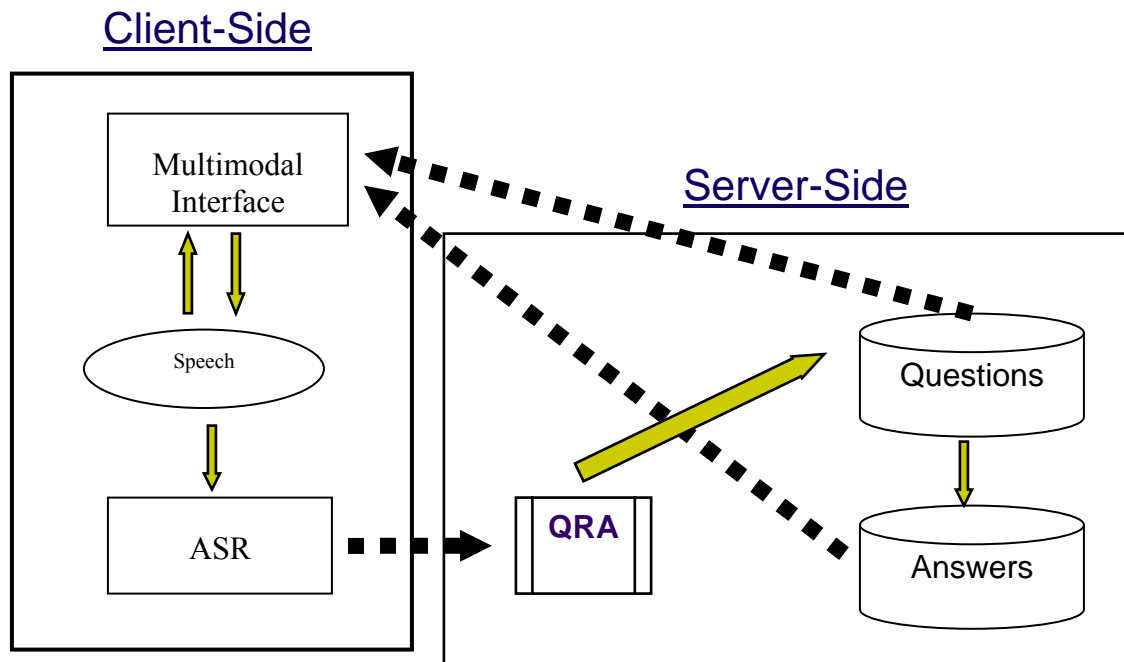
#### **3.3.1. SYSTEM ARCHITECTURE**

The iTech system has the typical client-server architecture as shown in *Figure 9 System Architecture*. On the client side, the user initiates the conversation by pressing the button to speak and asking a question. The built-in speech recognition engine, Microsoft English ASR Version 5 Engine, recognizes the user's question and passes the recognized speech to the browser environment of the page where the Speech Application Language Tags (SALT) is hosted [CCIM02]. Additional client-side scripts then manipulate the

SALT elements and the resulting text of the recognized speech is sent as a request to the server.

The server side consists of the Knowledge Repository (KR) and the Question Resolution Algorithm (QRA) module. The KR is populated with question-answer pairs generated from the chosen manual. The QRA module resolves the recognized question with the KR, identifies the ‘best-fit’ question-answer pair and retrieves the relevant answer. Control is then returned to the client, where the retrieved answer is displayed to the user.

The system works in the following way. A user initiates the system by opening up the application’s browser. Once loaded, iTech welcomes the user and instructs them of his purpose and to how to ask a question. The user presses the ‘Push 2 Speak’ button and asks their question. The browser interacts with the user and identifies the exact content of the question. The question is then translated into text and sent to the QRA module. This text is organized into word pairs<sup>2</sup>. The QRA module matches the question terms against the table of corresponding word pairs in the KR and identifies the question with the highest concentration of terms. The indexed answer to the identified question is retrieved and a request is passed back to the client containing the URL of the answer. Finally, the answer is displayed for the user. The following section will explain in detail how the architecture that supports these functions.



**Figure 9 System Architecture**

### 3.3.2. MULTIMODAL INTERFACE

The multimodal interface is the point of interaction between the user and I-Tech. It can be housed on any personal computing device with a microphone or the ability to add a microphone. The microphone is used to collect the user's speech. The graphical user interface (GUI) consists of two frames: the Navigation frame and the Content frame

The Navigation frame consists of the animated agent and the Speech Application Language Tags (SALT). The presence of a likeable animated pedagogical agent has been

---

<sup>2</sup> In this research, the phrase 'word pairs' and the word 'terms' are used interchangeably.

shown to improve student performance by enhancing the student's desire to learn [BR03]. This desire is increased as the student forges a personal connection with the agent, thereby making the learning experience more enjoyable. However, the agent must possess the following characteristic to be effective: engaging, person-like and credible. To develop an agent that is motivating; believable and trustworthy thereby promoting relationships with the learner requires the presence of these characteristics [BR03]. Additionally, iTech is male. This choice was deliberate and influenced by the findings that male pedagogical agents are perceived as more extraverted and agreeable resulting in a more satisfying experience by the learner [BK03]. The ethnicity of iTech was chosen as African-American. This choice was determined by study results that indicated African-Americans were more inclined to choose an agent of the same ethnicity than Caucasians [BSH03]. The agent was generated using SitePal and embedded into a HTML file [Sit06]. The SitePal application allows for greater developer control over the appearance of iTech. To enable the agent's perceived participation in conversations, SALT and JavaScript were used.

JavaScript provided text-to-speech (TTS) capabilities to the agent. Therefore, JavaScript was used to control iTech's speech. SALT is then used to enable iTech's hearing. SALT is embedded in a compliant browser and using Microsoft's recognition engine, allows iTech to listen to the user's questions. Once the question is recognized, the

question resolution algorithm is applied, an answer identified and retrieved, this answer is displayed in the Content Frame.

The Content frame is a HTML page that is used to display the solutions to the user's questions. When iTech is loaded for the first time, this frame displays the cover of the vi manual used to populate iTech and in its evaluation study. Once interactions begin and the user starts asking iTech questions, the Content frame dynamically displays the solutions retrieved by the question resolution algorithm. Application of the QRA occurs once speech is recognized. It is initiated by a PHP script that also connects to the MySQL database that houses the KR.

### 3.3.3. KNOWLEDGE REPOSITORY (KR)

iTech's knowledge repository is a MySQL Database. Its database design consists of six database tables. They are Answers, AnswerType, Categories, CategoryTerms, Questions and Terms. These tables maintain iTech's knowledge. This discussion will begin with a detailed description of each database table. When completed, an entity-relationship model will be presented.

#### 3.3.3.1. DATABASE TABLES

The SQL used to create each of the database tables can be found in Appendix C. This section will discuss each table with respect to its functionality, population and purpose.



#### 3.3.3.1.1. ANSWERS

The Answers table contains everything iTech knows about the vi editor. It is populated from the book manual, “Learning the vi Editor, 6<sup>th</sup> Edition” published by O’Reilly. The manual represents all of iTech’s knowledge about the vi editor. Based on the premise that iTech can only answer questions that he knows the answer to, the manual was separated into its delineated sections. These sections, accessed via their unique URLs, are identical to the actual pages in the book. This carbon copying was done to reduce the effect of any indirect variables during the evaluation study. Each section is stored in the Answers table as a unique answer. Each answer in the table has a unique id, the URL for the solution and an answer type. Another field in this table is the NumOfOccurrences field. This field is used when the question-answer resolution does not yield a unique answer but falls into one of the scenarios mentioned in section 3.1.2.

#### 3.3.3.1.2. ANSWERTYPE

AnswerType is used to distinguish the type of answer. For example, does the answer to the matched question give an amount, a command or an explanation? Each answer in the Answers table is indexed by an answer type.

#### 3.3.3.1.3. CATEGORIES

The Categories table contains information about the proper names in the Questions and Terms tables. Proper names tend to yield higher WER than common words.

Thus, to facilitate greater recognition accuracy, the proper names in the Terms table are replaced by a common genre. For example, if the Terms table contained ‘Atlanta airport’, it would be replaced by ‘city airport’ and the Categories table would have an entry for the category ‘city airport’ and Term ‘Atlanta airport’.

#### 3.3.3.1.4. CATEGORYTYPE

CategoryType is used to identify the different types of categories indicated in the Categories table. It provides the description for the CategoryID in the Categories table. These latter two tables will be instrumental in improving recognition accuracy when the assistant’s domain consists of many proper words.

#### 3.3.3.1.5. QUESTIONS

Questions contains the many ways that a user might request the information that iTech knows. It represents all of the questions that will illicit the answers stored in the Answers table. Each question is indexed by an entry in the Answers table. The data for the Questions table is manually generated. The preferred scenario consists of the developer coming up with the original question set, initially. This set could then be enhanced by the efforts of one or two other developers. For this research, one other developer was used. Finally, Wizard-of-Oz experiments are performed with subjects from the participant pool or expected users. These experiments should simulate interactions with a functional iTech and the interactions recorded. When completed, coverage of these

questions in the table should be verified. Here, the questions were transcribed and applied to iTech via text. This was done to eliminate any recognition errors. If the questions do not yield an answer, they are added to both the Questions and Terms tables. Those questions that yield answers, even if they are different from existing questions in the repository, should not be added as they are already covered. Wizard-of-Oz experiments were performed with five subjects from the identified participant pool. Within participation of the first three subjects, no new questions (question to be added to the database) were introduced by the participants.

#### 3.3.3.1.6. TERMS

Word pairs are stored in the Terms table for each term. These pairs are generated from the Questions table and they represent all the terms that would elicit a response from iTech's knowledge repository.

#### 3.3.3.2. ENTITY RELATIONSHIP MODEL

All of the tables discussed here are used by iTech to provide answers when the user poses a question to the application. *Figure 10: Entity Relationship Model* is the entity relationship model for iTech's knowledge repository.

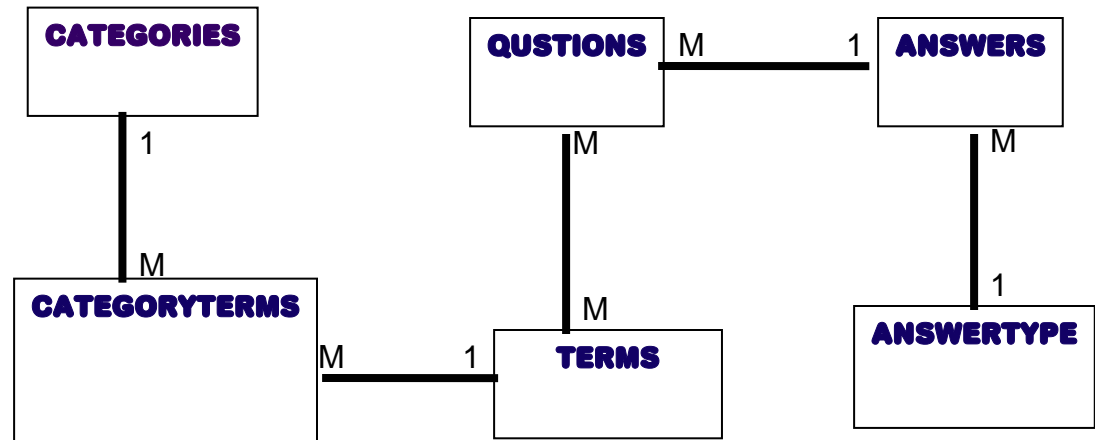


Figure 10: Entity Relationship Model

### 3.4. SAMPLE INTERACTIONS

Every time iTech is launched he introduces himself and informs the user of his purpose, to assist the user with the vi editor.

