Journal of Digital Imaging

Speech Recognition Interface to a Hospital Information System Using a Self-Designed Visual Basic Program: Initial Experience

Edward C. Callaway, Clifford F. Sweet, Eliot Siegel, John M. Reiser, and Douglas P. Beall

Speech recognition (SR) in the radiology department setting is viewed as a method of decreasing overhead expenses by reducing or eliminating transcription services and improving care by reducing report turnaround times incurred by transcription backlogs. The purpose of this study was to show the ability to integrate off-the-shelf speech recognition software into a Hospital Information System in 3 types of military medical facilities using the Windows programming language Visual Basic 6.0 (Microsoft, Redmond, WA). Report turnaround times and costs were calculated for a medium-sized medical teaching facility, a medium-sized nonteaching facility, and a medical clinic. Results of speech recognition versus contract transcription services were assessed between July and December, 2000. In the teaching facility, 2,042 reports were dictated on 2 computers equipped with the speech recognition program, saving a total of US $3,319 in transcription costs. Turnaround times were calculated for 4 first-year radiology residents in 4 imaging categories. Despite requiring 2 separate electronic signatures, we achieved an average reduction in turnaround time from 15.7 hours to 4.7 hours. In the nonteaching facility, 26,600 reports were dictated with average turnaround time improving from 89 hours for transcription to 19 hours for speech recognition saving US $45,500 over the same 6 months. The medical clinic generated 5,109 reports for a cost savings of US $10,650. Total cost to implement this speech recognition was approximately US $3,000 per workstation, mostly for hardware. It is possible to design and implement an affordable speech recognition system without a large-scale expensive commercial solution.

KEY WORDS: computer, speech recognition, picture archiving and communication systems, interface, composite health-care system

THE EFFORT to improve patient care by collapsing the diagnostic and therapeutic timeline has driven computer applications development in a variety of areas. Tremendous improvements in hardware and software over the last decade have stimulated this progress. With picture archiving and communication system (PACS) technology, images are available immediately throughout the health care system, but there continues to be a lag in the transmission of the corresponding completed radiology reports.[1,2] The purpose of this report is to relate our experience with the development and integration of an off-the-shelf speech/voice recognition application into a hospital information system (HIS) using a graphical interface program developed by one of the authors.

Speech recognition in the radiology department setting decreases overhead expenses by reducing or eliminating transcription services or as a means to improve patient care by reducing report turnaround times.[2,3] Significant problems can arise in facilities that attempt to integrate a speech recognition system into the HIS.[4] This can be difficult particularly in the setting of a training program with rotating residents,

From GI Imaging & Teleradiology, David Grant USAF Medical Center, Travis AFB, CA; Diagnostic Imaging/SGSAR, 72d Medical Group, Tinker AFB, OK; Dept. of Veterans Affairs Medical Center, Baltimore, MD; Diagnostic Imaging Associates, Sisters of Charity Hospital, Buffalo, NY; and Wilford Hall Medical Center, San Antonio, TX.

Correspondence to: Edward C. Callaway, GI Imaging & Teleradiology, David Grant USAF Medical Center, 60th MDG/SGQX, 101 Bodin Circle, Travis AFB, CA 94535; tel: 707-423-7182; fax: 707-423-7207; e-mail: edward.callaway2@60mdg.travis.af.mil

The views in this material are those of the authors, and do not reflect the official position of the U.S. Government, the Department of Defense or the Air Force.

Online publication 30 April 2002
doi:10.1007/s10278-001-0007-y

changing staff assignments, and independent electronic signature/verification processes for both groups.

The time it takes for a clinician to receive a complete dictated and signed report is based on 3 major waiting periods associated with finite time intervals: (1) the workload queue, (2) the transcription queue, and (3) the verification or signature queue (Fig 1). The workload queue is the time it takes from examination completion until the radiologist dictates the report and can be demonstrated physically as a stack of films or a PACS work list. The transcription queue is the time it takes to transcribe the report after it has been dictated. This can be represented physically as a stack of cassette tapes or digital voice files waiting for transcription. The verification queue is the time it takes the radiologist to edit and eventually sign the report once it has been transcribed. Mistakes in a transcribed report repeat the processes because the corrected report must go back to the transcriptionist, or the radiologist must manually edit it by typing or writing corrections. The application of speech recognition technology can decrease significantly the time to complete the transcription and verification queue, occasionally at the cost of an increase in the time it takes to complete the workload queue.[5]

All US military medical facilities use the Composite Health-Care System or CHCS (Science Applications International Corporation, Washington, DC), a comprehensive hospital information system that handles every facet of the medical facility operation, including order entry, results reporting, and all administrative record keeping and data mining. A complete radiology information system (RIS) is included as a subprogram of the HIS. Every radiology report, no matter how it is generated, must be entered into the CHCS for the permanent medical record.

