The Utility and Cost Effectiveness of Voice Recognition Technology in Surgical Pathology

Walter H. Henricks, M.D., Kavous Roumina, Ph.D., Bradley E. Skilton, Debra J. Ozan, Gwendolyn R. Goss

Department of Anatomic Pathology, The Cleveland Clinic Foundation, Cleveland, Ohio

Voice recognition (VR) technology in computer systems converts speech directly into electronic text. In pathology, VR holds promise to improve efficiency and to reduce transcription delays and costs. We investigated the utility and cost effectiveness of targeted VR deployment in surgical pathology. A VR system was deployed for entry of gross descriptions of biopsies and of low to moderate complexity specimens and for entry of final reports for specimens not requiring microscopic analysis. Templates for VR were developed for all reports. Free-text speech entry was used to enter information not covered by templates. Voice converted to text by VR crossed over an interface into the anatomic pathology laboratory information system. Tallies were kept of whether individual specimens were entered by VR or by conventional dictation. A computer program was written to analyze the number of lines of text entered through VR. Cost savings were calculated based on per-line transcription costs from an outside agency. Over 18 months, gross descriptions for an average of 5617 specimens per month were entered via VR, corresponding to 70% of all gross specimens processed by the laboratory. A mean of 106 gross-only final reports per month was entered through VR. VR facilitated same-day processing of specimens received after the previous day processing cutoff time (average 35 specimens per day). VR generated an average of 23,864 lines of text per month, translating to $2625 savings per month. Estimated payback period for VRT as implemented is 1.9 years. The use of VR for gross descriptions of biopsies and low to moderate complexity specimens and for gross-only final reports in surgical pathology facilitates data entry, reduces transcription costs, and contributes to improved turnaround time. Development of templates is important to successful implementation of VR in surgical pathology.

KEY WORDS: Computers, Informatics, Surgical pathology, Technology, Templates, Voice recognition.

Mod Pathol 2002;15(5):565–571

Pathology reports consist of observations and interpretations recorded as words and numbers in electronic and paper form. A transcription step is typically necessary for dictated observations and diagnoses to be typed as text into an anatomic pathology laboratory information system (APLIS), from which reports are produced. Voice recognition (VR) technology in computer systems offers the capability to convert human speech directly into electronic text.

In VR systems, a microphone converts human speech into an analog electrical signal that an electronic circuit board within a computer then converts to a digital signal (1–3). Speech recognition engine software then uses acoustic, language, and vocabulary models as well as complex statistical algorithms to transform the digital signal into words and punctuation marks. Language and vocabulary models specific to pathology have been developed and are available in commercially available VR systems. Such pathology-specific language models and vocabulary models improve the accuracy of word recognition and word prediction in the context of the language used pathology reports. Earlier generations of VR systems were termed "discrete" because the speaker had to separate each word by a short pause. VR systems now available allow "continuous" voice recognition in which a speaker may speak more naturally without pausing between each word.

In surgical pathology, voice recognition (VR) technology holds promise to improve efficiency of workflow and to reduce transcription delays and costs. Dictation in pathology that a VR system converts directly into electronic text does not require transcription into the APLIS if the VR system is either interfaced to or integrated as part of the

APLIS. Others (4, 5) have previously described implementation of VR across all stages of producing pathology reports, including gross descriptions, microscopic descriptions, and final diagnoses and comments. Using VR technology for microscopic descriptions and final diagnoses and comments, however, raises potential barriers to successful implementation that include the potential for use of VR technology to transfer greater clerical responsibility to pathologists (5). Also, gross descriptions are more amenable to the use of standardized templates, and the greater amount of free-form text (or "free-text") entry generally required in final diagnoses and comments compared with gross descriptions increases chances for inaccurate speech recognition when using VR for final diagnoses. We investigated the utility and cost effectiveness of deployment of a commercially available VR system that was targeted specifically in the grossing area in surgical pathology.