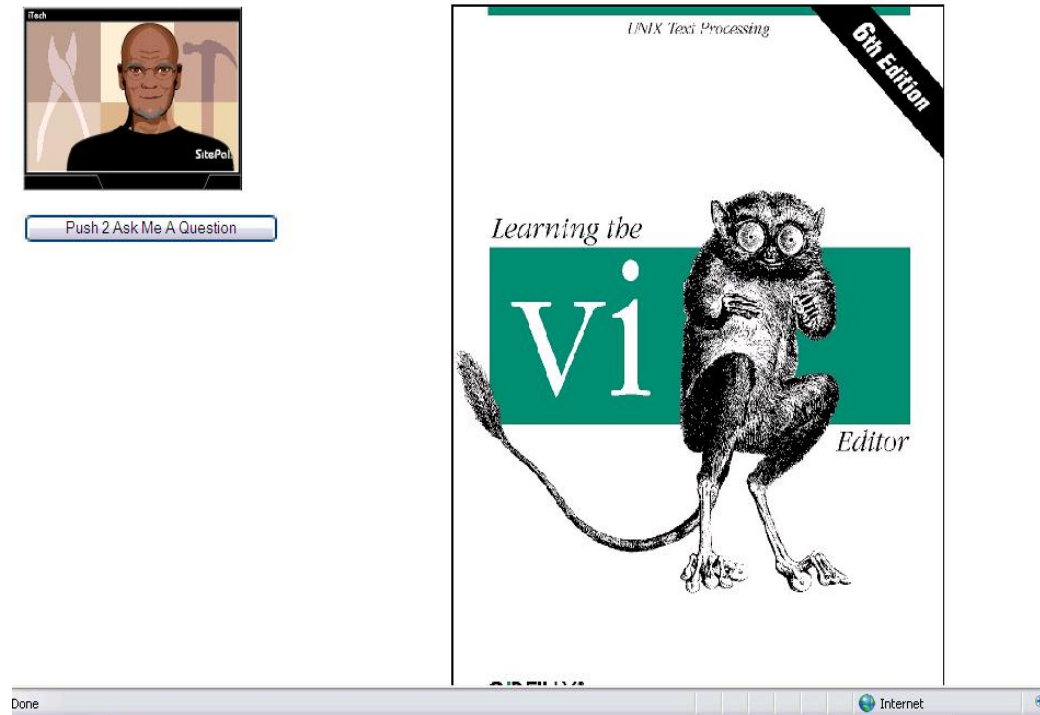


Figure 11: iTech Welcome Screen

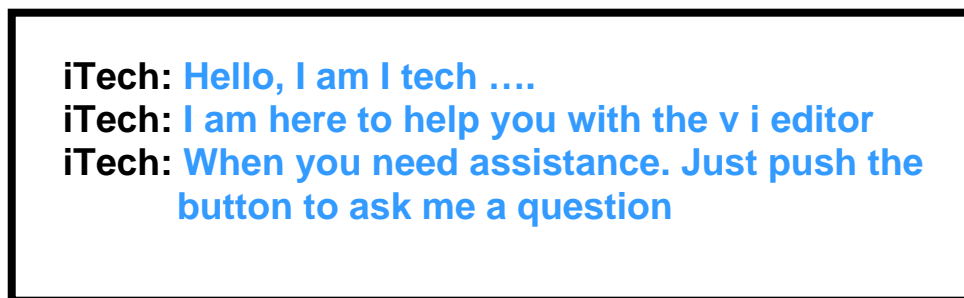
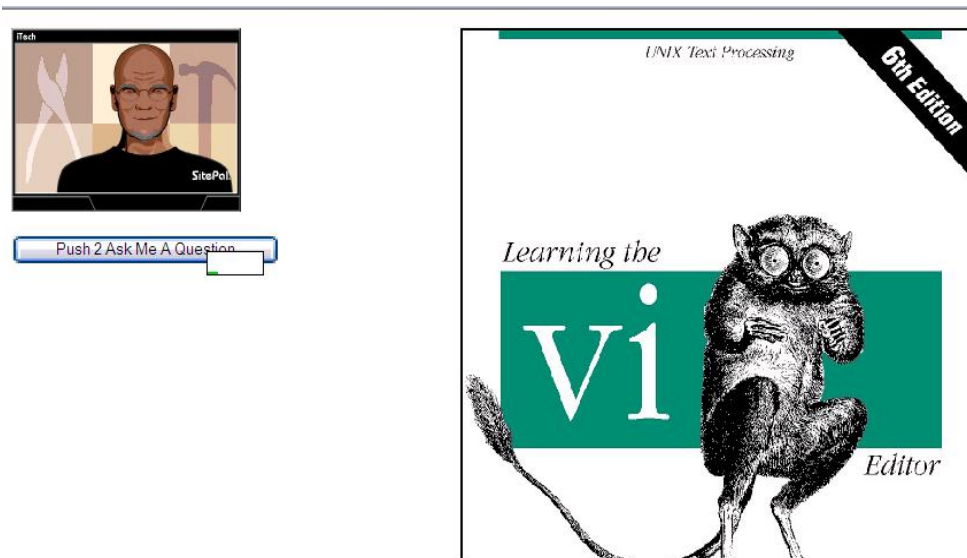


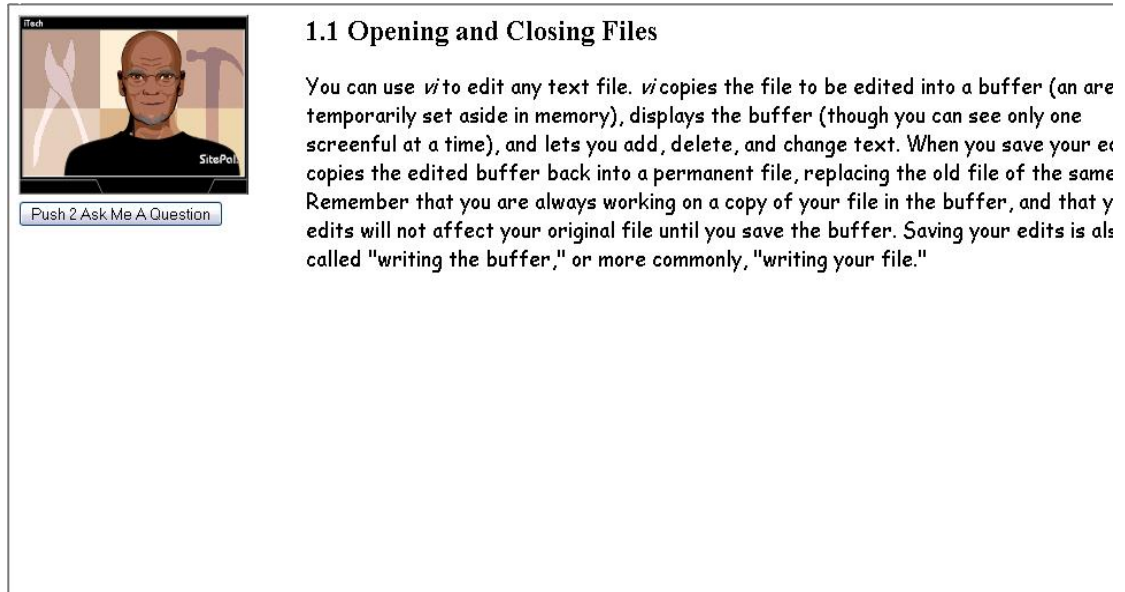
Figure 12: Example of iTech's Welcome

After iTech's introduction, the user now has autonomy. At his/her discretion, they can press the button and ask iTech a question.



**Figure 13: iTech is Listening**

If the question is matched, the solution is displayed and iTech responds with an affirmative statement that an answer was found.



**Figure 14: iTech displaying a Solution**

If a scenario in which more than two unique answers occurs, the Content frame stays blank while iTech encourages the user to rephrase their question. iTech also has event handlers for situations when no recognition occurs, or user presses the button and doesn't speak, or speaks too softly for iTech to hear. Examples of these dialogues are shown in *Figure 15: iTech Dialogue*.

No Recognition Event:-

**ITECH: I DIDNT QUITE UNDERSTAND WHAT YOU ASKED.**

Silence Event:-

**ITECH: YOU PRESSED THE BUTTON TO SPEAK, BUT I DID NOT HEAR ANYTHING.**

**ITECH: IF YOU SAID SOMETHING, COULD YOU PLEASE REPEAT IT?**

Solution found:

**ITECH: I HOPE THIS ANSWERS YOUR QUESTION**

**ITECH: DO NOT HESITATE TO CALL ON ME IF YOU HAVE ANY MORE QUESTIONS**

Figure 15: iTech Dialogue



## **4. IMPLEMENTATION RESULTS**

In the absence of experimentation, iTech's design is simply theory. This theory makes significant contributions to fields of technical communications, interactive assistants and conversational question answering. However, these contributions must be proven via experimentation. Upon completing the system implementation of iTech based on the approaches outlined in the previous chapter, a formal experiment was conducted to validate the claims specified in Sections 3.1 and 3.2.1. The objectives of this evaluation mainly focus on evaluating the system performance, with respect to search time for solutions and task completion time; measuring user satisfaction of the system and how accurate was the question resolution algorithm. Statistical methods such as Shapiro-Wilk tests and Kruskal-Wallis [MS01] were used to evaluate and analyze the experimental results.

This chapter reports the experiment with iTech, which incorporates the Answers First approach. Section 4.1 details the goals of this experiment including the benchmark for the success of iTech. Then the experiment settings, participants and procedure are specified in Section 4.2. Section 4.3 describes the data collection methods. Finally the experimental results are described and discussed in Section 4.4.

### **4.1. EXPERIMENT DESIGN**

The goal of this research experiment is to evaluate iTech with respect to task success. If iTech can allow users to complete their task more expediently, with less time

spent searching for solutions, less effort in producing their search query and a more enjoyable experience when compared to the online and paper versions of the vi manual, then the experiment will be viewed as a success. Recorded search times and user evaluations will provide the main data points for evaluation. The primary goal of this experiment is to prove that there is a significant difference in the search times of the different technical communication mediums presented in the experiment. However before any experiment can be performed, the correct hardware and software must be in place to support the experiment.

## **4.2. EXPERIMENTAL SETTINGS**

### **4.2.1. MATERIALS**

The study was conducted in a private room at Auburn University. The room was furnished with one large table, 5 chairs. Testing was conducted on a Gateway 2000e CPU with a 17” Sony Monitor running Windows XP equipped with a standard scroll mouse and a Logitech USB Headset. The following software was required and downloaded to the experiment machine:

- Internet Explorer 6.x
- Microsoft Internet Speech Add-in 1.0.
- SecureCRT 4.07

There was also a Sony 700x Digital Handy cam to record all user interactions via the monitor.

#### 4.2.2. PARTICIPANTS AND PROCEDURE

Seventy-four college level students were recruited as subjects. Since the domain for iTech involved using the vi editor, all of the subjects needed very limited or no exposure to vi. No other special skills were required. All of the subjects were Auburn University students. To ensure that all subjects had similar knowledge concerning the use of a PC and editing text documents, the subjects were enrolled in at least one course from the College of Engineering.

The usability evaluation was a controlled experiment. To reduce the causal effects of other factors, the following controls were applied:

- 1) All participants sat in the same chair in the same room with the researcher.
- 2) The task completed by the participants was the same. Therefore all participants were asked to do the same task in the same order. The only independent variable changed was the medium of the technical communication.
  - i. The participants were randomly selected to use the book manual, online manual or iTech.
  - ii. All participants who were assigned the iTech medium also used a Logitech USB Headset.

- 3) The delay time for each participant before starting the survey was the same. The pre-experiment survey was started upon arrival into the experiment room. The post-experiment survey was started immediately after the participants had finished their task.
- 4) All participants were told not to discuss the experiment with their classmates to ensure that all participants had an equal knowledge of the experiment.

The experiment was conducted for three different mediums. Medium I was the book entitled 'Learning the vi Editor' published by O'Reilly. Medium II was a search engine on top of an electronic version of the manual from medium I. To generate this medium, an electronic copy of medium I was used. Each section in the manual was separated and saved as an individual file. Once the copy was decomposed into its individual sections, Google Desktop [Goo06] was installed on the experiment computer. The preferences in Google Desktop were set to only search a specific folder on the experiment computer's hard drive. The folder contained electronic copies of the individual sections of the manual. To access the online medium, a floating desk bar positioned in the top right corner of the monitor was used. The participant would enter their search in the floating desk bar and a list of all relevant documents was presented to the participant. Medium III was iTech. iTech was populated with the knowledge from the book manual. The solutions/answers indexed in iTech are the same electronic copies of the individual sections of the manual used for the online medium. Consistency in content

was maintained across all three mediums to reduce the probability that any difference in search and/or task completion time were due to an independent variable other than medium.

Medium	Reference	Source
Paper	I	‘Learning the vi Editor’
Search Engine	II	Decomposed sections of ‘Learning the vi Editor’
iTech	III	Decomposed sections of ‘Learning the vi Editor’

**Table 1: Medium Type**

Twenty participants used Medium I, twenty-four used Medium II and thirty used Medium III. The experiment began by having each participant fill out the pre-experiment questionnaire. Once they completed the questionnaire, the participants were given the Information Sheet (see *APPENDIX A*) and the Instruction Sheet (see *APPENDIX B*) to read. When the participant completed the reading, they were assigned a medium. If Medium I was chosen, the participant was given the book and informed that they would be using the book to assist them in completing the task. If Medium II was chosen, the participant was directed to the floating desk bar and instructed that they would be using a search engine on top of an online manual to assist them in completing the task. Finally, if Medium III was chosen, the participant was instructed to put and adjust the headset comfortably. iTech was launched and the participant listened to his welcome prompt. When each participant was assigned a medium, they began their task.

The participant was directed to the SecureCRT application. In SecureCRT, the prompt is within the directory containing the file to be edited, *example.txt*. The participants were instructed by the experimenter that they would be accessing vi and the file from the current prompt. The video recorder, aimed at the computer monitor, was started and the participant began the task.

The task was selected from the Exploring Microsoft Office 2003 textbook [Gra03]. The participant must figure out how to open the specified file and perform edits on the file. The edits consist of deleting individual words, changing words, changing characters, deleting sentences and inserting sentences and paragraphs. When the edits are complete, the participant must save and exit the file. The participant is expected to use the medium provided for assistance in completing the task. Once the participant completes the task, they are instructed to fill out the post-experiment questionnaire. During the experiment, a set of resulting data was collected by following the data collection methods, which will be introduced next.

#### **4.3. DATA COLLECTION METHODS**

To achieve the objectives of the experiment, the following data were measured and collected:

1. Spoken query metrics based on the average number of spoken queries per participant (for Medium III).

2. Written query metrics based on the average number of solutions referenced (Medium II).
3. Task success metrics.
4. Interface Quality metrics [WLKA97], based on Recognition error rates of spoken queries, the number of ASR rejections.
5. User satisfaction

This section describes two different approaches used to collect the data: video recordings and user survey. *Table 2 Experimental Instruments and Measures* provides an overview of the experimental instruments and measures.

Instrument	Description
Pre-experiment Questionnaire	User background, demographics, computer literacy, etc.
Performance data	Time, QRA accuracy
User Observations	Qualitative and quantitative observations
Post-experiment Questionnaire	User satisfaction

**Table 2 Experimental Instruments and Measures**

#### 4.3.1. PRE-EXPERIMENT QUESTIONNAIRE

The pre-experiment questionnaire gathered general information about the participants to assess whether they met the criteria established for classification as a vi editor novice. Data gathered included such general identifiers as age, gender and major.

