Pergamon

Comput. Biol. Med. Vol. 24, No. 4, pp. 289–294, 1994
Copyright © 1994 Elsevier Science Ltd
Printed in Great Britain. All rights reserved
0010–4825/94 $7.00 + 0.00

0010-4825(94)00020-4

FEASIBILITY STUDY OF VOICE-DRIVEN DATA COLLECTION IN ANIMAL DRUG TOXICOLOGY STUDIES

MICHAEL A. GRASSO and CLARE T. GRASSO

Segue Corporation, 8265 Hammond Branch Way, Laurel, MD 20723, U.S.A.

(*Received* 23 *March* 1994; *received for publication* 21 *July* 1994)

Abstract—The object of this study was to determine the feasibility of using voice recognition technology to enable hands-free and eyes-free collection of data related to animal drug toxicology studies. Specifically, we developed and tested a prototype voice-driven data collection system for histopathology data using only voice input and computer-generated voice responses. The overall accuracy rate was 97%. Additional work is needed to minimize training requirements and improve audible feedback. We conclude that this architecture could be considered a viable alternative for data collection in animal drug toxicology studies with reasonable recognition accuracy.

Speech recognition   Voice recognition   Data collection
Animal Drug   Toxicology

INTRODUCTION

Data entry has become the bottleneck of many scientific applications designed to collect and manage information related to experimental studies. In animal drug toxicology studies, this is true because of the need to collect data in hands-busy or eyes-busy environments. For example, during microscopy, the operator's hands and eyes are occupied with the process of examining tissue slides. During necropsy, gross observations and organ weights must be collected while the operator's hands are busy. With in-life data collection, technicians must record daily observations while handling animals. An ancillary issue with in-life data collection is that it may not be practical to keep computer equipment in animal rooms, where it is most convenient to record observations. In these areas, there is a definite need to develop tools for computer-automated data collection which do not rely on a keyboard for input or a monitor for output.

Large volumes of pathology data are processed during animal drug toxicology studies. These studies are used to evaluate the long-term, low-dose effects of potentially toxic substances, including carcinogens. This information must be collected, managed, and analyzed according to Good Laboratory Practice regulations for animal studies [1]. Since the 1970s, several systems have been developed to automate this process [2, 3], and procedures for manual data entry have been set up. Others included interfaces to clinical chemistry and hematology analyzers to automate data collection [4]. Today, however, the collection of microscopic, gross, and in-life observations is still a limiting factor, owing to hands-busy and eyes-busy restrictions.

In this paper we describe a prototype voice-driven data collection system for the on-line collection of microscopic pathology data from animal drug toxicology studies. The system was developed to facilitate the collection of histopathology data using only voice input and computer-generated speech responses. After testing the prototype system we evaluated the results to determine the feasibility of this project and provide a basis for implementing voice-driven systems that support microscopic, gross, and in-life data collection.

CURRENT SPEECH RECOGNITION TECHNOLOGY

Speech recognition systems provide computers with the ability to identify spoken words and phrases. Research into automated speech recognition began in the 1950s. However, only recently have advances in computer architectures and computational throughput resulted in a number of successful commercial voice recognition systems. These systems can be categorized by speaker dependence, continuity, and vocabulary size.

Speaker-independent systems can recognize speech from any speaker. Speaker-dependent systems must be trained by each individual user, but typically have higher accuracy rates. Speaker-adaptive systems, a hybrid approach, start with speaker-independent templates and adapt them to specific users over time without explicit training. Continuous speech systems can recognize words spoken in a natural rhythm while isolated word systems require a deliberate pause between each word. Although more desirable, continuous speech is harder to process, because of the difficulty in detecting word boundaries. Vocabulary size can vary anywhere from 20 words to more than 40,000 words. Large vocabularies cause difficulties in maintaining accuracy, but small vocabularies can impose unwanted restrictions on the naturalness of communication. A more thorough review of this subject can be found elsewhere [5, 6].

Several voice recognition computer systems have been developed in the medical area. Voice-activated dictation has been applied to radiology, emergency medicine, and pathology [7–9]. These are large vocabulary, isolated word, speaker adaptive systems geared toward the generation of reports using fill-in forms, trigger phrases, and free-form speech. Other systems enable data collection in hands-busy situations, such as periodontal charting [10] and anesthesia record keeping [11]. Speech is being researched as an improved interface for medical expert systems [12]. Another effort used voice input in combination with a digitizing tablet to collect stereological data [13].