A number of articles have been written over the last 5 years describing the use of speech or voice recognition in the radiology department.[2,3,7,8] Several vendors have produced proprietary applications coupled with off-the-shelf speech recognition engines with interfaces into various HIS/RIS systems. Currently, however, there are no vendors with an adequate means of interfacing into CHCS. One recent article describes the problems that can occur with some proprietary applications, including lack of adequate vendor support, inability to electronically sign reports after dictation, and inability to change the name of the staff supervisor during the dictation session.[4]

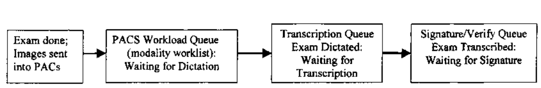

Fig 1. The 3 major waiting periods in the radiology reporting process—dictation, transcription, and signature.

Because of a complete absence of transcription support, one military radiologist working at an outpatient clinic developed a computer program that provides a bridge between the speech recognition program or voice engine (Dragon Medical Professional 4.0; Lernout and Hauspie, Burlington, MA) and CHCS. The interface program was written using the Windows programming language Visual Basic 6.0 (Microsoft, Redmond, WA). This program (named VoicePatch) addresses most, if not all, of the interface problems that have been identified thus far within the literature.[4]

It is recognized that considerable controversy exists about how speech recognition affects individual productivity. We did not choose to address this issue, but rather addressed the end result, the completed report, and the associated implementation costs.

FACILITIES

David Grant USAF Medical Center (DGMC) is a medium-sized military medical teaching facility. The department of radiology interprets approximately 85,000 inpatient, outpatient, and teleradiology examinations each year. The department is a completely filmless PACS environment except for mammography, using General Electric Pathspeed (version 7.12; General Electric, Milwaukee, WI) workstations. Eleven Pathspeed workstations are present in various reading areas of the department. During the period of this study, only 2 computers equipped with VoicePatch were available for dictation purposes. One was fixed in the gastrointestinal (GI) reading area and one was a "mobile" laptop system. There are 11 staff radiologists and 11 radiology residents. Of these, one staff member (ECC) uses the VoicePatch system almost exclusively on the laptop computer, whereas residents use the system when assigned to the GI reading area.

Wright Patterson USAF Medical Center (WP) is a medium-sized nonteaching medical facility that performs approximately 72,000 inpatient and outpatient examinations per year and also is filmless, using the same General Electric Pathspeed PACS. Seven reading workstations are present, there is one PC-based VoicePatch system at each workstation. All but one staff radiologist use the speech recognition system.

The Tinker Air Force Base (TAFB) medical facility is a medium-sized military outpatient clinic that performs approximately 10,000 outpatient examinations per year using film-chemistry–based imaging. There is one VoicePatch system located in the main reading room. No transcription-based service is available, and speech recognition is used exclusively.

HARDWARE

At DGMC 2 types of computers were used. One was a standard desktop PC with a Pentium II 500-MHz CPU, 128 MB of RAM, a 9-GB hard drive, a CD-ROM, and a headset with microphone. The other was a laptop PC equipped with a Pentium III 700-MHz CPU, 256-MB of RAM, a 20-GB hard drive, a CD-ROM, and a Universal Serial Bus headset microphone. Both computers have 10/100 Mb ethernet cards. The desktop PC used Windows NT, whereas the laptop used the Windows ME (Millennium Edition) operating system. At WP, standard desktop PCs equipped similar to the one listed above were used. The reading area at the TAFB clinic was equipped with a desktop PC using a 450-MHz CPU, 384-MB RAM, and a 20-GB hard drive. The system at TAFB uses a Shure TCHS (Shure Corporation, Evanston, IL) wireless headset microphone, whereas the systems at WP all use a Philips Speechmike (Philips Speech Processing, Vienna, Austria).

SOFTWARE

The VoicePatch program, with associated files, was distributed either via FTP download from a military Internet site or from a recordable CD mailed to the facility. Installation was relatively straightforward, requiring loading of 3 executable files and 1 or 2 datafiles (Microsoft Access database) that contained a list of users and a relationship table of template reports. No other software was necessary; however, the Microsoft Access program was installed to edit the templates and the "authorized user" database files. The Dragon Professional Medical Software was installed first, followed by the VoicePatch program. Using the base general medical vocabulary within Dragon, a partially edited and revised vocabulary was created to lower the misrecognition rate.