MATERIALS AND METHODS

A continuous voice recognition (VR) system (Clinical Reporter for Pathology Version 4.02, Lernout & Hauspie, Burlington, MA) was deployed in the grossing room in the Department of Anatomic Pathology at The Cleveland Clinic Foundation and assessed from January 2000 through June 2001. The VR system hardware consisted of a server computer (Main Board, Inc., Waltham, MA; Pentium Pro 200 megahertz [MHz] processor unit, 64 megabyte [MB] random access memory [RAM], and Windows NT Version 4.0 operating system; Microsoft Corp., Redmond, WA) and three personal computer (PC) workstations (Pentium MMX 333 MHz, 128 MB RAM, Windows NT Version 4.0; Main Board, Inc.; also, a Montego II sound card; Turtle Beach, Yonkers, NY) connected to the departmental Ethernet local area network. The server stored users' voice profiles, gross description templates, and text files of pathology reports generated by VR. The workstations were installed at three locations at which pathologist assistants (PAs) and surgical pathology technicians (SPTs) performed gross examinations of biopsies, low- to moderate-complexity specimens, and gross-only specimens not requiring microscopic analysis. Individual templates were developed for gross descriptions of all specimen types to be processed through VR data entry (Table 1). These templates were tailored to each specimen type and consisted of skeletons of fixed text with prompts for the user to speak descriptive phrases and words from predefined fill-ins ("trigger words") as well as prompts to dictate numerical measurements. Users could dictate additional free-text descriptions into all reports. An electronic interface enabled data from the VR system to transfer automatically into the appropriate fields within the AP Laboratory Information System (CoPath version 5.3, Dynamic Healthcare Technologies Incorporated, Waltham, MA). The usual voice dictation system was kept in place for larger and higher complexity specimens that were not processed using the VR system. A training station separate from the work area was established and was used for initial user enrollment and training.

Briefly, the process of generating gross reports using the VR system worked as follows (Fig. 1): Before using the VR system, each user (PA or SPT) underwent a 2–3 hour training session that included creation of an individual voice profile based on the characteristics of the individual's speech. To create reports using the VR system, a user first logged in to the VR application at a workstation, and his or her voice profile was downloaded from the server over the network to the PC on which he/she was working. Using voice commands, the user first called up the specimen accession number, and the demographic information for that particular case was then retrieved from the APLIS and presented on screen. The user then called up the template specific for the specimen type on which he or she was working. The user then dictated the appropriate information at the prompts within the template and added free text description as necessary (Fig. 2). Proofreading and editing occurred at the time of data entry into the VR system, as the user determined whether the VR system entered the correct data as the data appeared on the screen. When the description was completed and after voice activation of an electronic signature within the VR system, the gross description information

TABLE 1. Specimen Types for Which Templates Were Prepared for Voice Recognition–Based Entry of Gross Descriptions

Adenoids	Foreign Body	Products of conception
Aorta–aneurysm	Gallbladder	Prostate–transurethral resection
Appendix	Hemorrhoid	Prosthesis
Artery–plaque	Hernia sac	Teeth
Biopsies–all sites	Knee–arthroplasty	Tonsil
Breast tissue–reduction mammoplasty	Lipoma	Uterine cervix–LEEP excision, cone biopsy
Dermatological–biopsies and excisions	Lymph nodes–regional dissection	Uterus–leiomyoma, prolapse
Femoral head–arthroplasty	Placenta	Vas deferens
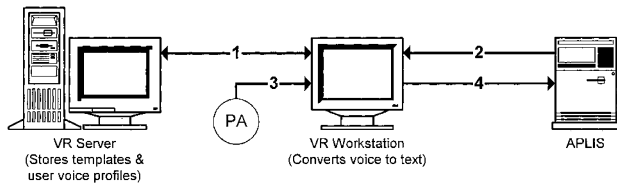
FIGURE 1. Steps in generating reports through voice recognition depicted in this figure: (1) user logs in, and user profile and updated templates are downloaded from server; (2) patient demographics and case accession information are transferred on command from APLIS; (3) report is dictated by user (PA) into VR system; (4) gross description is populated in APLIS after voice-activated release from VR system. VR = voice recognition; APLIS = anatomic pathology laboratory information system; PA = pathologist assistant.

crossed over the interface from the VR system into the APLIS, where it was immediately available.