The second group of questions ascertained the participant's familiarity with computers. It posed questions on how long they had used a computer, how often, computer programming experience and experience with specific computer applications like word processors. Details of the pre-experiment questionnaire can be found in *APPENDIX E*.

#### 4.3.2. PERFORMANCE DATA AND USER OBSERVATIONS

Performance data was collected via video tape. All the experiment interactions were videotaped. These recordings were mainly used to measure time for searches, reading and task completion. Also, some characteristics of spoken queries such as average number of spoken queries per search and per user, number of recognition errors and total number of spoken terms per query were derived from the participant's utterances. In addition to the performance data, informal user observations and formal user observations were collected.

#### 4.3.3. POST-EXPERIMENT QUESTIONNAIRE

The post-experiment questionnaire was designed to gather information about how the participants assessed the system. There are two post-experiment questionnaires. One is for the participants that used the iTech medium and the second is for all other participants. Part I of the questionnaire is the identical for both versions. It gathered overall participant ratings using six bi-polar rating scales. The second part of the



questionnaire included a series of Likert-type scales where participants rated their reactions to the system via statements concerning whether they found the medium easy to use, did they know what to do, etc. The version presented to the users of the iTech medium, includes statements with respect to iTech and the participants reactions to the agent. Finally, the participants were asked to share any suggestions or comments they had regarding the medium. Details of the post-experiment questionnaire can be found in *APPENDIX F* and *APPENDIX G*.

#### **4.4. RESULTS AND DISCUSSION**

This section summarizes and discusses the results from the empirical comparison of Mediums I, II and III, including both quantitative and qualitative data and analysis. A summary of the participant background obtained from the pre-experiment questionnaire will be presented first. This will be followed by the analyses of quantitative data collected during the experiment task with respect to the major aspects of spoken query metrics and task success metrics, considering the evidence with respect to the experimental hypothesis. A separate section will contain a comparison of participants' reactions to the three mediums.

##### **4.4.1. PARTICIPANT BACKGROUND**

A summary comparison of several quantitative measures appears in *Table 3: Participant Background Data*.

<b>Measurement</b>	<b>Medium I</b>	<b>Medium II</b>	<b>Medium III</b>	<b>Total</b>
	<b>N = 20</b>	<b>N = 24</b>	<b>N = 30</b>	
<b>Avg age</b>	19.15	19.22	22	20
<b>% female</b>	20%	37.5%	23.33%	26.67%
<b>Avg years computer use</b>	8.3	16.0	11.53	11.94
<b>English – 2<sup>nd</sup> Language</b>	N/A	N/A	6.67%	N/A

**Table 3: Participant Background Data**

The ages of all the participants ranged from 18 to 27 with a mean age of 20 years; 71% of all the participants were male and 29% were female. With respect to computer usage, the average number of years for all participants was 12 with a minimum of 8. These numbers validated the deduction that the majority of participants are comfortable using a computer. Also, with respect to creating or updating documents, all of the participants have updated or created more than 9 documents. This affords a second deduction that the task presented in the experiment did not warrant additional training. Therefore any learning that occurred during the completion of the task should be with respect to the medium used.

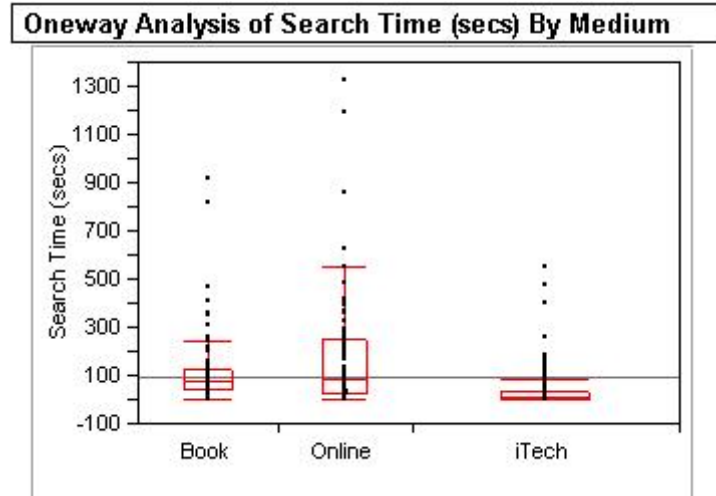
#### 4.4.2. PERFORMANCE DATA FINDINGS

The main goal of this research was to improve upon the search times for current manual usage, thereby improving upon the efficiency of manuals. To validate this improvement, there should be statistical significance in the data. The search times (the time the participant spent referencing their assigned medium until the correct solution is obtained) were analyzed and compared with respect to the mediums. As *Table 4: Mean Search Times by Medium* shows, Medium III was found to have the fastest average search time. This analysis was performed using Jmp In statistical software [SAS89].

First, the distribution of search times for each medium was investigated. The means and standard deviations calculated are displayed in *Table 4: Mean Search Times by Medium*. The Shapiro-Wilk test for normality was performed since it is resilient to the presence of outliers, which were present in the data, see *Figure 16: Side-by-Side Box Plots of Medium Search Times*.

Measurement	Book (secs)	Online (secs)	iTech (secs)
Mean	119.55	176.49	38.13
Standard Deviation	145.66	235.43	74.2

**Table 4: Mean Search Times by Medium**



**Figure 16: Side-by-Side Box Plots of Medium Search Times**

The Shapiro-Wilk's test provided very strong evidence to reject the null hypothesis which states that the means are normally distributed. With  $\alpha = 0.05$ , Book medium [ $W = 0.6195$ ,  $p = 0.0000$ ], Online medium [ $W = 0.6811$ ,  $p = 0.0000$ ] and iTech medium [ $W = 0.5174$ ,  $p = 0.0000$ ] all strongly support this deduction. This led to the application of the Kruskal-Wallis or Wilcoxon test to check for statistical significance.

The Kruskal-Wallis nonparametric analysis of variance provides a method for coping with data that contain extreme outliers and that have more than 2 independent variables. It does this by replacing the observation values by their ranks in a single sample and applying a one-way analysis of the F-test on the rank-transformed data [RS02]. The result of this test [ $F(1,2) = 106.9946$ ,  $p < .0001$ ] was a Kruskal-Wallis test statistic of 106.9946 with a p-value  $< .0001$  from a chi-square distribution with 2 degrees of freedom, The null hypothesis for this test states that the search time means for the

mediums are equal. The Kruskal-Wallis test provided strong evidence to reject this null hypothesis. Thereby, there is statistical significance that the search time means are different. Thus, there was strong evidence to reject the null hypothesis which stated that the means were equal. While the Kruskal-Wallis test allows for the comparison between three or more unpaired groups, it does not allow for deductions between specific pairs or means. The resulting p-value, which is very small, indicates that the confident deduction that the difference in the group means is not a coincidence can be made. However, this does not mean that every group differs from every other group. The Kruskal-Wallis test only determines that at least one group differs from one of the others. Thus a post test was applied to determine which groups differ from which other groups.

The post test applied was the Tukey-Kramer procedure. This test analyzes data of unequal sample sizes and determines whether the differences between all existing pairs are due to coincidence [RS02]. The results of the Tukey-Kramer test provided very strong evidence that the differences in the pairs of means were statistically significant see *Figure 17: Tukey-Kramer Test Results*. The positive values between each pair of means indicate that their differences are significantly different. Thus, there is sufficient evidence to deduce that the independent variable of medium type had a statistically significant effect on the search times, with the search times for the iTech medium being the most expeditious. Analysis continued to determine if the positive effects of search time were able to influence task completion time.

	Online	Book	iTech
Online	-54.024	3.698	91.582
Book	3.698	-52.466	35.528
iTech	91.582	35.528	-38.200

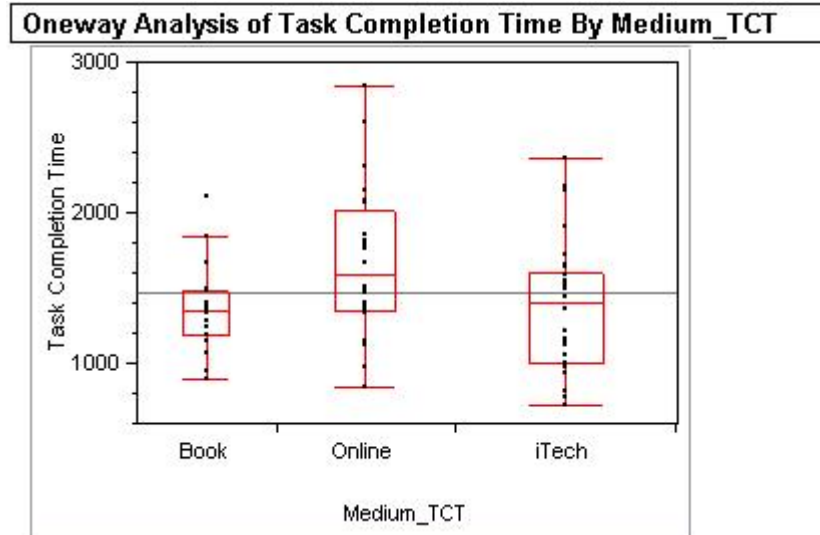
Positive values show pairs of means that are significantly different.

**Figure 17: Tukey-Kramer Test Results**

The same tests applied to the medium search times were applied to task completion times. The mean times are displayed in *Table 5: Mean Task Completion Time by Medium* and the normality spreads in *Figure 18: Side-by-Side Box Plots of Medium Task Completion Time*. Preliminary observations indicate that Medium I (Book medium) had the fastest average task completion time. The Shapiro-Wilk's test did not provide sufficient evidence to either strongly reject or fail to reject the null hypothesis which states that the means are normally distributed. With  $\alpha = 0.05$ , Book medium [ $W = 0.9395$ ,  $p = 0.229$ ], Online medium [ $W = 0.9664$ ,  $p = 0.582$ ] and iTech medium [ $W = 0.9501$ ,  $p = 0.199$ ] all recommend the failure to reject the null hypothesis, indicating that the distributions are fairly normal. As a result, the Kruskal-Wallis was applied to check for statistical significance.

Measurement	Book (secs)	Online (secs)	iTech (secs)
Mean	1360.58	1666.63	1377.87
Standard Deviation	290.95	500.1	420.87

**Table 5: Mean Task Completion Time by Medium**

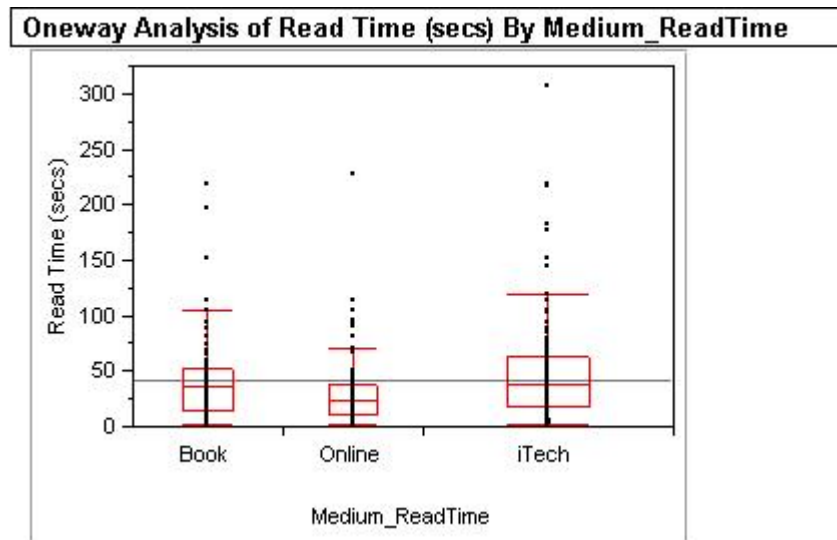


**Figure 18: Side-by-Side Box Plots of Medium Task Completion Time**

Application of the Kruskal-Wallis yielded no significant differences. The result of this test [ $F(1,2) = 5.7065$ ,  $p = 0.0577$ ] suggests a failure to reject the null hypothesis which states that the differences in the mean times is due to coincidence. Therefore, there is no statistical significance that indicates the task completion time means are different. An investigation as to the cause of this effect, led to the effect of reading times. Reading time was recorded as the time the participant spent reading and understanding the solution once it was presented to the user by the respective medium. This recorded time represented the time from the appearance of the solution on the monitor to the time the participant touched the keyboard. Analysis of the reading times yielded the following results.

Measurement	Book (secs)	Online (secs)	iTech (secs)
Mean	42.36	33.09	47.77
Standard Deviation	38.34	33.08	45.34

**Table 6: Mean Read Times by Medium**



**Figure 19: Side-by-Side Box Plots of Medium Read Times**

On preliminary observations, there is very little difference between the means for the different mediums. Application of the Shapiro-Wilks test for normality yielded these results; Medium I (book medium) [ $W=0.7854$ ,  $p<.0000$ ], Medium II (online medium) [ $W = 0.7166$ ,  $p < .0000$ ] and Medium III (iTech medium) [ $W = 0.7694$ ,  $p < .0000$ ] provide very strong evidence that they are not normally distributed. As the Kruskal-Wallis test is just as effective on normal distributions and to allow for consistency of applied tests, it was applied to the read time data. On application of the test, the result [ $F(1,2) = 9.1906$ ,  $p$



= 0.0101] indicates that there is evidence that the difference between the means is statistically significant. However, further investigation as to which pairs were significantly different was required. The Tukey-Kramer test was applied and demonstrated that only the difference between the online and iTech mediums was statistically significant. Thus the effect of read time nullified the improvements in search time generated by the iTech medium. Personal observations revealed that when the solution was found, several participants encountered difficulties understanding the text. It should be noted that the solutions were all identical regardless of the medium used. Also, a significant proportion of participants did not read the solution carefully and as a result either had to return to the solution several times, or implemented an incorrect action that led them further away from the correct action. These results suggest improvements in the content and understandability of technical communications would increase the improvements in search time provided by the iTech medium.