DATA

At DGMC, data were collected for all reports dictated by first-year residents over 6 month trial period to examine the interface with the HIS and evaluate for bugs in the program. Report turnaround times (TAT) were calculated and averaged from July to December 2000 for the VoicePatch speech recognition system (SRS) and compared with a corresponding transcription service (TRANS) using the Remote Telephone Access System (RTAS Sudbury Systems, Sudbury, MA) during the same period. TAT were calculated as the time from "examined" status (designated by the radiologic technologist in the HIS log as the time the examination was finished and images were submitted to the PACS for interpretation) until "complete", defined as the time stamp for the final electronically signed or verified report within the HIS). Because only 2 speech recognition units were available for 12 PACS workstations, one fixed within the relatively low volume GI reading area, only the TAT for 4 types of examinations were evaluated. To give a representative cross section of the department workload, we included all chest, all abdomen, some bone (hand and foot studies only), and all fluoroscopic studies. Any examination result with a TAT of greater than 165 hours (1 week) was excluded, because this was most likely caused by the resident or staff radiologist not signing reports because of vacation, holidays, or waiting for peer consultations. Staff radiologists at DGMC using speech recognition without a resident were excluded, because this would eliminate 1 of the 2 tiers of the verification process, and the data for this situation would be demonstrated by the nonteaching facilities.

An informal calculation of accuracy rates was performed with the DGMC residents by extracting the average words per report from the HIS and dividing by the number of errors per report, leaving a words-per-error value. This was performed for the first 10 to 20 reports for each resident and demonstrated recognition rates greater than 90%. Several previous studies have confirmed recognition rates for speech recognition software from 90% to 98%.[5,6,7,9]

At WP, all dictated reports by all radiologists were included. At this site, only one tier of verification is required, that of the dictating radiologist. TAT were calculated and averaged by month for both traditional transcription and speech recognition.

At TAFB, all reports were generated and electronically signed at the time of interpretation. Because the radiologist at this site had no transcription support, no comparative data could be obtained. Because reports were signed immediately, TAT would reflect only delays in the worklist queue, and these data were excluded.

PROCESSES

To design and evaluate how the Voicepatch program interacts with the HIS, a thorough understanding of the dictation/transcription process and institutional workflow is necessary. The conventional process for transcription entry of reports at the DGMC teaching institution is structured as follows (Fig 2): the resident or staff radiologist dictates a report using a digital recording transcription service and at the same time types a preliminary report into the PACS system for immediate availability to clinical users thereby increasing work by as much as 100%. This dictated report enters a recording queue to be typed into CHCS by a transcriptionist. The resident may not know which staff member will be overreading. In the former circumstance, the resident notes this during the original dictation. In the latter circumstance, the staff member must dictate a separate note into the RTAS system. If the resident is not present to discuss the interpretation, the staff member must rely on what was typed into the preliminary report screen on the PACS during overread sessions. After the transcriptionist enters the report into CHCS, the report enters an electronic signature or verification queue. In a nonteaching facility, the report reaches a final or "complete" status once the radiologist has viewed the report within CHCS, made any necessary corrections, and electronically signs the report. In a teaching facility, the resident must proofread, edit and sign the report, at which point the report enters a second verification queue for the staff radiologist. The resident has 5 days to verify the report and, if not verified,

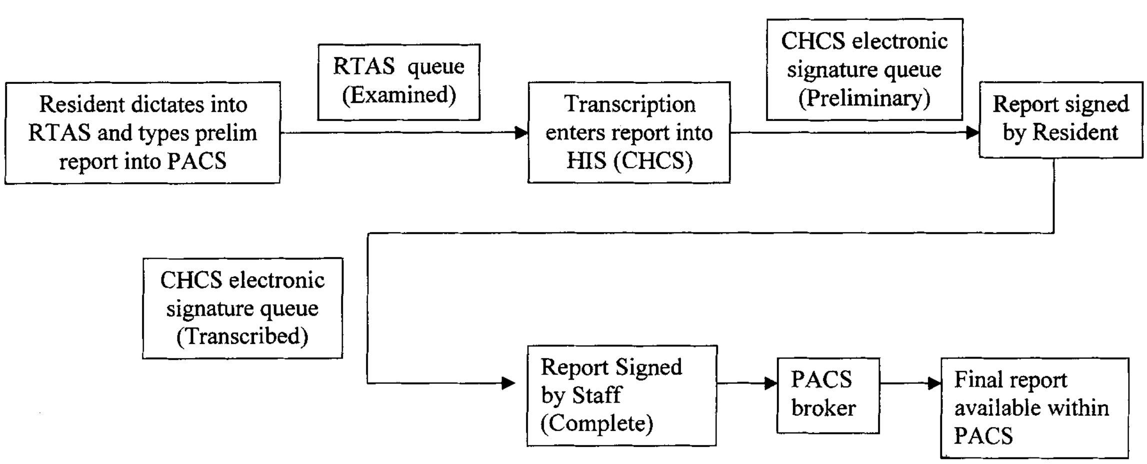

Fig 2. Flowchart depicts conventional transcription process with dual electronic signature requirements, one for residents and one for staff radiologists, within the HIS/RIS. HIS/RIS status tag denoted in parentheses.

the report is forwarded automatically to the staff radiologist's verification queue. The final corrected report is not available within CHCS to all physicians until electronically signed by both the resident and the staff radiologist. Unfortunately, this may take days and, in some instances, weeks, especially when the resident staff member takes vacation. Although this is not critical within the hospital because of the preliminary typewritten report within the PACS, this can be a significant problem for teleradiology sites that do not have access to a PACS terminal.