Tallies were kept of the number of specimens for which descriptions and gross-only final reports were entered through the VR system and for those processed using the conventional dictation method. A computer program (Visual Basic version 6.0, Microsoft Corporation, Redmond, WA) was written that used information on the VR server to analyze the text that was entered through the VR system. The capital costs are summarized in Table 2. The payback period on capital investment was calculated using per-line cost from the regular transcription agency that was employed by the department.

RESULTS

Over the 18-month period, gross descriptions for an average of 3864 accessioned cases consisting of 5617 individual specimens per month were entered using the VR system. A mean of 106 gross-only final reports per month was entered through the VR system. Gross descriptions for 98% of all biopsies and low-complexity specimens, including >99% of all biopsies, were generated using the VR system. For specimens entered using VR, >99% of the data entered were part of a template, and <1% were free-text. Overall, 70% of all gross specimens (individual parts) processed in the laboratory had gross descriptions generated through the use of the VR system, and 30% of gross descriptions (those for higher complexity specimens) were generated using conventional transcription. The immediate availability of gross descriptions in the APLIS after entry through the VR system facilitated same-day processing of average of 35 specimens per day that were received after the previous day's processing cutoff time. The amount of error correction required varied depending on speaker and/or exact words. Biopsy reports in which the voice entries were mostly numbers (for measurements) generally required less correction than did low-complexity specimen reports, for which more trigger words were used. Recognition accuracy for words, numbers, and triggers ranged generally from 70–90%.

The VR system generated an average of 23,864 lines of text per month, translating into an average cost saving of $2625 per month in transcription costs. Estimated capital payback period for the VR system as implemented is 1.9 years.

Approximately 120 hours of computer system analyst and PA time were required to develop and to enter templates into the VR system, and approximately 40 hours were spent testing the system and templates. Establishing a new voice profile for a new user required 2–3 hours per user. Loading an individual's voice profile from the server required 30–60 seconds at the time of user log-in. Approximately 4 weeks, on average, were required for users to become proficient with the VR system.

Minor adjustments to the equipment and work environment were necessary. Overall, background noise created little interference with speech recognition. One workstation was moved from an area close to louder noises to a quieter area in the laboratory. Moving PC units off workbenches to shelves conserved bench space. The installation of flat-panel LCD (liquid crystal display) computer monitors conserved bench space and enabled monitor placement to be at users' eye level (Fig. 3). Wireless headset microphones replaced initial units that required wires to connect to the system and facilitated user mobility necessary to perform other tasks.

DISCUSSION

Our study reports the utility and cost-effectiveness of the use of a voice recognition (VR) system for gross description reports and gross-only final reports in surgical pathology. VR systems convert speech into text in computer systems. VR software can now achieve acceptable or better performance and is readily available as a consumer product. In medicine, the use of VR technology is being actively pursued in a number of fields, including radiology, where the use of VR to generate reports is becoming more commonplace (6).