The next performance measurement analyzed was task success. Task success was determined by comparing the file updated by each participant to a correct version of the updated file. 95% of all participants successfully completed the task using one of the three mediums provided. The final performance measurement analyzed was the accuracy of the question resolution algorithm which is influenced by the spoken query metrics.

#### 4.4.2.1. SPOKEN QUERY METRICS

There were 298 spoken queries submitted by 30 participants with an average of 10 queries per participant. Of those participants, 23 were male and 7 were female. English was the second language for only two participants, however through personal observation it was discovered that the majority of participants spoke with heavy Southern accent. The average number of queries spoken per participant is very high. This was due to high recognition errors. The recognition errors were the result of:

1. Heavy southern accents
2. Participants not waiting on recording box before speaking
3. Delayed processing of recognized speech

To circumvent the effect of the recognition errors on the QRA, all spoken queries were transcribed and applied to iTech via text. This was done to verify that if the question had been correctly recognized, the correct solution would have been presented to the participant. 38.59% or 115 of the spoken queries were unique from each other and from the questions already residing in the KR. Yet, they yielded a success rate of 89.60%. Thereby contributing not only to the accuracy of the QRA, but also to the deduction that not every question or enunciation must reside in the KR for the answers first approach to be successful. Further investigation continued on the unsuccessful queries – 10.4%. It was found that for every 3 out of 4 queries in this set had a length of 3 or less words. This provided evidence to the hypothesis that the A1 approach is better suited for longer

questions. Another trend revealed from personal observation was the difference in question lengths for the online and iTech mediums. A distinct difference in the length and type of query entered for the online manual and the questions posed to iTech was observed. The online medium received keyword queries, while iTech received more naturally flowing questions, as shown in the excerpt of transcribed queries below. This increased length contributed to the high accuracy of the A1 approach and the question resolution algorithm. A discussion of the users' reactions to the individual mediums will follow.

<i><b>Query</b></i>
How do I create a new paragraph
How do you open a file
How do you edit text
How do you edit a file
How do I delete text from a file
How to dump the current file to text

**Figure 20: Excerpt of Transcribed Queries**

<i><b>Keywords</b></i>
new paragraph
file open
add text
delete text from file

**Figure 21: Examples of keyword searches**

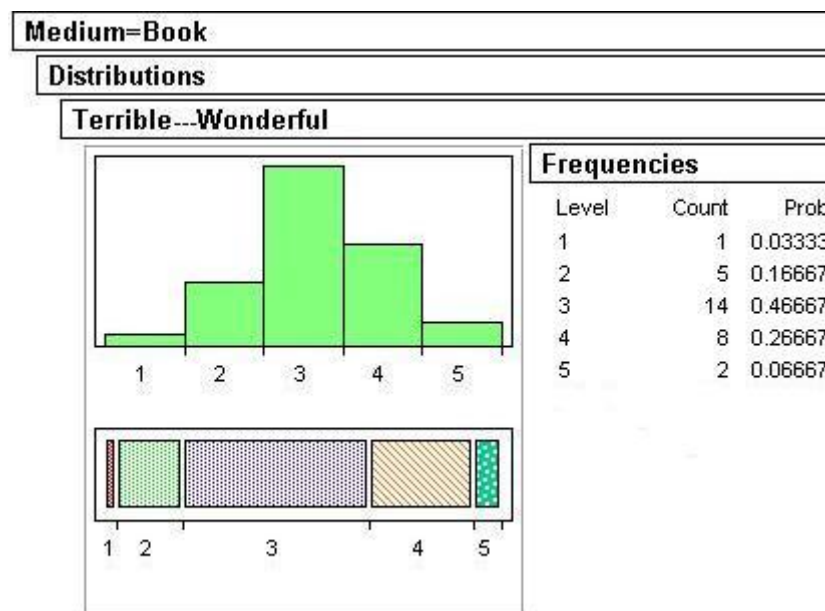
#### 4.4.3. USER SATISFACTION

The post-experiment questionnaire collected users' reactions via two rating scales. The first rating scale included a five-point bi-polar scale. This scale presented several qualities that might influence usability to be rated on the bi-polar scale. The means are shown in *Table 7: Bi-polar Rating Scales Assessing General Usability*. For each of these scales a higher rating indicates a number closer to the positive side except for the anchor of usable to not usable. For this anchor a higher rating indicates a number closer to the negative side. A quick review suggests that the participants' reactions to iTech were generally more favorable than the other two mediums. However, investigation of just the means did not provide a complete picture of the users' evaluations. A review of the entire distribution for each rating was required.

Bi-Polar Scale Anchors	Book Ratings (Mean)	Online Ratings (Mean)	iTech Ratings (Mean)
Terrible – Wonderful	3.17	3.55	3.29
Frustrating – Satisfying	3.23	2.90	3.0
Dull – Stimulating	2.93	3.62	3.79
Usable – Not Usable	2.4	2.31	2.57
Boring – Fun	2.9	3.38	3.64

**Table 7: Bi-polar Rating Scales Assessing General Usability**

The most interesting results were found with respect to three scales: 1) terrible – wonderful, 2) dull – stimulating and 3) boring – fun. The five-point rating inherently assigns the score of 3 a neutral rating, with scores 1 and 2 being negative and scores 3 and 4 positive. Subsequently, the positive ratings were of particular interest. With respect to the book medium, one third of the participants rated that medium with a score of 4 and higher for the scale of terrible to wonderful. The online medium received 53.33% and iTech 65.52% for the same score values. These results are displayed in *Figure 22: Book Medium Wonderful--Terrible Bi-polar Distribution*, *Figure 23: Online Medium Wonderful--Terrible Bi-polar Distribution* and *Figure 24: iTech Medium Wonderful--Terrible Bi-polar Distribution*.



**Figure 22: Book Medium Wonderful--Terrible Bi-polar Distribution**

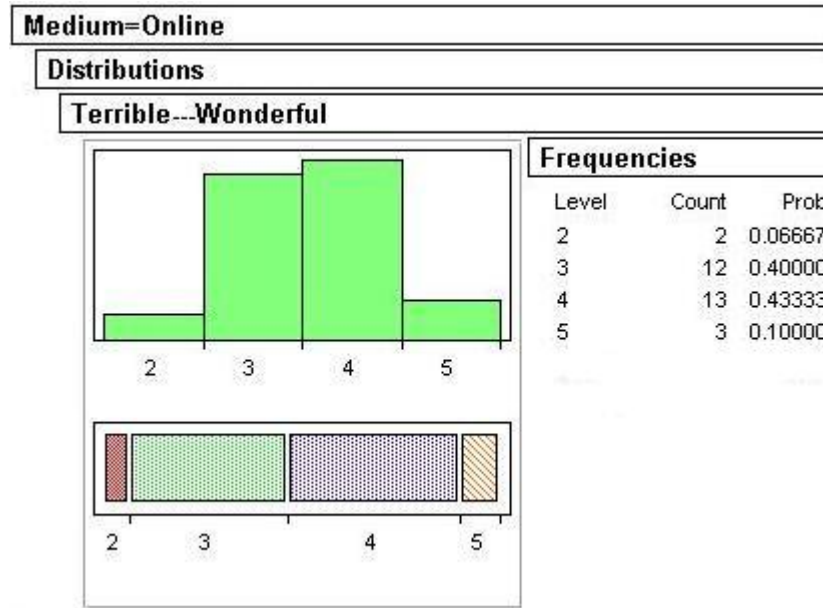


Figure 23: Online Medium Wonderful--Terrible Bi-polar Distribution

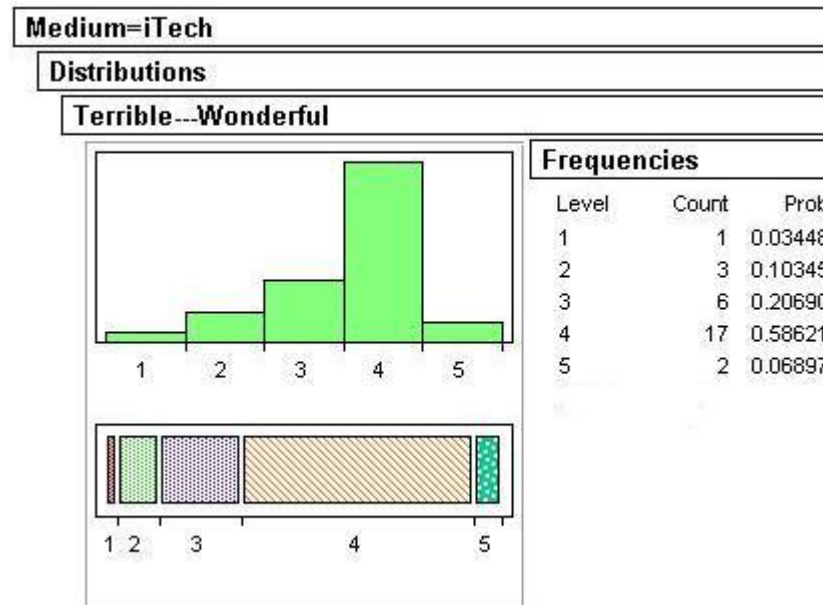
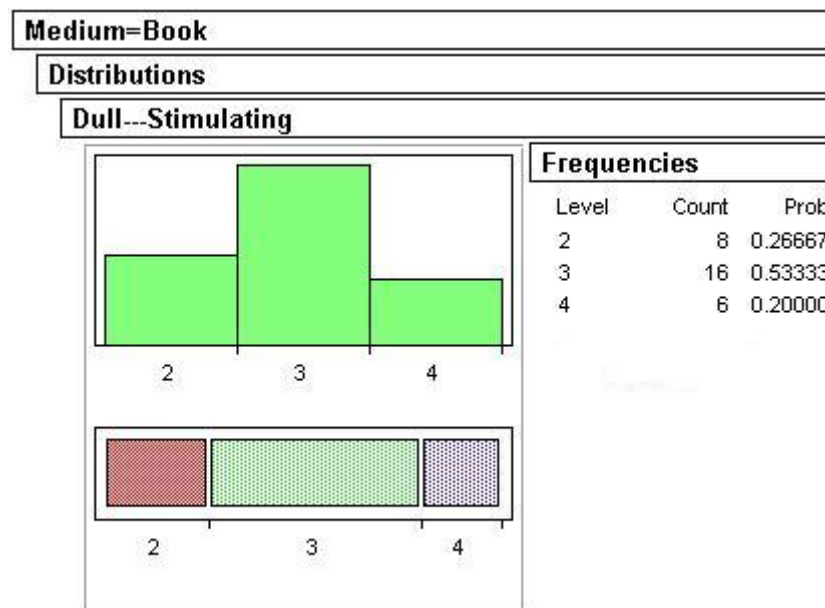


Figure 24: iTech Medium Wonderful--Terrible Bi-polar Distribution

In the scales of dull to stimulating and boring to fun, iTech also received the highest scores with respect to the other two mediums. For these scales, a much larger disparity in the distribution of scores is observed with respect to the book medium verses the online and iTech mediums. These results reinforce the popularity trends of the internet and its subsequent applications. These distributions are displayed in the following six figures.