Different problems exist with traditional transcription methods. One problem can occur when a staff member dictates a countersignature note into the RTAS system. The original examination report usually is entered by giving the patient identification information and an accession number. The recorded RTAS report can be accessed via telephone by any clinician entering the same accession number. When the clinician hears only the words "add staff signature" or a similar statement, it is a waste of time. The clinician may have to step through a number of these unhelpful sound bites before hearing an actual dictated report. Another problem is "lost" dictations. Occasionally, whole reports are either not recorded or somehow dropped from the RTAS system. Monthly lists are generated for examinations that need to be redictated. This is a particularly painful event for long dictations such as computed tomography (CT) examinations or interventional procedures. Occasionally, transcription personnel have difficulty hearing or understanding dictations, especially when the radiologist values speed over enunciation and frequently when reports are dictated by nonnative English speakers. This usually results in a series of asterisks and exclamation points within the dictation. Lastly, a lag in the transcription "tape" may cause important modifiers such as "no" to be lost. This may not be apparent when proofreading the report anywhere from hours to days later.

With the VoicePatch/Dragon program, an examination report can be entered into the HIS, completely bypassing the transcription and signature process (Fig 3). This allows for instantaneous transmission of a report immediately after an examination which is invaluable in urgent situations. There are 4 alternatives in the dictation matrix in a teaching hospital: a resident dictating without knowing who the supervising staff will be; a resident dictating and knowing who the staff will be, but the staff is not present; a resident and staff dictating together; and a staff dictating on his/her own. The latter 2 alternatives show that a report goes to "complete" at the time of dictation, bypassing all intermediate steps and becoming real-time reporting. The only physical delay in reporting is how long it takes to get through the worklist.

The VoicePatch system communicates, in a fashion, with the CHCS HIS. The program navigates through the menus of the HIS, duplicating keystrokes that would otherwise be entered by a human transcriptionist. To begin the dictation, the user simply says "enter exam" followed by speaking the examination accession number or, alternatively, by typing the accession number or using a barcode scanner. As an option, the text-to-voice feature within Dragon can be activated, and the patient data retrieved by the program can be read back to the user. This allows the radiologist to confirm the correct patient data by listening rather than removing his/her eyes from the images. This is helpful to some radiologists who dislike looking back and forth from the PACS monitor to the dictation screen. At completion of the report, the radiologist uses the command "send report" and can see (and hear) within the VoicePatch GUI that the report is crossing into the HIS. This is done by tying text-based events within CHCS to a sound bite and a colored graphics level that corresponds to each step of the report signature process. When the report disappears from the screen, it has made it successfully into the HIS. In the rare event of an unsuccessful entry, the report remains on the screen and an error box appears.

The user can enter reports for a single examination or multiple examinations or add an addendum to previously completed reports. If the examination has not been verified previously (for instance when a resident dictates the case), the program will pull up the complete body of text that was entered previously into the HIS, either by the VoicePatch program, or the transcription service. This is especially