Although voice recognition technology is becoming more widely used, there are two important caveats. First, successful voice application development requires an adequate understanding of the technology's limitations. True natural language processing by computers is still several years away and there is little conclusive evidence that speech is superior to the keyboard or other input devices. It is important therefore to apply voice recognition where there is an additional motivation to offset its limitations. This includes the use of moderately sized, well-defined vocabularies in hands-busy and eyes-busy situations, such as during microscopy and necropsy. Second, voice input is an entirely new modality that should be appraised without the bias of other input paradigms. Simply adding voice to an existing user interface can decrease system integrity or create integration discontinuity [14]. Therefore, it is best to design voice-driven interfaces from scratch, to enable examination of user interaction from this new perspective.

MATERIALS AND METHODS

The hardware for this study consisted of an IBM-compatible 486/33 computer with a 14-inch Super VGA monitor, 8 Mb of memory, 200 Mb hard disk, a mouse, Microsoft Windows 3.1, and Microsoft DOS 6.0. Software was developed under Microsoft Windows using Borland C++3.1 and the Borland Object Windows Library 1.0 (Borland International, Inc., Scotts Valley, CA, U.S.A.).

The Verbex 6000 AT31 Model 0637 Voice Input Module with 3 Mb of memory, 40 MHz processor, and text-to-speech synthesis was used for voice recognition and computer-generated voice responses (Verbex Voice Systems, Inc., Edison, NJ, U.S.A.). Although several voice recognizers were considered, we selected the one by Verbex for a number of reasons. The unit is contained on a single full-sized AT bus board, making it compatible with the ISA 16-bit bus of our 486/33 computer. It comes with an application programming interface for the development of Microsoft DOS and Windows applications using C and C++. The recognizer's memory and co-processor can be upgraded to handle vocabularies from 300 to over 10 000 words, depending on configuration. Finally, the unit can recognize continuous speech with accuracy rates above 98%.

We chose Watcom SQL for Windows 3.1 (Watcom International Corporation, Waterloo, Ontario, Canada) over other more established relational databases because of its low resource demands, client/server scalability, and support of the ANSI SQL89 standard. The resulting environment was therefore able to run on a single windows workstation with as little as 4 Mb of memory and also support multiple users in a networked configuration. Adoption of the ANSI SQL89 standard will make porting the system to other databases relatively easy.

Two separate interfaces were developed for data collection. One used the keyboard, mouse and computer monitor with standard Windows techniques such as dialog boxes, push buttons, and pulldown menus. The other used only voice input and computer-generated voice responses with no visual feedback. Note that these were two distinct user interfaces. We did not simply add voice-driven capabilities to the Windows user interface.

Grammar development

Grammar definitions are used by the Verbex Voice Input Module to identify the vocabulary of possible word utterances, the rules that determine which words or word orders are valid in various contexts, and the responses the system will take when a word or phrase is recognized. Our grammar contained around 900 words, mostly due to the large number of nomenclature terms. However, grammar rules were used in such a way that no more than 100 words, and most often less than 15, were valid at any time. We divided the list of possible words into functional subsets for navigation, voice response, error correction, nomenclature terms, and data collection.

(1) *Navigation.* FIRST, LAST, NEXT, PREVIOUS, CLOSE. These commands are used to navigate through studies and the animals assigned to each study. For example, FIRST STUDY, LAST ANIMAL opens the last animal in the first study. Alternatively, ANIMAL 5 opens the fifth animal in the current study. CLOSE STUDY closes the current study.

(2) *Voice response.* LIST, REPEAT, QUIET. The LIST command is used to list valid studies, animals or nomenclature terms using one of three forms. For example, LIST STUDY lists all studies; LIST ANIMAL 1 DASH 10 lists the first ten animals on the current study; LIST ORGAN 5 lists the fifth organ in the nomenclature. REPEAT restates the last computer response. QUIET is used to end prematurely a voice response.

(3) *Error correction.* HELP INFORMATION, CANCEL, STATUS. HELP INFORMATION initiates a voice response containing context sensitive help information. CANCEL aborts the current data collection transaction. STATUS uses voice responses to keep track of information previously entered.

(4) *Nomenclature terms.* The nomenclature was adapted from the Pathology Code Table which was designed for use by the National Toxicology Program [15]. It contains terms for both microscopic and gross observations that are organized into topographies, morphologies, and qualifiers. The topographies are comprised of a hierarchy of systems, organs, and sites. Qualifiers are organized into groups. Sites and morphologies may be associated with a specific organ observation. Qualifiers may be used in conjunction with a morphology or a gross site, allowing further description of the abnormality.