The benefits of using VR technology to automate the entire process of producing a pathology report have been previously reported (4, 5). Teplitz and colleagues (4) deployed a department-wide system that involved residents and attending pathologists and used VR for gross descriptions, microscopic descriptions, and final diagnoses and comments. We investigated the utility of a less universal, targeted deployment of VR for generating gross descriptions of biopsies and low to moderate complexity specimens that pathologist assistants (PAs) handle. The appeal of this approach compared with
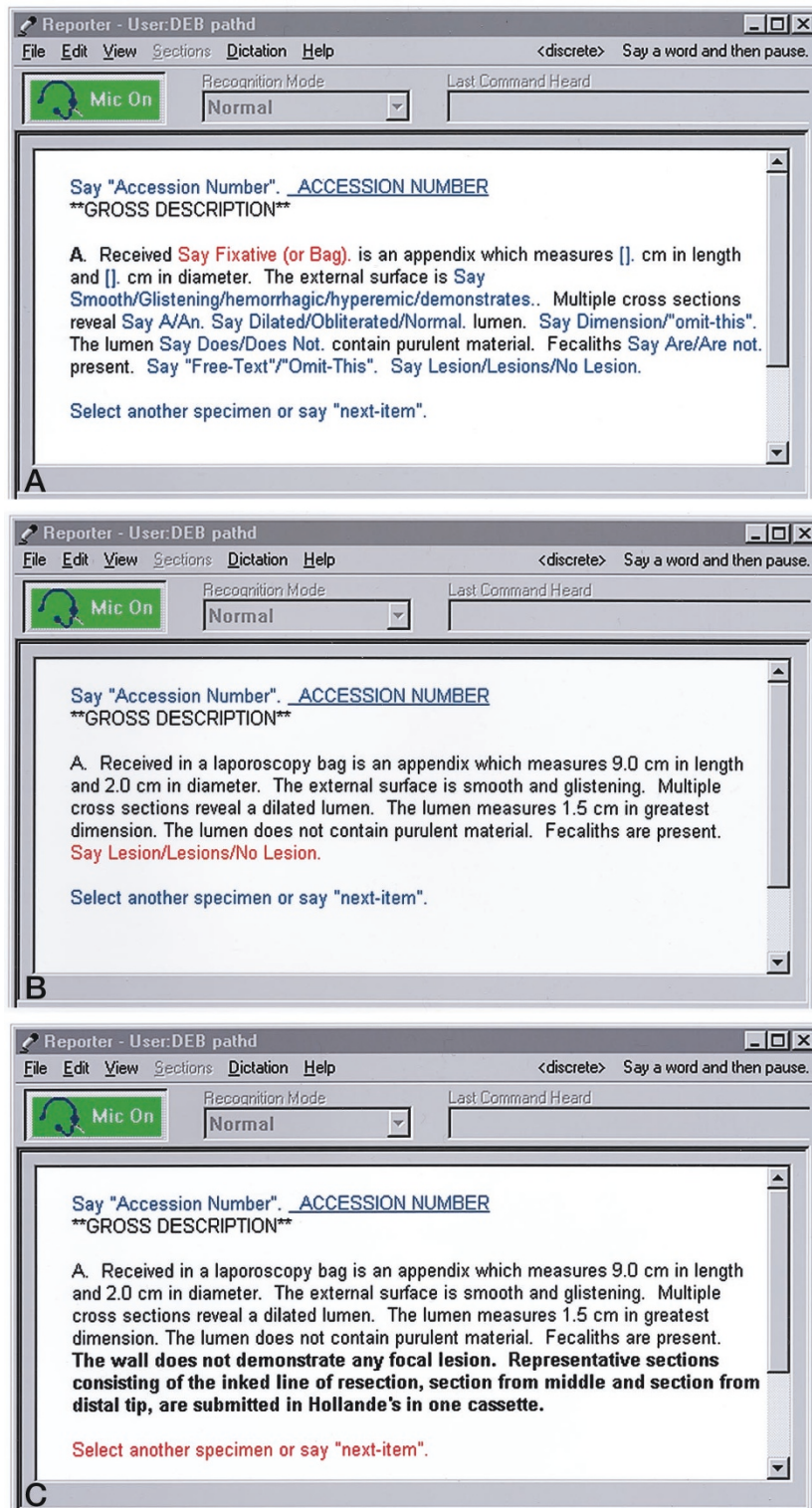
FIGURE 2. Screen captures from the voice recognition system illustrate report generation and use of templates and trigger phrases. **A,** a template (for appendix in this example) consists of a skeleton of fixed text (**black text**) with prompts (**blue text)** to speak trigger words and phrases and prompts to dictate numerical measurements (**blue brackets**). The current active prompt—"Say Fixative (or Bag)" is displayed in **red. B**, This screen illustrates the appearance after completion of the template up to the paragraph's final prompt, which is now displayed in **red. C,** after the user speaks the trigger "no lesion," the bold text displayed is inserted. The report now represents the completed gross description. The system prompts the user for the next step in the process (**red text**).