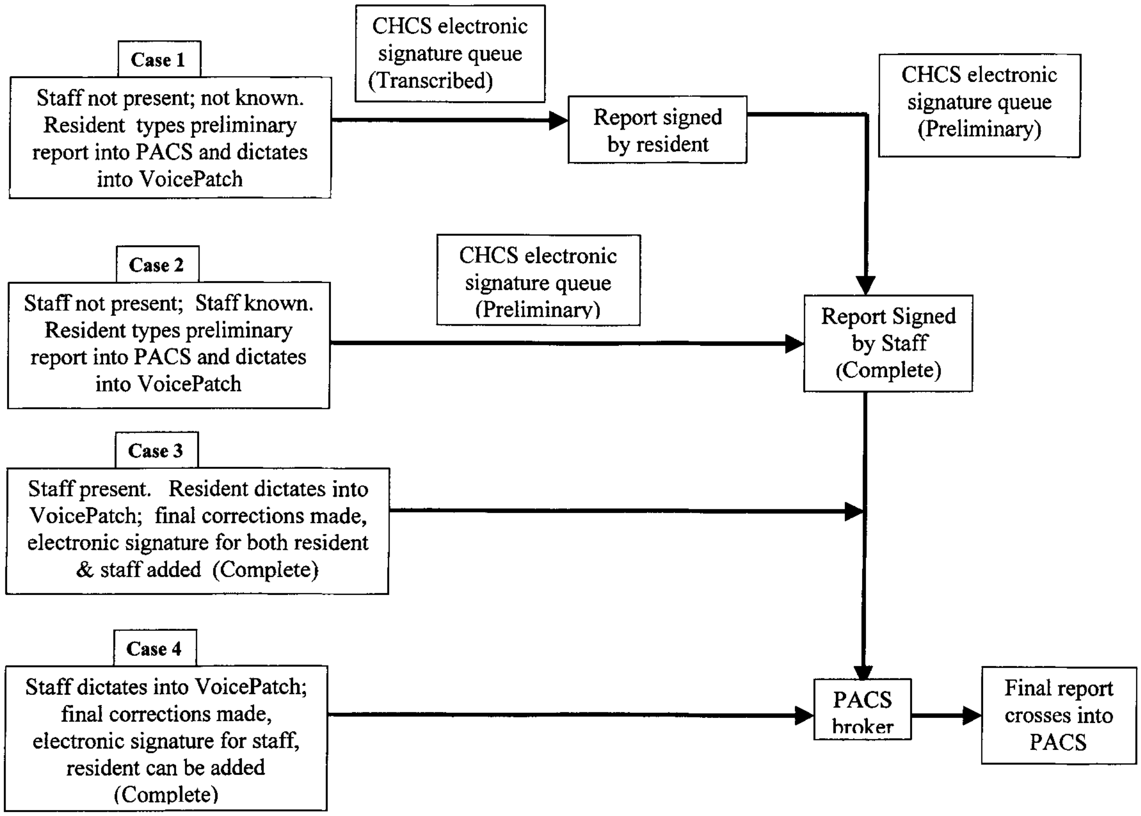

Fig 3. Flowchart depicting VoicePatch speech recognition (SRS) process. There are 4 alternatives in the dictation matrix: a resident dictating without knowing who the supervising staff will be; a resident dictating and knowing who the staff will be, but staff is not present; a resident and staff dictating together; and a staff dictating on his/her own. HIS/RIS status tag denoted in parentheses. The last 2 alternatives show that the report goes to complete at the time of dictation, bypassing all intermediate steps, ie, real-time reporting. The only physical delay in a report is how long it takes to get through the worklist to a given examination. Urgent reports can be dictated and signed immediately.

valuable because what the residents dictate and what they type into the PACS for a preliminary report can be very different in terms of content and detail. Unfortunately, if the resident used the RTAS system, the report may not be available because of delays in the transcription queue. Using links to a separate database file, the VoicePatch program can enter standardized template reports. Another database file is maintained that separates users as staff or resi-

Table 1. DGMC Workload for Transcription (Trans) Versus the Speech Recognition System (SRS), VoicePatch, Jul-Dec 2000

	July	Aug	Sep	Oct	Nov	Dec	Total
Reports							
Total Trans	5,099	5,485	5,231	3,403	5,020	4,796	29,034
Total SRS	180	331	339	318	291	323	1,782
Combined Total	5,279	5,816	5,570	3,721	5,311	5,119	30,816
SRS % Total	3	6	6	9	5	6	6
Lines							
Total Trans	84,550	120,577	103,653	70,482	103,588	100,868	583,718
Total SRS	2,190	3,897	3,903	3,309	3,893	3,894	21,086
Combined Total	86,740	124,474	107,556	73,791	107,481	104,762	604,804
SRS % Total	3	3	4	4	4	4	3

Note. Using only 2 (of 12) workstations equipped with the VoicePatch interface, over $3,300 savings were generated at only 6% of the total lines dictated ($0.13/line).

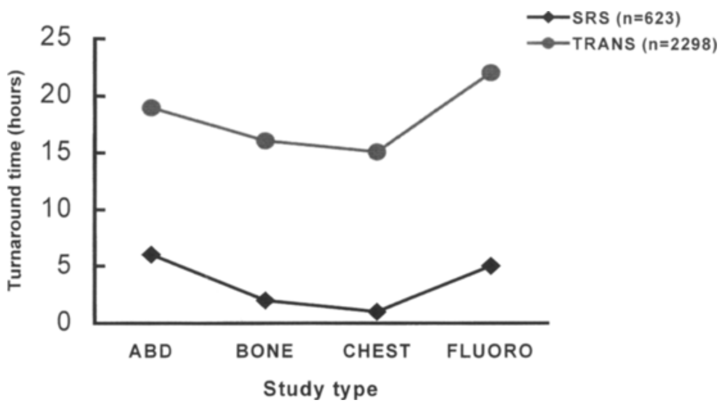

Fig 4. DGMC breakdown of average turnaround time in hours by examination type, speech recognition system (SRS) versus transcription (Trans). Bone and chest examinations are signed off quickly, usually after. they have been performed; however, abdomen and fluoro cases are read on the gastrointestinal service in the afternoon during a separate reading session. This creates an artificial delay in the time to completion.

dents and allows dictation with any matrix combination of resident or staff. A resident user may select a staff as supervisor, or change the supervisor name, at any time during the dictation. However, once the staff supervisor has been selected and the resident electronically signs or verifies the report, the supervisor name cannot be changed using the program. The staff radiologist can countersign the report at any time using either the CHCS, the RTAS system, or by VoicePatch.