**Figure 25: Book Medium Dull--Stimulating Bi-polar Distribution**

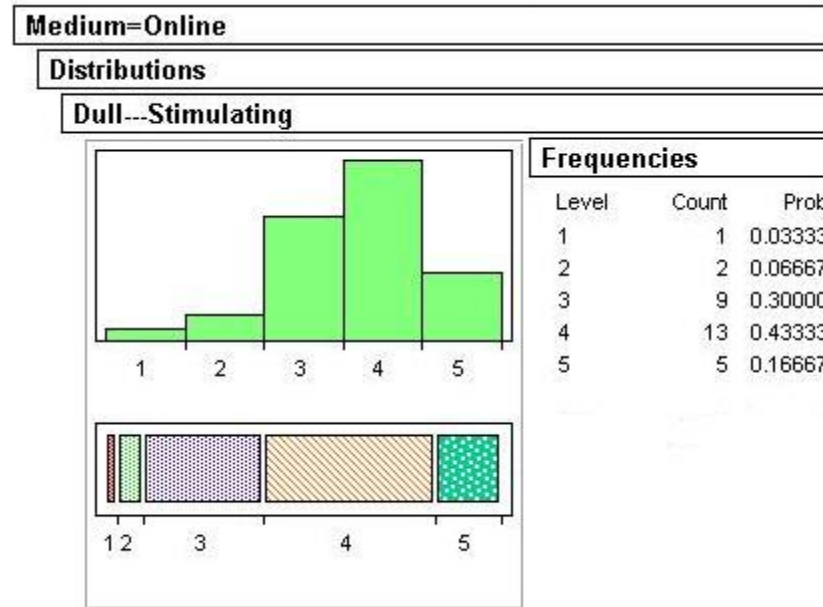


Figure 26: Online Medium Dull--Stimulating Bi-polar Distribution

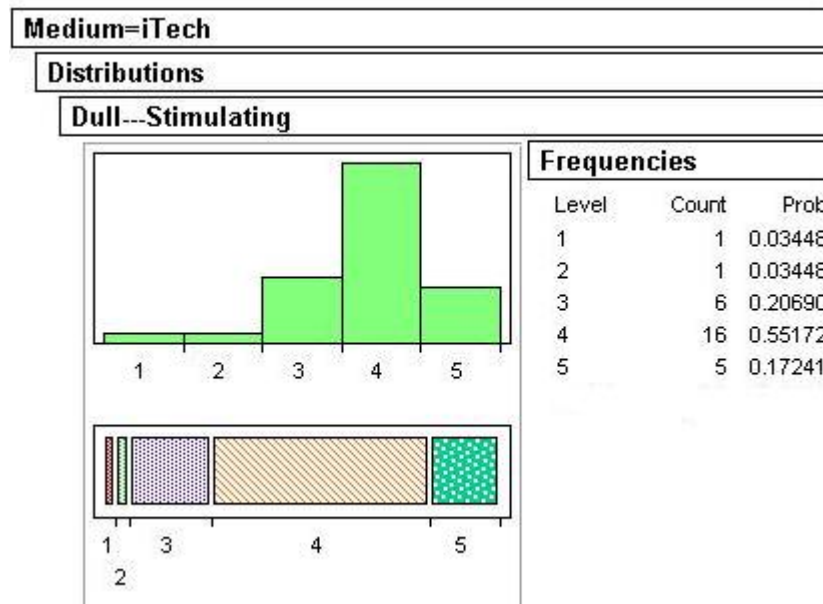


Figure 27: iTech Medium Dull--Stimulating Bi-polar Distribution



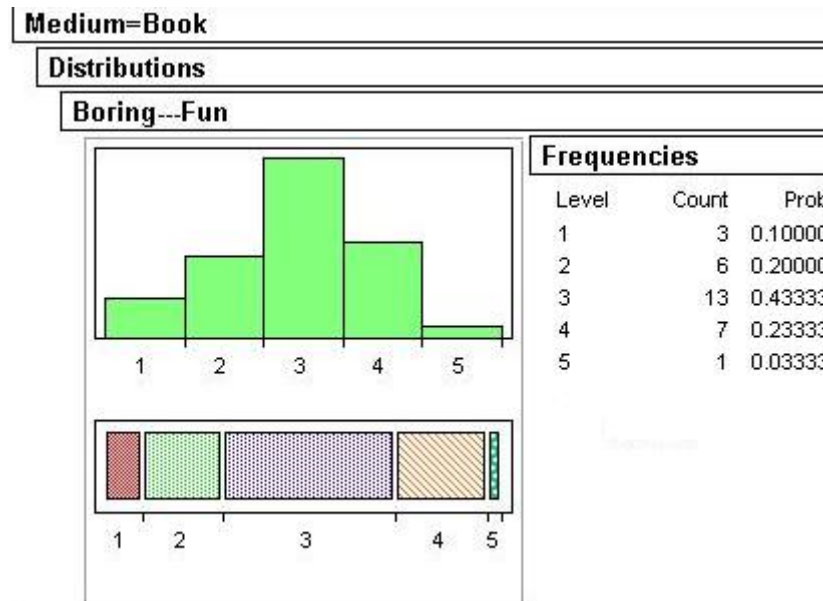


Figure 28: Book Medium Boring--Fun Bi-polar Distribution

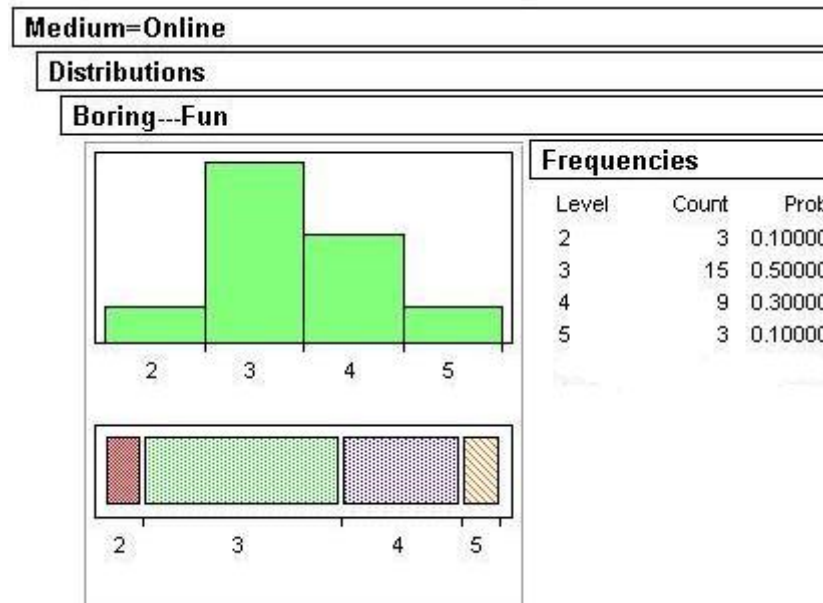
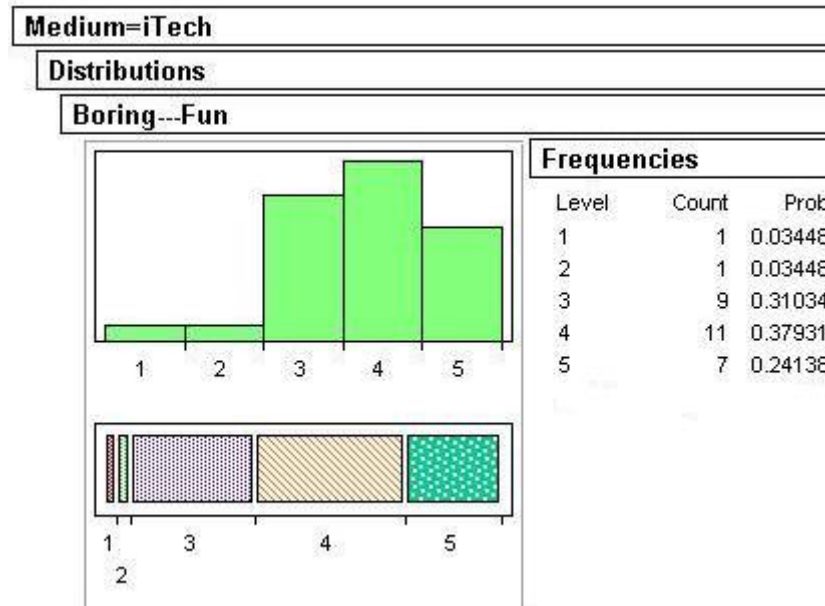


Figure 29: Online Medium Boring--Fun Bi-polar Distribution



**Figure 30: iTech Medium Boring--Fun Bi-polar Distribution**

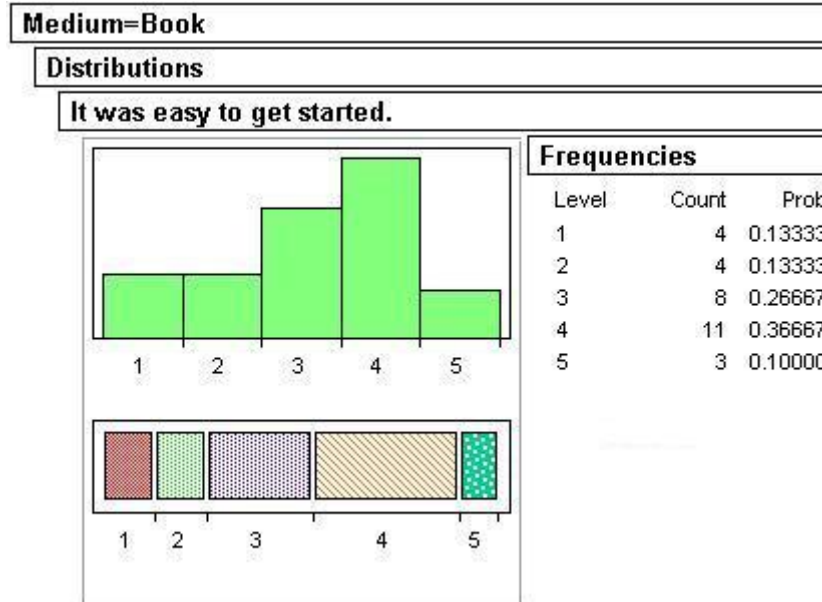
The second set of rating scales consisted of items designed to assess reactions to specific aspects of the participants' interaction experience. These scales each contain an assertion e.g. 'The medium was easy to use', to which the participants responded using a five-point scale. This scale contained the following ratings: Strongly Agree, Agree, Neutral, Disagree and Strongly Disagree. Each rating was assigned a weight. This weight was used for statistical analysis. The weighting was as follows:

- Strongly Agree = 5
- Agree = 4
- Neutral = 3

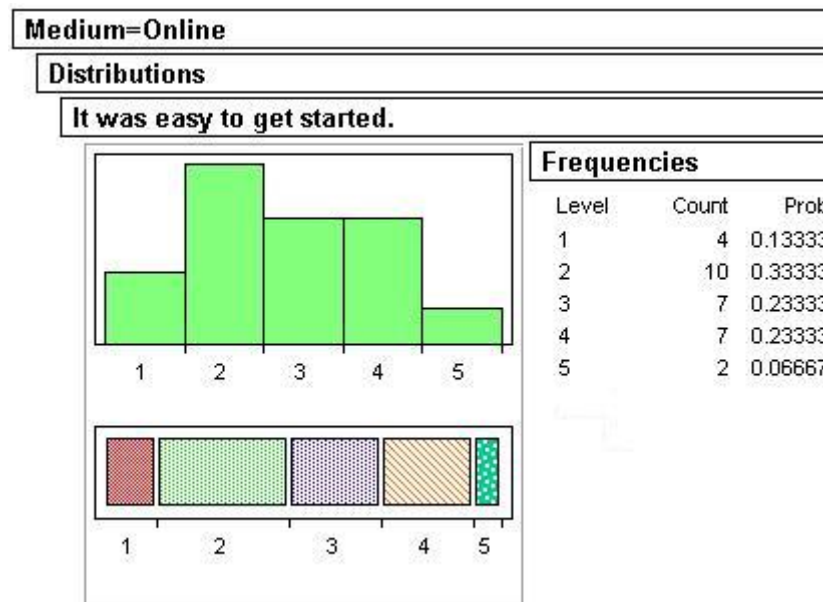
- Disagree = 2
- Strongly Disagree = 1.

There were two versions of the post-experiment questionnaire. One version contained additional questions unique to iTech that are not applicable to the book and online mediums therefore a separate questionnaire was generated. This version will be referred to as Version II (see *APPENDIX F*) and the other questionnaire as Version I (see *APPENDIX G*). Version I contains 10 Likert-style ratings and Version II contains 22. The first 9 ratings for each questionnaire are identical and as a result were compared across all three mediums.

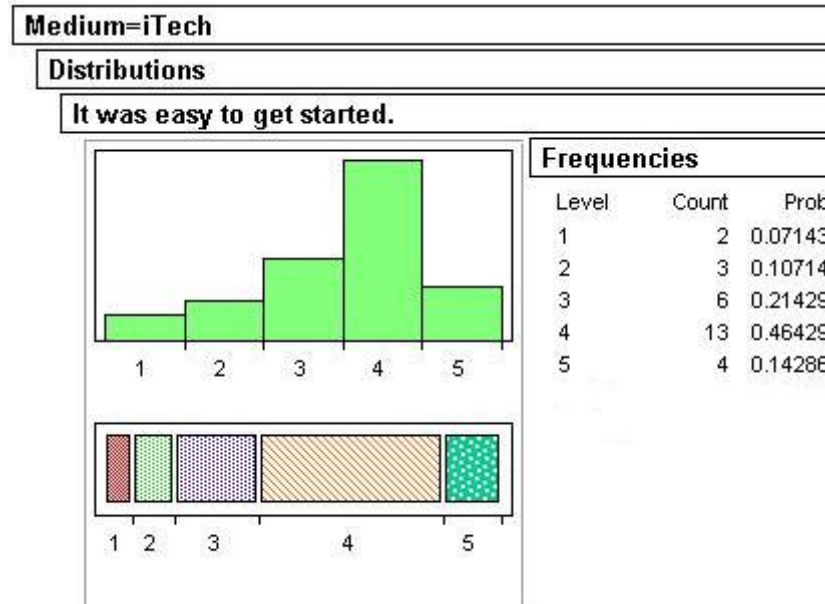
The first property investigated was the affordance of the mediums. This property was retrieved from the question, “It was easy to get started.”. Results show that iTech received a score of 4 or higher from 60.7% (see *Figure 33: iTech Medium Affordance Distribution*) of the participants, while the book and online mediums received 46.67% (see *Figure 31: Book Medium Affordance Distribution*) and 30.0% (see *Figure 32: Online Medium Affordance Distribution*) respectively. This data is in concordance with the trends found in the mediums’ search times. The online medium had the worst average search time with iTech having the best, reinforcing the guideline that an application’s affordance is an important feature of the application’s success.