implementing a department-wide VR system for the entire process is that (1) less capital investment is required; (2) fewer users are involved; (3) gross descriptions for biopsies and low- to moderate-complexity specimens are amenable to standardization and templates, whereas microscopic de-

TABLE 2. Capital Acquisition Costs for Targeted Voice Recognition Technology (VRT) Implementation in Surgical Pathology

Item	$ Cost
Hardware and equipment	22,900
VRT software licenses	13,000
Interfaces between VRT and laboratory information system	11,000
Implementation/training	9,600
Maintenance (annual)	3,250
Total	59,750

FIGURE 3. Pathologist assistant at voice recognition grossing station with space-saving arrangement as described in the text. The flat-panel monitor (*right*) also allows the monitor to be at eye level. The wireless headset microphone permits mobility without having to remove the headset. The receiver for the wireless microphone sits atop the PC unit (*upper right*).

scriptions and final diagnoses and comments tend to require more free-text (non-templated) verbiage. Before our study, however, it was unclear whether undertaking a project that limited the use of VR to only a subset of specimens and to generating only the gross description portion of reports would be worthwhile. Indeed, much of the benefit reported for VR in pathology has been attributed to pathologists generating the final reports through the VR system. Our data and experience indicate that targeted deployment of VR for gross descriptions of biopsies and low- to moderate-complexity specimens, as well as for gross-only final reports, facilitates data entry, reduces transcription cost, and contributes to improved turnaround time in surgical pathology.

In our laboratory, VR has been successfully integrated into the workflow of surgical pathology. Overall, 70% of all gross specimens had gross descriptions generated through the use of the VR system, and 30% of gross descriptions (for higher complexity specimens) were generated using conventional dictation. Gross descriptions for 98% of all biopsies and low complexity specimens (Table 1), including >99% of all biopsies, were generated using the VR system. High, or "overflow", specimen volumes on certain days necessitated the use of conventional dictation and transcription for a minority (approximately 2%) of low-complexity specimens, reflecting the greater availability of dictation devices in the laboratory (compared with the number of VR workstations). More than 99% of the data for VR-processed specimens were part of a template, with <1% free-text. Because the implementation focused on biopsies and low complexity specimens with largely uniform characteristics, the need for free-text data entry was minimal and necessary only for unusual findings not covered in the choices for voice prompts. After 18 months of experience, conventional transcription of gross reports for these specimens has been virtually eliminated, because the information is sent automatically to the APLIS from the VR system. Transcription cost savings based on per-line costs from our outside agency has been $2625 per month. We estimate that our experience would correspond to a saving of at least one full-time equivalent position if these reports were typed in-house. The payback period on the capital investment is <2 years.

The use of VR as implemented has contributed to improved turnaround time. Because of the geographically distributed nature of our institution, biopsies often arrive late on the day on which they have been procured. These biopsies are processed early the following morning and can be signed out that same date. Entry of gross descriptions for these specimens through the VR system facilitates the process in that the gross descriptions are immediately available in the APLIS after dictation instead of the typical 2- to 4-hour turnaround of dictation by the transcription agency. Additionally, the transcriptionists at that time are typing gross descriptions on larger specimens (those not entered through the VR system) and final reports on all specimens, so the use of the VR systems for these biopsies reduces the concurrent workload affecting the transcriptionists. In essence, the use of the VR system improves the ability to do more work in parallel without a transcription bottleneck.

Loading an individual's voice profile from the server required 30–60 seconds at the time of user log-in. This amount of time did not pose a problem because the same user worked at a workstation for a relatively long period. Also, the voice profile need only be loaded from the server at the beginning of each log-in session and not with each case.

Key to successful VR implementation has been the use of the templates. Although a controversial topic, the potential benefits of standardized reports have been described by others (5–7), and many of these benefits hold for gross descriptions. Templates can ensure completeness of descriptions, standardize measurements and reports, and streamline and focus descriptions while eliminating excess verbiage. Gross descriptions of small specimens are especially amenable to standard descriptions coupled with trigger phrases (*i.e.*, text strings filled in based on a single trigger word) and numerical fill-ins. Not only does the use of templates in this manner (hard-coded text and trigger words) reduce the length of time it takes to create a report, but also hard-coded text within the template is not spoken by the users and thus does not require software recognition. Furthermore, trigger phrases consisting of multiple words require recognition only of the trigger word. In these ways, templates and trigger words reduce the impact of inaccurate speech recognition by the VR system. In our experience, development and testing of templates required considerable time, approximately 160 hours of system analyst and PA effort.