Errors in the program that cause a fatal crash occur occasionally and are dealt with by simply closing the program and restarting, which takes about 90 seconds. If the user was in the middle of a report, the text automatically is copied onto the Windows clipboard and can be retrieved by pasting it back into the report text field after entering the examination number again.

RESULTS

During the time frame of the study, a total of 32,448 reports (607,815 lines) were generated at DGMC. Of these, 1,782 reports were entered into the HIS using the VoicePatch system or 6.3% of the total for a total of 25,533 lines of dictation or 4.2% of the total lines generated (Table 1). First-year radiology residents dictated 2,921 reports in the categories, chest, abdomen, bone and fluoroscopy. A total of 623 reports were completed using the VoicePatch speech recognition system (SRS) and 2,298 reports using the traditional transcription method (TRANS). Summation of the TAT for the 4 examination types evaluated at DGMC are shown by month (Fig 4). Despite the 2-tiered electronic signature verification scheme, TATs for the VoicePatch SRS were consistently better than transcription, achieving an average time of 4.7 hours compared with 15.7 hours for transcription (Fig 5). The number of examinations reaching a complete status within 2 hours of the examination also was better for the SRS when compared with transcription (25% versus 7%), with 91% of dictations reaching "complete" status during the same workday (Table 2).

At WP, of 26,651 total reports dictated, 19,405 were done using speech recognition. This generated a total of $255,026 lines of text. Because transcription charges 17 cents per line, a total savings of approximately US $42,079 at this facility was realized (Table 3), not including expenses. For the VoicePatch speech recognition reports, the average TAT was approximately 19 hours (0.83 days) compared with 89 hours (3.49 days) for the transcription service (Table 4). The transcription/verification queues

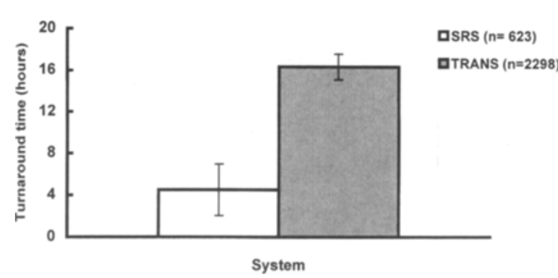

Fig 5. DGMC summary of average turnaround time (in hours): for the traditional transcription system compared with the VoicePatch speech recognition system (SRS) interface. Error bars designate the upper and lower limits of the 95% confidence interval. Transcription average is 15.7 hours compared with 4.7 hours for the VoicePatch system.

Table 2. DGMC Turnaround Time Within 24 hours: Number and Percent of Reports Completed Within The Time Specified

	<1 Hr	<2 Hr	<3 Hr	<4 Hr	<5 Hr	<6 Hr	<7 Hr	<8 Hr	<24 Hr
SRS	74	156	234	299	380	458	526	573	612
%	11.8	25.0	37.4	47.8	60.8	73.3	84.2	91.7	97.9
Trans	35	157	358	602	797	990	1083	1137	1976
%	1.5	6.8	15.6	26.2	34.7	43.1	47.1	49.5	86.0

Note. The VoicePatch speech recognition system (SRS) enables a significantly greater number of completed reports within 2 hours of performance of examination and by end of workday than conventional transcription method (Trans). At 24 hours and beyond, transcription catches up as reports in the HIS/RIS are signed off by the radiologist.

or the "transcribed to complete' time was just over 2 hours for speech recognition reports compared with approximately 16 hours for transcription and verification. Only "urgent" studies were interpreted at the time of examination completion. The remainder of studies were read in the PACS worklists depending on the priorities of the radiologist. That is why there is still a slight time lag overall for report completion; efficiency reverts to depending primarily on the radiologist physically completing the workload queue.

The TAFB clinic generated 67,246 lines of text in 5,109 reports for a total of US $10,650 saved in 6 months, not including expenses. All reports were verified automatically at the end of the radiologist's dictation, which essentially combines the transcription and the verification queue. No TAT comparison could be made to transcription because all reports were dictated using speech recognition.