**Figure 31: Book Medium Affordance Distribution**



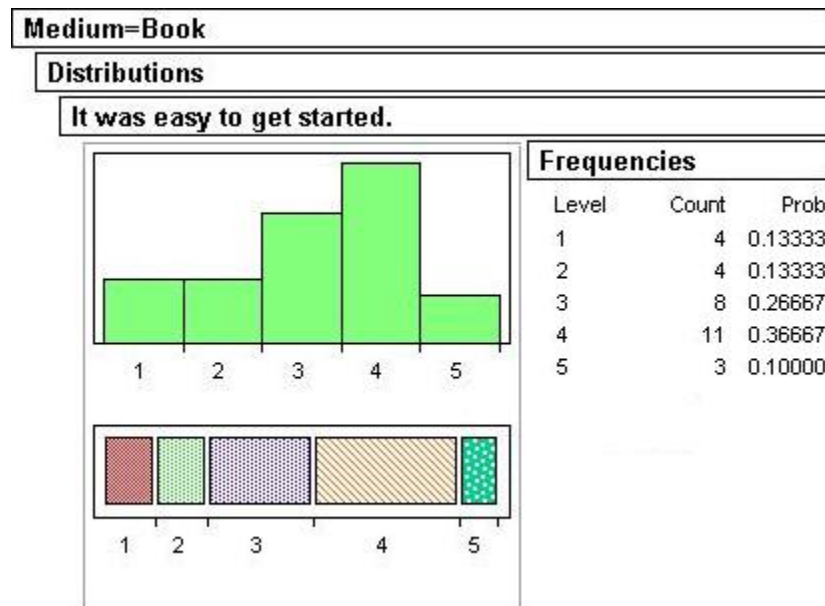
**Figure 32: Online Medium Affordance Distribution**



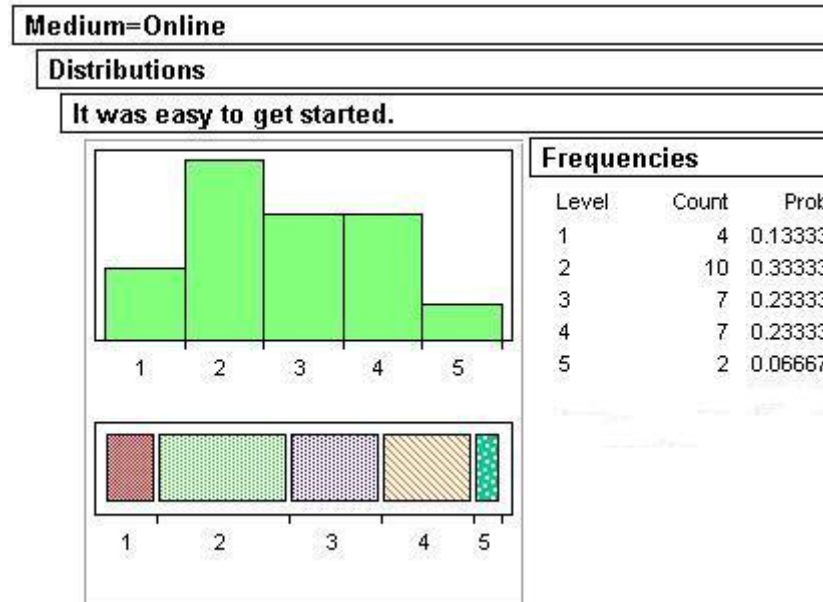
**Figure 33: iTech Medium Affordance Distribution**

The next property investigated was the cognitive demand of the task to be completed. With the scores for the understanding of document updates being all over 80.0% for the mediums, suggests that task selected to decrease any additional training required to complete the task was accomplished. The results for the property of ‘ease of use’ reflect the problems with speech recognition accuracy. As mentioned previously, there were problems with recognition accuracy due to heavy southern accents and incorrect usage of the recording box. Subsequently, though the range for the medium averages is small, the scores for the iTech medium are the lowest in response to the statement, “It was easy retrieving an answer”. The results are as follows: book medium – 63.3%, online medium – 50.0% and iTech medium – 48.27% for scores of 4 or higher. In spite of the recognition accuracy issues, iTech received the highest ratings with respect to knowing how to use

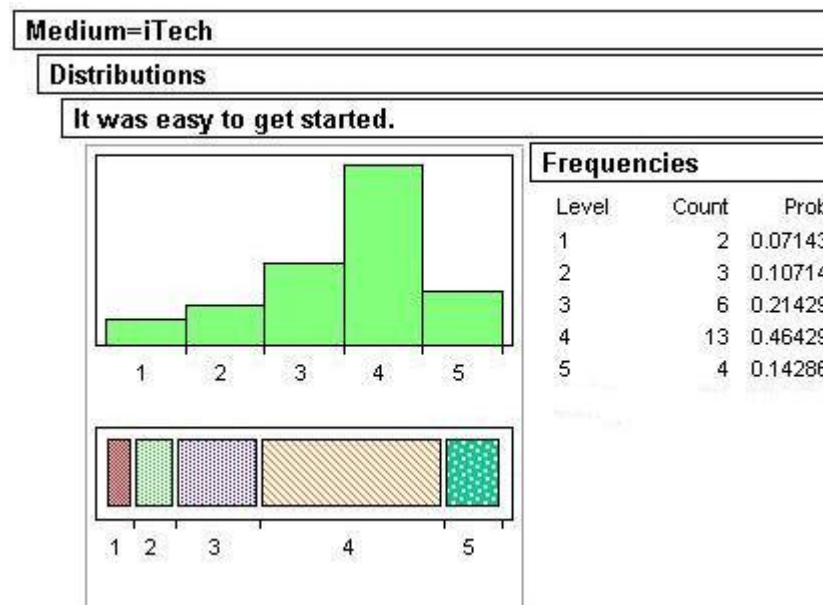
the medium. The distributions are displayed below in *Figure 34: Book Medium Getting Started Distribution*, *Figure 35: Online Medium Getting Started Distribution* and *Figure 36: iTech Medium Getting Started Distribution*. Analysis then shifted towards the users' reactions of the iTech medium.



**Figure 34: Book Medium Getting Started Distribution**



**Figure 35: Online Medium Getting Started Distribution**



**Figure 36: iTech Medium Getting Started Distribution**

Prior to analysis, the statements unique to the iTech medium were placed into one of possible six categories. These categories represent the six factors investigated in user attitude to speech systems [HG01]. The statistics of these categories are shown below.

<b>Likert-type Scale Item</b>	<b>Mean</b>	<b>% with Score of 4 or higher</b>
iTech worked as I expected it during the task.	3.41	55.1
I was confident that iTech would be able to help me.	3.86	72.4

**Table 8: Rating Scales Assessing Cognitive Modeling**

<b>Likert-type Scale Item</b>	<b>Mean</b>	<b>% with Score of 4 or higher</b>
iTech gave me the correct answers.	3.69	72.41
iTech had problems understanding me.	3.52	68.97

**Table 9: Rating Scales Assessing Perceived System Response Accuracy**

<b>Likert-type Scale Item</b>	<b>Mean</b>	<b>% with Score of 2 or lower</b>
I had problems understanding iTech.	2.32	71.14

**Table 10: Rating Scale Assessing Cognitive Demand**

<b>Likert-type Scale Item</b>	<b>Mean</b>	<b>% with Score of 4 or higher</b>
I would use iTech again.	3.82	78.57
I liked the appearance of iTech.	4.18	89.29
I would have preferred a female technician.	2.97	13.79
I would have preferred iTech having no face, just a voice.	2.10	7.9

**Table 11: Rating Scale Assessing Likeability**



<b>Likert-type Scale Item</b>	<b>Mean</b>	<b>% with Score of 4 or higher</b>
iTech would be easy to use by people who don't know a lot about computers.	3.62	75.86

**Table 12: Rating Scale Assessing Habitability**

<b>Likert-type Scale Item</b>	<b>Mean</b>	<b>% with Score of 4 or higher</b>
iTech was fast enough in response to my question.	4.03	86.21

**Table 13: Rating Scale Assessing Speed**

Analysis of the data representing the users' reactions with respect to each medium yielded some interesting results. The users liked the appearance of iTech and would reuse the application. They were able to understand iTech, thought that the application retrieved their answers in an expedient fashion and agreed that computer novices would be able to use the application. The high user satisfaction ratings were solidified by their additional comments. These comments included: "Worked greater than expectations based on previous speech help programs...", "Pretty easy to use. User friendly", "I really enjoyed iTech ...the layout and technology used was great" and "It was overall very helpful and would be useful for people whom are computer literate".

#### **4.5. EXPERIMENT SUMMARY**

The main experiment goal was to observe an improvement in search times for iTech as compared with the mediums of paper and online. At the end of the experiment, there was statistical significance that supported the improved search time rates for the

iTech medium. During the experiment, it was observed that spoken queries are longer than their written counterparts. This fact contributed to the success of iTech. Also, in spite of the problems with speech recognition accuracy, the majority of participants were able to complete the task successfully and also enjoyed the experience. As a result, the generation of a medium for technical communication whose interaction style more closely matches human's natural information-seeking process was feasible and generated promising results.

## **5. SUMMARY**

iTech is an interactive technical assistant that uses a new methodology for conversational question answering. Its main goal is to provide users with a medium for technical communications that accommodates their natural process for fulfilling their information needs. Through experimentation, iTech has shown that its improvements in the search time are statistically significant. It has also proven to be a desirable medium for technical assistance. iTech accomplished these via the Answers First approach. Therefore, iTech makes several contributions to various field of research.

### **5.1. CONTRIBUTIONS**

This research made the following broad contributions to the fields of Technical Communications and Conversational Question Answering:

- Identified and addressed the limitations of technical documentation.
  - Research into the improvement of current technical documentation is ongoing.
- Advancements in technology are being recognized as potential solutions to the current problems. However, attention must be directed towards making the usage of these communications more closely match our natural information-seeking process. iTech accomplishes this using the Answers First approach.
- Introduced a new medium for technical communications that improves on existing mediums.

- iTech provides a multimodal interactive interface. Thereby affording a more natural interaction style between the user and technical communication.
- Introduced a new methodology for conversational question answering called Answers First.
  - Traditional question answering techniques involve language processing to be performed on the question before the query is executed. Answers First removes this additional processing and suggests that all the words in the original questions are vital towards the retrieval of the accurate solution.
- Provided more evidence for a multi-disciplinary approach towards technical communications.
  - The development of iTech included input from several fields including but not limited to information retrieval, spoken user interfaces, interactive assistants and conversational question answering.

## **5.2. DIRECTIONS FOR FUTURE RESEARCH**

There are a number of areas that warrant further investigation. These areas either describe aspects of this work that are proposed for further research and are based upon insights from the experimental results.

1. The introduction of a context-aware agent. Currently iTech has no knowledge of the user's progress or intentions other than the questions posed by the user. While iTech retrieved the user's answer more expediently, there was not a significant

improvement in the user's task completion time due to problems encountered reading and understanding the presented solution. Giving iTech the capability to know the status of the user would result in iTech recognizing whether the user is following the instructions provided. If not, iTech could alert the user and wait for confirmation on whether the user's actions are deliberate or not.

2. Currently the Knowledge Repository in iTech is manually generated. First the developer generates questions to the answers present. Then an adaptation of the user edit process is used. Here subjects from the representative user pool use the system (a working system or via Wizard-of-Oz), to accomplish a task. Any gaps in the question knowledge repository are subsequently filled. Development of an automatic question generator would eliminate this effort.
3. The main issue with iTech is the recognition accuracy. While improvements in continuous speech recognition engines would help reduce this problem, improvements in iTech can also provide contributions. With the addition of a context-aware agent, iTech would be able predict the user's future directions. As a result, the grammars that covered the most highly expected direction would be utilized first. If misrecognitions continue, then the grammar is expanded to include other directions within the application.

## REFERENCES

- [AC06] Answers Corporation. Speech Recognition. [Online] Available: <http://www.answers.com/speech%20recognition> 2006.
- [Alo05] ALon, G. Key-Word Spotting – The Base Technology for Speech Analytics White Paper, Natural Speech Communication Ltd, July 2005.
- [AR00] Androutsopoulos, I., Ritchie, G. Database Interfaces. *Chapter 9 in handbook of Natural Language Processing*, ed. R.Dale, H. Moisl and H. Somers. New York: Marcel Dekker Inc, 2000.
- [Atl81] Atlas, M. The User Edit: Making Manuals Easier to Use. ACM SIGDOC Asterisk *Journal of Computer Documentation*, Vol 22, pages 5 -6, 1981.
- [Bab91] Baber, C. Human factors aspects of automatic speech recognition in control room environments. In *Proceedings of IEEE Colloquium on Systems and Applications of Man-Machine Interaction Using Speech I/O*, 10/1 – 10/3, 1991.
- [Bar04] Barlow, L. The Spider's Apprentice—A Helpful Guide to Search Engines. <http://www.monash.com/spidap.html> Last updated May 11<sup>th</sup>, 2004.
- [Bay03] Bay, S. Cellular phone manuals: users' benefit from spatial maps. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems*. ACM Press, New York, NY, pages 662-663, 2003.

- [BEM97] Beskow, J., Elenius, K., McGlashan, S. Olga A. Dialogue System With An Animated Talking Agent. *Proc. Eurospeech '97*, 1997.
- [BK03] Baylor, A. & Kim, Y. The Role of Gender and Ethnicity in Pedagogical Agent Perception. In G. Richards (Ed.), *Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2003* (pp. 1503-1506). Chesapeake, VA: AACE, 2003.
- [BP01] Berglund, E. and Priestley, M. Open-source documentation: in search of user-driven, just-in-time writing. In *Proceedings of the 19th Annual international Conference on Computer Documentation* (Sante Fe, New Mexico, USA, October 21 - 24, 2001). SIGDOC '01. ACM Press, New York, NY, pages 132-141, 2001.
- [BR03] Baylor, A., & Ryu, J. The effect of image and animation in enhancing pedagogical agent persona. *Journal of Educational Computing Research*, 28(4), pages 373-395, 2003.
- [BSH03] Baylor, A., Shen, E. & Huang, X. Which Pedagogical Agent do Learners Choose? The Effects of Gender and Ethnicity. In G. Richards (Ed.), *Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2003* (pp. 1507-1510). Chesapeake, VA: AACE, 2003.
- [CCBH02] Chen, L., Chen, S., Birnbaum, L. and Hammond, K.J. The Interactive Chef: A Task Sensitive Assistant. *7th International Conference on Intelligent User Interfaces*, San Francisco, CA, USA, ACM Press New York, NY, USA, 2001.