Several environmental and ergonomics issues required resolution for a successful implementation. First, the use of a VR system requires the addition of additional equipment, including a PC unit, computer monitor, and keyboard, to each individual grossing station. The equipment had to be incorporated into a limited amount of available space at each site. Bench space-saving solutions included moving the PC unit off of the workbench and onto a shelf (the floor was also considered), attaching a retractable keyboard drawer to the edge of the bench surface, and installing flat-panel LCD (liquid crystal display) monitors that have a narrower profiles than the original CRT (cathode ray tube) monitors.

Second, the necessity for the user to look at the monitor while dictating reports meant that the location of the monitor relative to the user's line of sight required consideration. The small footprint of the flat-panel LCD monitors allowed placement directly in front of the users on the workbench and eliminated the need for constant back-and-forth or up-and-down head movements (Fig. 3). The use of adjustable-height chairs also helped ensure proper alignment.

Third, those individuals performing gross examinations in our laboratory often must leave the workbench to perform multiple other tasks. Providing wireless headset microphones greatly facilitated user mobility by eliminating the tethering wire and also eliminated the need to disconnect and reconnect each time a user left the grossing station. Another advantage of wireless microphone is that there is no wire present to interfere with specimen manipulation or to come in contact with blood, body fluids, or fixative.

The optimal placement of VR grossing stations within the laboratory required consideration of several factors, the most important of which was the potential level of background noise. Generally, background noise did not have an appreciably negative effect on speech recognition. Only when a station was placed in an area closer to intermittent, louder noises, for instance, persons announcing "frozen section", did the system's recognition suffer from interference. Low-level continuous background noise did not interfere with recognition; in fact, the workstations in our laboratory reside in close proximity to a vacuum-fume hood. Other than placing stations in areas less affected by sporadic noises or background conversations, no specific measures such as installation of special cubicles or sound-insulated walls (4) were necessary. Others have also reported little impact of ambient noise on effectiveness of current VR systems available for clinical use (3).

Our VR implementation project devoted much attention to user training. First, two "super users" (supervisor and experienced PA) initially underwent training and then became the primary trainers for others who would use the system. This "train-the-trainer" approach, with the trainers then being readily available in the work environment, greatly facilitated other users' proficiency. Second, a training station was established in a location that was separate from the work area. The main reason that the training station initially was placed separate from the work area in a relatively noise-free area was to eliminate distractions and to improve focus while the greatest number of users was being enrolled and trained. It was also thought that such a location would ensure the creation of an optimal new voice profile during the initial enrollment phase. Once the system was fully implemented and all initial users were trained, enrollment and training of new individual users was able to proceed at a workstation in the laboratory environment. The VR software corrects for any background noise, regardless of the level, during individual enrollment sessions. Crucially important was the allocation of time for new users to be away from daily duties to receive adequate training. During the transition phase for new users the routine dictation system was kept available as a backup. As expected, the learning curve among users was variable. Most users were proficient in the system after about four weeks experience. Some users required a longer time to gain familiarity and an acceptable comfort level.

Several significant challenges were encountered during the implementation of VR, some of which persist as drawbacks today. When using VR to generate reports, users must perform proofreading and editing functions that transcriptionists (or pathologists) would otherwise perform. Error correction occurred at the time of report creation in the VR system, when the PA or SPT saw that an incorrect word, number, or phrase was inserted. In this manner, error correction and proofreading did add some additional steps to the PAs' and SPTs' work compared with conventional human transcription, because with conventional dictation the PA or SPT simply speaks into a microphone and does not typically review gross descriptions after dictating them. On the other hand, because PAs or SPTs could easily review and edit gross descriptions in the VR system and because the information in the gross reports was templated, pathologists needed to do little to no editing of them. The amount of error correction required varied depending on speaker and/or exact words. Biopsy reports in which the voice entries were mostly numbers (for measurements) generally required less correction than did low complexity specimen reports for which more trigger words were used. In general, recognition accuracy for words, numbers, and triggers ranged from 70–90%. As mentioned earlier, templated descriptions and trigger phrases greatly reduced the reliance on the accuracy of recognition by the VR system, because the number of words that the system inserts into the report was actually much greater than the number of words that the VR system must recognize.