DISCUSSION

Turnaround times improved at both the medium-sized medical facilities. DGMC showed 75% reduction in TAT for users despite requiring two separate electronic signatures. The TAT at WP improved by 60%, dropping from almost 3.5 days to less than 1 day. This correlates well with prior published results.[2,3,7,11] For sites with only single electronic signature by the dictating radiologist, results are "complete" immediately after entry of the report into the HIS/RIS (if desired; automatic verification can be turned off). At these sites, the only delay in interpretation is the physical delay for the radiologist to look at the images. Routine examinations may sit in the PACS worklist queue (or in the film stack) for some time before being read, depending on other priorities or tasks for the radiologist such as procedures, meetings, or administrative duties.

WP spent US $16,800 for 8 workstations equipped with small flat screen monitors. In the 2 other radiology departments all software was loaded on previously existing computers. The only new hardware required was for microphones of various types, which ranged from US $20 to $200 each. The least expensive were simple headsets, whereas the most expensive

Table 3. Wright Patterson Total Workload by Month for Speech Recognition System (SRS) and Transcription Services (Trans)

	July	Aug	Sep	Oct	Nov	Dec	Total
Reports							
SRS	1,346	2,560	3,126	4,438	3,968	3,967	19,405
Trans	1,786	2,262	1,516	696	538	448	7,246
Total	3,132	4,822	4,642	5,134	4,506	4,415	26,651
SRS % total	43	53	67	86	88	90	73
Lines							
SRS	25,312	25,746	38,439	59,216	53,412	52,901	255,026
Trans	23,611	29,904	20,042	9,201	7,112	5,923	95,793
Total	48,923	55,650	58,481	68,417	60,524	58,824	350,819
SRS % total	52	46	66	87	88	90	73
$ saved	4,176	4,248	6,342	9,770	8,812	8,728	42,079

Note. Cost estimates based on transcription charges of $0.13/line. Seventy-three percent of total lines generated were done using the VoicePatch speech recognition interface, saving over $42,000 in transcription costs.

Table 4. Wright Patterson Summary of Average Turnaround Time (in Days) During Evaluation Period Jul-Dec 2000 for The Speech Recognition System (SRS) Versus Conventional Transcription (Trans)

	July	Aug	Sep	Oct	Nov	Dec	Days	Hours
SRS	0.62	0.90	1.00	0.92	0.68	0.87	0.83	19
Trans	3.52	3.00	2.50	2.46	4.10	5.38	3.49	89

microphones were programmable hand-held units. No special maintenance or support for the Dragon software was purchased, and support for the VoicePatch program itself was provided by the author on the few occasions necessary. The rare occurrence of a fatal error simply required restarting the program.

The purchase of a proprietary integrated system may be very expensive in a large facility with multiple dictation stations. The total cost for complete implementation throughout our department will be approximately US $30,000 for 11 PCs and all accompanying software. The cost to implement a commercial vendor product may run as high as US $5000 to $15,000 per dictation station. The only special software required for this system is the Dragon Professionally Speaking Medical Edition (a newer version, 5.0, is available and is compatible with this program), which ranges from US $400 to $800. Nonetheless, the savings can be significant and have been documented in departments with both limited and widely distributed speech recognition systems.[2,7,11]

Because of the large amount of computer processing required for speech recognition programs, it has been only within the last few years that computer hardware capability has caught up to make off-the-shelf systems affordable. Minimum requirements listed by the speech recognition manufacturers usually are just that, the minimum required to operate a voice system slowly. To make the system usable requires, and is confirmed in our experience, CPU processors of at least 450 to 600 MHz and 128 to 384 MB of RAM. Our initial experience using computers ranging from 233 to 400 MHz and 64 to 96 MB of RAM created delays in response time before text appeared, sometimes as long as 5 to 10 seconds. With the hardware specifications described above, text appears virtually instantaneously. Fortunately, these specifications are standard on new desktop PCs. Two other important components are the microphone and soundcard used.[2,4] The manufacturers of the Dragon software maintain a list of compatible hardware components on their web site (http://www.lhsp.com). Background noise cancellation properties for the microphone are very important, and another helpful tool is an in-line muting switch on the microphone cord that allows the user to quickly turn off the microphone when engaged in other conversation. Two excellent discussions of speech recognition technology have been published previously.[9,10]

The main disadvantage of any speech recognition system is the resistance of staff and residents to the time-intensive learning curve. Despite initial recognition rates of 90%, correcting 1 in 10 words while trying to maintain productivity is at the least disconcerting, especially if the correction process is not mastered. Resistance is also noted secondary to a change in the normal habits of dictation or learning yet another computer system or program. Although continuous speech recognition systems are fast, they cannot accurately interpret the radiologist whose speaking style is to set the record for the world's fastest dictation, nor can it cope well with poor enunciators. Despite the most prodigious efforts, there still is a modestly high probability that words ending and beginning with the same consonant or vowel may be misrecognized. These problems inevitably decrease initial efficiency during readout sessions, and it takes 2 or 3 months of constant use to become comfortable and proficient with the system. New users have a tendency to try to run the speech recognition system and the VoicePatch program strictly using voice commands; however, we believe using a combination of voice, keyboard, and mouse is actually more practical and efficient.