(5) *Data collection.* OBSERVE, ORGAN, SITE, MORPHOLOGY, QUALIFIER. During data collection, the user can make microscopic observations about organs in the current animal and study. To observe the liver, the user would say OBSERVE LIVER. From here a site, morphology, and qualifier can be added to the current observation using phrases like SITE LEFT LATERAL LOBE, MORPHOLOGY INFLAMMATION, and SEVERITY MARKED.

Testing

In each test, the subject was either a pathology assistant, medical technologist or software engineer. The first test was to train the system to recognize each user's voice by reading each word twice, followed by reading representative words in context. Owing to time restrictions, some users trained on only a subset of the vocabulary consisting of the more common nomenclature terms.

The second test was used to validate the accuracy of the voice recognition system apart from our application. Each user was asked to read series of 100 randomly generated phrases. The number of correctly recognized phrases was used to compute the recognition accuracy. If a phrase was accidently read incorrectly, it was not counted as an error, and the user was given a second chance to read the phrase again.

In the next test, each user was asked to navigate to various animals and enter several microscopic observations. Here, voice recognition was not used. Instead, the keyboad and mouse were used for input and a computer monitor for visual responses. This test was to allow the users to familiarize themselves with the environment and provide a comparison for data entry using voice input.

The final test required each user again to navigate to various animals and enter several microscopic observations. This time, however, the mouse, keyboard, and monitor could not be used. Instead, each user relied on voice input and computer-generated voice output.

RESULTS

Overhead associated with training was a limiting factor. Roughly, 6–8 hr were required for each user to train on the entire vocabulary of 900 terms. Training time tended to increase for those who never used a voice recognition system before. Invariably, users needed to repeat the training for specific words that the system was not recognizing consistently. The mean recognition rate was 97% in the accuracy test, which was similar to that proposed by the manufacturer of the recognizer hardware.

In the last test, most participants felt uncomfortable at first when entering observations without any visual feedback. This was due in part to difficulty in understanding computer-generated speech. After a few practice runs, they were entering data without assistance. However, many felt the system should provide more feedback during data collection, be it visual or audible. The mean recognition rate in this test was also 97%.

DISCUSSION

Our main goal was to test the feasibility of developing voice-driven software that supports the collection of data in animal toxicology studies. To meet this goal, we developed a prototype voice-driven system that supported histopathology data collection, incorporated a grammar based on the pathology code table, and tested it with several users in a controlled environment.

Grammar development proved to be the most difficult task. Initially, we created a set of subgrammars that exceeded the capacity of the equipment we were using. We resolved this problem by organizing the nomenclature in a simpler format, and, thus, were able to work within the parameters of the equipment, but at the expense of naturalness of communication.

For reasonable response time, there was a practical upper limit of no more than 100 candidate words possible at one time. Note that this limit was only on the first word in each phrase. Using multiple word combinations the system could in fact recognize more than 100 phrases and still have acceptable performance. Also, no more than 20 subgrammars could be active, which limited the number of branches a grammar could have.

Another limitation was that the grammar rules were static and therefore had to be anticipated beforehand. This required us to store nomenclature terms both in the database for data validation and in the grammar file for voice recognition. This redundancy could have been eliminated if grammar rules could dynamically be created and downloaded to the recognizer in real-time.

The initial training requirements are a potential hindrance to the acceptance of a system of this type. In a time when few people, if any, read the user's guide, it is difficult to envision a pathologist or technician spending hours training the system to recognize his or her voice. An alternative that warrants further study is a speaker adaptive approach. Here, instead of training the system, operators would use a set of generic voice recognition templates, which would automatically be adapted for each person with continued use.

Another interesting observation has to do with word conflicts in the vocabulary. Such conflicts can occur with short, similar sounding words like "tree" and "three". With a vocabulary of complex medical terms, we thought we would be immune to such problems. However, there were a few conflicts with phrases like "inferior vena cava" and "superior vena cava".

The area of computer feedback requires additional research. Since the system operated in an eyes-busy environment, there could be no visual computer feedback. We anticipated several areas where audible confirmation would be appropriate, such as when a word was recognized by the system or when an observation was saved in the database. However, occasionally there were moments of "dead air time" when the computer was involved in a large database transaction or the recognizer was in a complex recognition event. Here, additional audible feedback is necessary so the user knows the computer is busy, similar to displaying an hourglass when a Windows program is busy. This is not always easy to do. For instance, a software application can only determine when a recognizer event ended, not when it began, which makes it difficult to know when to transmit a busy signal.