- [CCIM02] Cisco Systems Inc., Comverse Inc., Interl Corporation, Microsoft Corporation, Philips Electronic N.V., SpeechWorks International Inc., SALT Speech Application Language Tags (SALT) 1.0 Specification, July 15, 2002.
- [Col91] Coleman, V. Hardcopy to hypertext: putting a technical manual online. In *Proceedings of the 9th Annual international Conference on Systems Documentation* (Chicago, Illinois, United States). SIGDOC '91. ACM Press, New York, NY, pages 67-72, 1991.
- [DB01] Dybkjar, L. & Bernsen, N. O. Usability Evaluation in Spoken Language Dialogue Systems. In *Proceedings of the ACL workshop on Evaluation Methodologies for Language and Dialogue Systems*, Toulouse, France, 6-7 July, 2001.
- [DC04] Du, H., Crestani, F. Retrieval Effectiveness of Written and Spoken Queries: an Experimental Evaluation In *Proceedings of 6th International Conference On Flexible Query Answering Systems*, Lyon, France. June 2004.
- [FBBO92] Ferguson, W., Bareiss, R., Birnbaum, L., Osgood, R. ASK Systems: An Approach to the Realization of Story-Based Teachers. Institute for the Learning Sciences Report #22, Northwestern University, April, 1992.
- [FBH99] Franklin, D., Bradshaw, S., and Hammond, K. Beyond "Next slide, please": The use of content and speech in multi-modal control. *AAAI-99 Workshop on Intelligent Information Systems*, 1999.



[FBH00] Franklin, D., Bradshaw, S. and Hammond, K. Jabberwocky: You don't have to be a rocket scientist to change slides for a hydrogen combustion lecture. *International Conference on Intelligent User Interfaces*, New Orleans, Louisiana, USA, ACM Press New York, NY USA, 2000.

[FH01] Franklin, D. and Hammond, K. The Intelligent Classroom: Providing Competent Assistance. *5th International Conference on Autonomous Agents*, Montreal, Quebec, Canada, ACM Press, 2001.

[Goo06] Google Desktop Beta [Online] Available:  
<http://desktop.google.com/en/index.html> 2006.

[Gra03] Grauer, R. *Exploring MS Office XP and Exploring FrontPage 2003 plus the Train and Assess IT Generation*, Prentice Hall Publishing Co. ISBN: 0-536-83155-6, 2003.

[GWCP04] Glass, J., Weinstein, E., Cyphers, S., Polifroni, J., Chung, G., Nakano, N. A Framework for Developing Conversational User Interfaces. In *Proceedings of CADUL*, Funchal, Isle of Madeira, Portugal, Jan 2004.

[GZ03] Gilbert, J. E., Zhong, Y. Speech user interfaces for information retrieval. In *Proceedings of the twelfth international conference on Information and knowledge management*, New Orleans, LA, 2003.

[Hai04] Hailey, D.E. A Next Generation of Digital Genres: Expanding Documentation into Animation and Virtual Reality. In *Proceedings of the 22<sup>nd</sup> Annual International Conference on Design of Communication*, Memphis, October, 2004.

- [HBHD98] Horvitz, E., Breese, J., Heckerman, D., Hovel, D and Rommelse, K. The Lumiere Project: Bayesian User Modeling for Inferring the Goals and Needs of Software Users. *Fourteenth Conference on Uncertainty in Artificial Intelligence*, Madison, WI, 1998.
- [HG01] Hone, K.S., Graham, R. Subjective Assessment of Speech-System Interface Usability. Eurospeech 2001.
- [HTSR04] Hakulinen, J., Turunen, M., Salonen, E. and Räihä, K. Tutor design for speech-based interfaces. *2004 Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*, Cambridge, MA, USA, ACM Press, 2004.
- [JJ93] Johnson, H. and Johnson, P. Explanation Facilities and Interactive Systems. *First International Conference on Intelligent User Interfaces*, Orlando, Florida USA, ACM Press, 1993.
- [KBBO93] Kedar, S., Baudin, C., Birnbaum, L., Osgood, R., Bareiss, R. ASK How It Works: An Interactive Manual for Devices. *Conference on Human Factors in Computing Sciences*, pages 171 – 172, 1993.
- [KLM97] Keim, M., Lewis, D. D. and Madigan, D. Bayesian information retrieval: Preliminary evaluation. *Sixth International Workshop on Artificial Intelligence and Statistics*, Ft. Lauderdale, FL, USA, 1997.
- [KM92] Konstantinou, V. and Morse, P. Electronic documentation system: using automated hypertext techniques for technical support services. *10th annual international conference on Systems documentation*, Ottawa, Ontario, Canada, ACM Press, 1992.

- [KR01] Kirste, T., Rapp, S.: Architecture for Multimodal Interactive Assistant Systems. *Proceedings of the status conference "Mensch-Technik Interaktion"*, Saarbrücken, Germany, 2001.
- [Kra01] Krahmer, E. J.: The Science and Art of Voice. Interfaces Philips Research Report, Philips, Eindhoven, The Netherlands, 2001.
- [KSCL00] Klemmer, S. R., Sinha, A. K., Chen, J., Landay, J. A., Abookbaker, N., Wang, A. SUEDE: A Wizard of Oz Prototyping Tool for Speech User Interfaces. In *Proceedings of UIST*, page 110, 2000.
- [KWK03] Kvale, K., Warakagoda, N. D. and Knudsen, J. E. Speech centric multimodal interfaces for mobile communication systems. *Teletronikk* nr.2, 2003.
- [LSBH99] Leake, D., Scherle, R., Budzik, J., and Hammond, K. Selecting task-relevant sources for just-in-time retrieval. *Proceedings of the AAAI-99 Workshop on Intelligent Information Systems*, (Menlo Park, CA) AAAI Press, 1999.
- [Maj85] Major, J. H. Pulling it all together: a well-designed user's manual. In *Proceedings of the 13th Annual ACM SIGUCCS Conference on User services: pulling it all together*, pages 69 -76, Toledo, OH, 1985.
- [MS01] McMillan, J. H., Schumacher, S. *Research in Education, A Conceptual Introduction*. Addison Wesley Longman, Inc, 2001.
- [OC00] Oviatt, S. L., Cohen, P. R. Multimodal Interfaces That Process What Comes Naturally. *Communications of the ACM*, Vol 43, No. 3, pages 45 – 53, March, 2000.

- [OCVD00] Oviatt, S. L., Cohen, P., Vergo, J., Duncan, L., Suhn, B., Bers, J., Holzman, T., Winograd, T., Landay, J., Larson, J., and Ferro, D. Designing the User Interface for Multimodal Speech and Pen-based Gesture. Applications: State-of-the-Art Systems and Future Research Directions, *Human Computer Interaction* 15,4, 263-322, 2000.
- [Ovi03] Oviatt, S. L. Multimodal Interfaces. In the *Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*, J. Jacko and A. Sears Eds. Lawrence Erlbaum Assoc., Chapter 14, pages 286 – 304, Mahwah, NJ, 2003.
- [Ovi99] Oviatt, S. L. Ten myths of multimodal interaction. *Communications of the ACM*, Vol. 42, No. 11, pages 74 - 81, November, 1999.
- [Pie03] Pierce, R. Optimizing your documentation with the help of technical support. In *Proceedings of the 21st Annual international Conference on Documentation* (San Francisco, CA, USA, October 12 - 15, 2003). SIGDOC '03. ACM Press, New York, NY, pages 6-11, 2003.
- [RP81] Relles, N., Price, L.A. A User Interface for Online Assistance. In *Proceedings of 5<sup>th</sup> International Conference on Software Engineering*, Institute for Electrical and Electronic Engineers, pages 400 -408, New York, 1981.
- [RFQW05] Radev, D., Fang, W., Qi, W., Wu, H. and Grewal, A. Probabilistic Question Answering on the Web. *Journal of the American Society for Information Science and Technology*, Vol 56(6), pages 571 – 583, 2005.

- [RQZB01] Radev, D., Qi, H., Zheng, Z., Blair-Goldensohn, S., Zhang, Z., Fan, W. and Prager, J. M. Mining the Web Answers to Natural Language Questions. *CIKM*, pages 143 – 150, 2001.
- [RS97] Rich, C. and Sidner, C. L. COLLAGEN: When agents collaborate with people. *First International Conference on Autonomous Agents*, Marina del Rey, Ca USA, ACM Press, 1997.
- [RZBZ01] Radev, D., Qi, H., Zheng, Z., Blair-Goldensohn, S., Zhang, Z., Fan, W. and Prager, J. M. Mining the Web for Answers to Natural Language Questions. *CIKM*: pages 143 – 150, 2001.
- [Sam04] Sampson, G. The CHRISTINE Project. [Online] Available: <http://www.grsampson.net/RChristine.html>, Jun 12th, 2004.
- [Sam05] Sampson, G. The SUSANNE Analytic Scheme. [Online] Available <http://www.grsampson.net/RSue.html>, Jan 5th, 2005.
- [SAS89] SAS Institute JMP IN: Software for statistivcal visualization on the Apple Macintosh Cary, NC, 1989.
- [Shn98] Shneiderman, B. *Designing the User Interface: Strategies for Effective Human-Computer Interactio.*, 3<sup>rd</sup> edition, Addison-Wesley, Reading, 1998.
- [Sit06] SitePal: Now you're talking business. [Online] Available: <http://www.oddcast.com/sitepal/> 2006.
- [SK92] Saddler, H. J. and Kaplan, L. E. Choosing a medium for your message: what determines the choice of delivery media for technical documentation? *10th Annual*

*International Conference on Systems Documentation*, Ottawa, Ontario, Canada, ACM Press, 1992.

[SKD03] Stratica, N., Kosseim, L. and Desai, B. C. NLIDB Templates for Semantic Parsing. *Proceedings of Applications of Natural Language to Data Bases (NLDB 2003)*. pages 235 – 241, Burg, Germany, June 2003.

[SRZT01] Shriver, S., Rosenfeld, R., Zhu, X., Toth, A., Rudnicky, A., Flueckiger, M. Universalizing Speech: Notes from the USI Project. In *Proc. Eurospeech 2001*.

[Thi96] Thimbleby, H. Creating user manuals for using in collaborative design. In *Conference Companion on Human Factors in Computing Systems: Common Ground* (Vancouver, British Columbia, Canada, April 13 - 18, 1996). M. J. Tauber, Ed. CHI '96. ACM Press, New York, NY, pages 279-280, 1996.

[THRS05] Turunen, M., Hakulinen, J., Räihä, K., Salonen, E., Kainulainen, A. and Prusi, P. An architecture and applications for speech-based accessibility systems. *IBM Systems Journal*, Vol. 44, No 3: pages 485-504, 2005.

[THTS05] Tomko, S., Harris, T. K., Toth, A., Sanders, J., Rudnicky, A., and Rosenfeld, R. Towards efficient human machine speech communication: The speech graffiti project. *ACM Trans. Speech Lang. Process.* 2, 1, Feb., 2005.

[Ven88] Ventura, C. A. Why Switch from Paper to Electronic Manuals? *Proceedings of ACM Conference on Document Processing Systems*, ACM, New York, pages 111 – 116, Santa Fe, New Mexico, 1988.

[ZCCF01] Zachary, C., Cargile-Cook, K., Faber, B., Zachary, M. The Changing Face of Technical Communication: New Directions for the Field in a New Millennium. *Proceedings of the 19th Annual International Conference on Systems Documentation*, pages 248 – 260.

[ZLBD03] Zschorn, A., Littlefield, J. S., Broughton, M., Dwyer, B., Hashemi-Sakhtsari, A. Transcription of Multiple Speakers Using Speaker Dependent Speech Recognition. DSTO Technical Report, DSTO\_TR\_1498, 2003.

## APPENDIX A

# Information Sheet

### **Instructions:**

1. You are required to complete the task of updating a text document “example.txt”.
2. You will be given an information sheet that lists the updates that must be performed.
3. The task is complete when all of the specified updates have been performed.
4. You will be using the vi editor to update the document.
5. If you are familiar with vi, let the experimenter know now.....
6. You will be given one of three manual mediums to be used as reference for the vi editor.
7. Use the manual medium provided as you typically would.
8. If you are using iTech, you will communicate with iTech via speech.
9. When you have completed the task, let the experimenter know.
10. The experimenter will **not** answer any questions on the manual medium or the task during the process. The experimenter is there for observation ONLY.
11. When the updates have been completed, you will be given a questionnaire to be completed in the lab. This questionnaire must be returned to the observer.



## APPENDIX B

### Instructions:

You are required to update a document by completing the following tasks:

1. Open the file named “example.txt” using the vi editor.
  - a. Change the word “**worry**” to “**be concerned**” in the first sentence of the second paragraph. Change the word “**position**” to “**location**” in the last sentence of the second paragraph.
  - b. Delete the word “**simply**” from the second sentence in the third paragraph. Delete the word “**also**” in the second line of the last paragraph.
  - c. Go to the end of the third paragraph, be sure you are in insert mode, and add the sentence, **The insert mode adds characters at the insertion point while moving existing text to the right in order to make room for the new text.**
  - d. Change the “**s**” in the word “**test**” found in the fourth line of the second paragraph to “**x**”.
  - e. Delete the two sentences in paragraph three that describe the OVR indicator.
  - f. Create a new paragraph between paragraphs three and four, entering the following text: **There are two other keys that function as toggle switches of which you should be aware. The Caps Lock key toggles between upper- and lower case letters. The Num Lock key alternates between typing numbers and using the arrow keys.** Add blank lines as needed.
  - g. Save and exit the file.