When evaluating the success of implementation, workflow delays caused by having to proofread and correct reports "on the fly" should be comparable with time spent when going back at the end of the day to proofread and correct reports in a traditional dictation setting. It is difficult to estimate the amount of time spent proofreading reports among users as the degree of diligence, and therefore the time penalty incurred, differs significantly. This averages approximately 30 to 60 minutes per day at our facility for each radiologist or resident. However, the speech recognition user must proofread the text during report generation and, conceivably, this could be considered an advantage. Being able to correct mistakes while still looking at the image is a significant advantage over trying to remember, based on surrounding word context, what one observed on any particular previous case, which was sometimes read days before.

No special training was required to run the Dragon program beyond spending approximately 1 hour learning to navigate the editing tools. Voice enrollment also should be performed for at least 30 minutes. The Voicepatch and Dragon programs can be managed by one trained "superuser" to troubleshoot problems. This person can maintain the database tables, or each individual user can maintain a separate database for their own standardized reports, if desired.

Integrating the system into the PACS would be very helpful. Using a separate computer system requires simultaneous use of a separate mouse and keyboard for each system. This can be cumbersome, especially when desktop space is at a premium. Maintaining a central server for voice files would be helpful, and software is available for that purpose from the Lernout & Hauspie corporation; however, we have elected not to implement such a solution at this time. One inadvertent benefit of partial implementation of a speech recognition system may be an improvement in turnaround times by the transcription service secondary to a reduction in the volume of dictated examinations waiting to be transcribed.

CONCLUSIONS

Our speech recognition workstation can be installed complete for less than US $3,000. No special service contract is needed; however, we do train 1 or 2 in-house "superusers" who can do virtually all of the troubleshooting. We expect these workstations to be functional for at least 5 to 7 years. The cost savings at DGMC is expected to be approximately US $25,000 in the first year of use for a partially distributed system performing 30% of the transcription workload. This is confirmed by the WP experience that achieved savings of approximately US $44,500 over the 6-month period, with a crossover point 3 months into the project as all of the radiologists became trained on the system. TAFB has saved approximately US $21,000, the cost of 1 full-time-equivalent employee. TAT in all instances improved a great deal in each of the tested work environments and CHCS variations.

It is possible to integrate an off-the-shelf speech recognition program into an existing HIS and PACS environment using a self-developed interface with standard PC hardware and operating systems. This system is equally compatible in both teaching and nonteaching facilities. We were able to achieve both a cost savings and an improvement in report turnaround time with the VoicePatch system. As we continue to progress to on-line radiology reporting, the adequate integration of any speech recognition system into the HIS/RIS is critical and can be achieved without an expensive, large-scale commercial solution.

ACKNOWLEDGMENTS

The authors thank Dr. Jennifer Thornton, DGMC Clinical Investigations Facility for her preliminary statistical help.

REFERENCES

1. Mehta A, Dreyer K, Boland G, et al: Do picture archiving and communications systems improve report turnaround times? J Digit Imaging 13:105-107, 2000 (Suppl 1)

2. Rosenthal D, Chew F, Dupuy D, et al: Computer-based speech recognition replacement for medical transcription. AJR 170:23-25, 1998

3. Lemmie P, Morin R: The implementation of speech recognition in electronic radiology practice. J Digit Imaging 13:153-154, 2000 (Suppl 1)

4. Houston J, Rupp F: Experience with implementation of a radiology speech recognition system. J Digit Imaging 13:124-128, 2000

5. Zafar A, Overhage J, McDonald C: Continuous speech recognition for clinicians. J Am Med Inform Assoc 6:195-204, 1999

6. Kanal K, et al: Evaluation of the accuracy of a continuous speech recognition software system in radiology. J Digit Imaging 13:211-212, 2000 (Suppl 1)

7. Ramaswamy M, Chaljub G, Esch O, et al: Continuous speech recognition in MR imaging reporting: Advantages, disadvantages, and impact. AJR 174:617-622, 2000

8. Pavilicek W, Muhm J, Collins J, et al: Quality-of-service improvements from coupling a digital chest unit with integrated speech recognition, information and picture archiving and communications systems. J Digit Imaging 12:191-197, 1999

9. Devine E, Gaehde S, Curtis A: Comparative evaluation of three continuous speech recognition software packages in the generation of medical reports. J Am Med Inform Assoc 7:462-468, 2000

10. Schwarz L, Kijewski P, Hertogen H, et al: Voice recognition in radiology reporting. AJR 169:27-29, 1997

11. Mehta A, Dreyer K, Schweitzer A, et al: Voice recognition—An emerging necessity within radiology: Experiences of the Massachusetts General Hospital. J Digit Imaging 11:20-23, 1998 (Suppl 2)