As testing of the prototype progressed, we began to feel that allowing no visual feedback was too restrictive. Audible feedback is at least 10 times slower than reading, which limits the amount of information that can be giving to the user [16]. Ideally, most of the time data entry would progress with audible feedback alone. There will, however, be times when the user would be better served by looking up at a monitor to evaluate the state of the system. This is especially true during error detection and resolution.

SUMMARY

In summary, we developed and tested a prototype voice-driven data collection system for histopathology data related to animal drug toxicology studies using speaker-dependent, continuous speech recognition technology. Recognition accuracy was satisfactory, as was the ability of operators to enter data using only voice input and computer-generated voice responses. Limitations that should be addressed in the future include minimizing voice training requirements and improving audible feedback. Our study demonstrates that this architecture could be considered a viable alternative for hands-free and eyes-free data collection in animal drug toxicology studies with reasonable accuracy.

Acknowledgement—Supported by SBIR grant #1R43RR07978-01 from the National Center for Research Resources.

REFERENCES

1. U.S. Food and Drug Administration, Good laboratory practice regulations for non-clinical laboratory studies, *Fed. Regulat.* **43**, 60015–60019 (1978).
2. M. F. Cranmer, L. R. Lawrence, A. J. Konvicka and S. S. Herrick, NCTR computer systems designed for toxicologic experimentation. I. Overview, *J. environ. Path. Toxic.* **1**, 701–709 (1978).
3. J. M. Faccini and D. Naylor, Computer analysis and integration of animal pathology data, *Arch. Toxic.* **2** (Suppl.), 517–520 (1979).
4. A. M. Daly, R. A. Martin, E. J. McGuire and C. J. DiFonzo, A microcomputer-based data acquisition and reporting system for clinical pathology data from animal drug toxicology studies, *Drug Inf. J.* **23**, 285–296 (1989).
5. B. Bergeron and S. Locke, Speech recognition as a user interface, *M. D. Comput.* **7**, 329–334 (1990).
6. R. D. Peacocke and D. H. Graf, An introduction to speech and speaker recognition, *Computer* **23**, 26–33 (1990).
7. R. A. Reed, Voice recognition for the radiology market, *Topics Hlth Rec. Management* **12**, 58–63 (1992).
8. J. A. Hollbrook, Generating medical documentation through voice input: the emergency room, *Topics Hlth Rec. Management* **12**, 49–57 (1992).
9. E. C. Klatt, Voice-activated dictation for autopsy pathology, *Comput. Biol. Med.* **31**, 429–433 (1991).

10. C. A. Feldman and D. Stevens, Pilot study on the feasibility of a computerized speech recognition charting system, *Commun. Dentist. Oral Epidemiol.* **18**, 213–215 (1990).
11. N. T. Smith, R. A. Brian, D. C. Pettus, B. R. Jones, M. L. Quinn and A. Sarnat, Recognition accuracy with a voice-recognition system designed for anesthesia record keeping, *J. clin. Monitor.* **6**, 299–306 (1990).
12. K. Johnson, A. Poon, S. Shiffman, R. Lin and L. M. Fagan, A history-taking system that uses continuous speech recognition, *Proc. 16th A. Symp. Comput. Applic. Med. Care*, November (1992).
13. P. J. McMillan and J. G. Harris, Datavoice: a microcomputer-based general purpose voice-controlled data-collection system, *Comput. Biol. Med.* **20**, 415–419, November (1990).
14. C. E. Wulfman, E. A. Isaacs, B. L. Webber and L. M. Fagan, Integration discontinuity: interface users and systems. Tech. Report KSL-88-12, Knowledge Systems Laboratory, Stanford University, Palo Alto, CA (1988).
15. National Center for Toxicological Research, Post experiment information system pathology code table reference manual, TDMS Document #1118-PCT-4.0, Jefferson, Arkansas (1985).
16. N. C. Maberly, *Mastering Speed Reading*. New American Library, New York (1966).

About the Author—MICHAEL A. GRASSO is the President of Segue Corporation. His major interests include medical informatics, database systems, and software engineering. He received a B.S. degree in Microbiology from the University of Maryland in 1983, an M.S. degree in Computer Science from the American University in 1986, and is currently completing his Ph.D. in Computer Science at the University of Maryland Baltimore County.

About the Author—CLARE T. GRASSO, Vice President of Segue Corporation, graduated Summa Cum Laude from Towson State University with a B.S. in Applied Mathematics in 1984. She specializes in systems programming, instrumentation interfaces, and information systems. Before co-founding Segue, she worked for the National Institute of Standards and Technology and NASA.