Designing and formatting templates required considerable time and meticulous attention to detail. Over 160 hours were necessary to design and test the templates. Numerous edits were necessary to bring the formatting for items such as paragraph numbering and specimen part designation into uniformity with the departmental format for surgical pathology reports in the APLIS. In some instances, the precise formatting and/or punctuation desired were not possible to create.

Because pathology reports are ultimately generated from the APLIS, proper functioning of the interface that transfers information between the VR and APLIS is crucial. Considerable difficulties occurred in stabilizing the interface between our two systems. Frequently, errors returned when the VR system would query the APLIS for patient demographics and accession information. Correction of the problems required extensive consultation among the departmental system analyst and the VR and APLIS vendors. As also noted by others (4), working with different vendors who do not necessarily share the same priorities represents a significant challenge in this type of VR installation. Of note is that some APLISs now have available options for VR capability integrated into the system; such a feature eliminates the need for an interface (although such integration does not necessarily affect positively or negatively the other aspects of the particular VR functionality).

The VR system requires expert support and maintenance in three respects: the VR application itself, technical aspects of the VR software, and hardware requirements. Application maintenance includes mostly creating and editing templates. A "superuser" PA who has received necessary training performs most of these tasks in our laboratory. Technical support includes troubleshooting network problems, fixing PCs or components such as sound cards, monitoring the server, maintaining users' accounts, and assuring overall system performance. In our department, individuals of the laboratory computing unit with expertise in the VR application, PCs, networks, and hardware perform these technical support tasks.

In summary, this study documents the utility and cost effectiveness of voice recognition technology implementation in surgical pathology targeted at generation of gross descriptions and reports. The use of VR technology for gross descriptions of biopsies and of low- to moderate-complexity specimens and for gross-only final reports facilitated data entry, reduced transcription costs, and contributed to improved report turnaround time. Use of templated gross descriptions and trigger phrases was important to successful implementation of the VR system in pathology. Such a targeted implementation requires less capital outlay compared with the case of complete replacement of conventional dictation and may be attractive to pathology departments that either are interested in gaining experience with VR technology or are investigating methods to reduce transcription costs.

REFERENCES

1. De Bruijn LM, Verheijen E, Hasman A, Van Nes FL, Arends JW. Speech interfacing for diagnosis reporting systems: an overview. Comput Methods Programs Biomed 1995;48:151–6.
2. Makhoul J, Schwartz R. State of the art in continuous speech recognition. Proc Natl Acad Sci U S A 1995;92:9956–63.
3. Zafar A, Overhage JM, McDonald CJ. Continuous speech recognition for clinicians. J Am Med Inform Assoc 1999;6:195–204.
4. Teplitz C, Cipriani M, Dicostanzo D, Sarlin J. Automated speech-recognition anatomic pathology (ASAP) reporting. Semin Diagn Pathol 1994;11:245–52.
5. Meijer GA, Baak JPA, van Diest PJ, van Hattum AH, van der Linden HC, Koevoets JM. Text processing by digital voice recognition. Anal Quant Cytol Histol 1996;18:261–6.
6. Mehta A, Dreyer KJ, Schweitzer A, Couris J, Rosenthal D. Voice recognition—an emerging necessity within radiology: experiences of the Massachusetts General Hospital. J Digit Imaging 1998;11(2 Suppl):20–3.
7. Leslie KO, Rosai J. Standardization of the surgical pathology report: formats, templates, and synoptic reports. Semin Diagn Pathol 1994;11:253–7.