## APPENDIX C

```
CREATE TABLE `Answers` (  
  `AnswerID` varchar(12) NOT NULL default "",  
  `Answer` tinytext NOT NULL,  
  `AnswerType` varchar(24) NOT NULL default "",  
  `NumOfOccurrences` int(11) NOT NULL default '0',  
  PRIMARY KEY (`AnswerID`)  
);
```

```
CREATE TABLE `AnswerType` (  
  `AnswerTypeID` varchar(12) NOT NULL default "",  
  `AnswerType` varchar(24) NOT NULL default "",  
  PRIMARY KEY (`AnswerTypeID`)  
);
```

```
CREATE TABLE `Categories` (  
  `CategoryID` varchar(12) NOT NULL default "",  
  `Term` varchar(255) NOT NULL default "",  
  `Description` longtext NOT NULL,  
  PRIMARY KEY (`CategoryID`, `Term`)  
);
```

```
CREATE TABLE `CategoryTerms` (  
  `CategoryID` varchar(12) NOT NULL default "",  
  `Category` varchar(255) NOT NULL default "",  
  PRIMARY KEY (`CategoryID`)  
);
```

```
CREATE TABLE `Questions` (  
  `QuestionID` int(11) NOT NULL default '0',  
  `Question` longtext NOT NULL,  
  `Length` int(11) NOT NULL default '0',  
  `AnswerID` text NOT NULL,  
  `NumOfOccurrences` int(11) NOT NULL default '0',  
  PRIMARY KEY (`QuestionID`)  
);
```

```
CREATE TABLE `Terms` (  
  `Term` varchar(255) NOT NULL default "",  
  `QuestionID` int(11) NOT NULL default '0',  
  `NumOfOccurrences` int(11) NOT NULL default '0',  
  PRIMARY KEY (`Term`, `QuestionID`)  
);
```

## **APPENDIX D**

### Interview Questions for Manual Shortcomings

1. Have you ever used a manual before?
2. What was the circumstance? Building furniture, troubleshooting software?
3. Do you often use manuals?
4. What are the shortcomings you have experienced with manuals?
5. Is the language easy to understand?
6. Is the font easy to read?
7. Are the diagrams realistic?
8. Are the diagrams helpful?
9. Are the diagrams easy to follow?
10. Were there sufficient diagrams?
11. Was it easy to identify the different parts?
12. Are the steps clearly indicated?

13. Is the order of steps clearly indicated?
14. Are necessary tools easily identified before actually required?
15. Are the tasks appropriately divided? Are you required to do too much per step?
16. How do you think the manuals could be improved? With respect to:
- a. Language
  - b. Presentation
  - c. Colors
  - d. Diagrams
17. Would a more interactive experience be more beneficial?

## APPENDIX E

### iTech Pre-Experiment Survey

Participant ID: \_\_\_\_\_

Age: \_\_\_\_\_

Gender: \_\_\_\_\_

Major: \_\_\_\_\_

**Race/Ethnicity:**

- ☐ Caucasian
- ☐ Hispanic
- ☐ African American
- ☐ Native American
- ☐ Pacific Islander
- ☐ Other: \_\_\_\_\_

**Citizenship:**

\_\_\_\_\_

**Highest Degree Obtained**

- ☐ High School
- ☐ B.S.
- ☐ B.A.
- ☐ M.S.
- ☐ M.A.
- ☐ Ph.D.
- ☐ Other: \_\_\_\_\_

**Zip Code of current residence**

\_\_\_\_\_

**Disabilities**

- ☐ Yes   ☐ No

Estimated annual income:

**Is English your native or second language?**

- ☐ Native language      ☐ Second language

**For approximately how many years have you been using a computer?**  
# of years

**Do you use a word processor, such as Microsoft Word or Word Perfect?**

- ☐ Yes  
☐ No

**If yes, how many documents have you created or updated?**

- ☐ 0 – 4      ☐ 5 – 8      ☐ 9 – 12      ☐ more than 12

**Have you ever used the vi editor before?**

- ☐ Yes      ☐ No

**If yes, how many documents have you created or updated?**

- ☐ 0 – 4      ☐ 5 – 8      ☐ 9 – 12      ☐ more than 12

**On average, how many times a week do you use a computer?**

- ☐ 0 – 1      ☐ 2 -3      ☐ 4 – 5      ☐ 6 or more

**Have you ever done any programming?**

- ☐ Yes      ☐ No

**If yes, what editor did you use?**

---

**In the section below, choose the response that most accurately describes you.**

**1. I frequently read computer magazines or other sources of information that describe new computer technology.**

- ☐ Strongly Agree      ☐ Agree      ☐ Neutral      ☐ Disagree      ☐ Strongly Disagree

**2. I know how to recover deleted or lost data on a computer or PC.**

- ☐ Strongly Agree      ☐ Agree      ☐ Neutral      ☐ Disagree      ☐ Strongly Disagree

**3. I know what a LAN is.**

- ☐ Strongly Agree      ☐ Agree      ☐ Neutral      ☐ Disagree      ☐ Strongly Disagree

**4. I know what an operating system is.**

- ☐ Strongly Agree      ☐ Agree      ☐ Neutral      ☐ Disagree      ☐ Strongly Disagree

**5. I know how to install software on a personal computer.**

- ☐ Strongly Agree      ☐ Agree      ☐ Neutral      ☐ Disagree      ☐ Strongly Disagree

**6. I know what a database is.**

- ☐ Strongly Agree      ☐ Agree      ☐ Neutral      ☐ Disagree      ☐ Strongly Disagree

**7. I am computer literate.**

- ☐ Strongly Agree      ☐ Agree      ☐ Neutral      ☐ Disagree      ☐ Strongly Disagree

**8. I am good with computers.**

- ☐ Strongly Agree      ☐ Agree      ☐ Neutral      ☐ Disagree      ☐ Strongly Disagree

**Submit**

## APPENDIX F

### iTech Post-Experiment Survey

Participant ID: \_\_\_\_\_

**Medium for technical communication used:**

☐ Book manual      ☐ Online manual      ☐ iTech

**Please respond by circling the reaction that best reflects your reaction to the technical communication medium:**

**Terrible.....Wonderful**

☐ 1    ☐ 2    ☐ 3    ☐ 4    ☐ 5

**Frustrating.....Satisfying**

☐ 1    ☐ 2    ☐ 3    ☐ 4    ☐ 5

**Dull.....Stimulating**

☐ 1    ☐ 2    ☐ 3    ☐ 4    ☐ 5

**Usable.....Not Usable**

☐ 1    ☐ 2    ☐ 3    ☐ 4    ☐ 5

**Boring.....Fun**

☐ 1    ☐ 2    ☐ 3    ☐ 4    ☐ 5

**Please respond by selecting the reaction that best reflects your impressions:**

**1. The medium was easy for me to use.**

☐ Strongly Agree    ☐ Agree    ☐ Neutral    ☐ Disagree    ☐ Strongly Disagree

**2. It was easy to get started.**

☐ Strongly Agree    ☐ Agree    ☐ Neutral    ☐ Disagree    ☐ Strongly Disagree

**3. It was easy retrieving an answer.**

☐ Strongly Agree    ☐ Agree    ☐ Neutral    ☐ Disagree    ☐ Strongly Disagree

**4. I knew what to say or do during a task.**

☐ Strongly Agree    ☐ Agree    ☐ Neutral    ☐ Disagree    ☐ Strongly Disagree

**5. I have a good understanding of how to edit documents on word processors.**



☐ Strongly Agree      ☐ Agree      ☐ Neutral      ☐ Disagree      ☐ Strongly Disagree

**6. If you had errors, it was hard to recover from them.**

☐ Strongly Agree      ☐ Agree      ☐ Neutral      ☐ Disagree      ☐ Strongly Disagree

**7. I was able to successfully complete the task.**

☐ Strongly Agree      ☐ Agree      ☐ Neutral      ☐ Disagree      ☐ Strongly Disagree

**8. I was intimidated by the medium used.**

☐ Strongly Agree      ☐ Agree      ☐ Neutral      ☐ Disagree      ☐ Strongly Disagree

**9. The medium I used, helped me to complete the task.**

☐ Strongly Agree      ☐ Agree      ☐ Neutral      ☐ Disagree      ☐ Strongly Disagree

**10. I would prefer having a technician present that I could personally ask questions to.**

☐ Strongly Agree      ☐ Agree      ☐ Neutral      ☐ Disagree      ☐ Strongly Disagree

**Additional comments or suggestions on the medium for technical communication used:**

---

---

---

---

**Submit**

## APPENDIX G

### iTech Post-Experiment Survey

Participant ID: \_\_\_\_\_

**Medium for technical communication used:**

- ☐ Book manual      ☐ Online manual      ☐ iTech

**Please respond by circling the reaction that best reflects your reaction to the technical communication medium:**

**Terrible.....Wonderful**

- ☐ 1    ☐ 2    ☐ 3    ☐ 4    ☐ 5

**Frustrating.....Satisfying**

- ☐ 1    ☐ 2    ☐ 3    ☐ 4    ☐ 5

**Dull.....Stimulating**

- ☐ 1    ☐ 2    ☐ 3    ☐ 4    ☐ 5

**Usable.....Not Usable**

- ☐ 1    ☐ 2    ☐ 3    ☐ 4    ☐ 5

**Boring.....Fun**

- ☐ 1    ☐ 2    ☐ 3    ☐ 4    ☐ 5

**Please respond by selecting the reaction that best reflects your impressions:**

**1. The medium was easy for me to use.**

- ☐ Strongly Agree    ☐ Agree    ☐ Neutral    ☐ Disagree    ☐ Strongly Disagree

**2. It was easy to get started.**

- ☐ Strongly Agree    ☐ Agree    ☐ Neutral    ☐ Disagree    ☐ Strongly Disagree

**3. It was easy retrieving an answer.**

- ☐ Strongly Agree    ☐ Agree    ☐ Neutral    ☐ Disagree    ☐ Strongly Disagree

**4. I knew what to say or do during a task.**

- ☐ Strongly Agree    ☐ Agree    ☐ Neutral    ☐ Disagree    ☐ Strongly Disagree

**5. I have a good understanding of how to edit documents on word processors.**

☐ Strongly Agree      ☐ Agree      ☐ Neutral      ☐ Disagree      ☐ Strongly Disagree

**6. If you had errors, it was hard to recover from them.**

☐ Strongly Agree      ☐ Agree      ☐ Neutral      ☐ Disagree      ☐ Strongly Disagree

**7. I was able to successfully complete the task.**

☐ Strongly Agree      ☐ Agree      ☐ Neutral      ☐ Disagree      ☐ Strongly Disagree

**8. I was intimidated by the medium used.**

☐ Strongly Agree      ☐ Agree      ☐ Neutral      ☐ Disagree      ☐ Strongly Disagree

**9. The medium I used, helped me to complete the task.**

☐ Strongly Agree      ☐ Agree      ☐ Neutral      ☐ Disagree      ☐ Strongly Disagree

**10. I knew what to say to iTech.**

☐ Strongly Agree      ☐ Agree      ☐ Neutral      ☐ Disagree      ☐ Strongly Disagree

**11. iTech was fast enough in response to my question.**

☐ Strongly Agree      ☐ Agree      ☐ Neutral      ☐ Disagree      ☐ Strongly Disagree

**12. iTech worked as I expected it to during the task.**

☐ Strongly Agree      ☐ Agree      ☐ Neutral      ☐ Disagree      ☐ Strongly Disagree

**13. I had problems understanding iTech.**

☐ Strongly Agree      ☐ Agree      ☐ Neutral      ☐ Disagree      ☐ Strongly Disagree

**14. iTech had problems understanding me.**

☐ Strongly Agree      ☐ Agree      ☐ Neutral      ☐ Disagree      ☐ Strongly Disagree

**15. I liked the appearance of iTech.**

☐ Strongly Agree      ☐ Agree      ☐ Neutral      ☐ Disagree      ☐ Strongly Disagree

**16. I would have preferred a female technician.**

☐ Strongly Agree      ☐ Agree      ☐ Neutral      ☐ Disagree      ☐ Strongly Disagree

**17. I was confident that iTech would be able to help me.**

☐ Strongly Agree      ☐ Agree      ☐ Neutral      ☐ Disagree      ☐ Strongly Disagree

**18. I would have preferred iTech having no face, just a voice.**

☐ Strongly Agree      ☐ Agree      ☐ Neutral      ☐ Disagree      ☐ Strongly Disagree

**19. I would use iTech again.**

☐ Strongly Agree      ☐ Agree      ☐ Neutral      ☐ Disagree      ☐ Strongly Disagree

**20. iTech would be easy to use by people who don't know a lot about computers.**

☐ Strongly Agree      ☐ Agree      ☐ Neutral      ☐ Disagree      ☐ Strongly Disagree

**21. If I had errors, it was hard to recover from them.**

☐ Strongly Agree      ☐ Agree      ☐ Neutral      ☐ Disagree      ☐ Strongly Disagree

**22. iTech gave me correct answers.**

☐ Strongly Agree      ☐ Agree      ☐ Neutral      ☐ Disagree      ☐ Strongly Disagree

**I would improve iTech by:**

---

---

---

**Additional comments or suggestions on the medium for technical communication used:**

---

---

---

---

**Submit**