Voice recognition software: psychiatrist as transcriber

Malarvizhi Babu Sandilyan,[1] Jonathan Darley[1]

The Psychiatrist (2013), **37**, 130–134, doi: 10.1192/pb.bp.112.038950

[1]Newtown Centre, Huntingdon

Correspondence to Malarvizhi Babu Sandilyan (dr.malar@gmail.com)

First received 10 Feb 2012, final revision 27 Jul 2012, accepted 16 Oct 2012

Aims and method Voice recognition software is promoted to improve clinician efficiency and decrease overall costs. Our aim was to compare its efficiency against the traditional method of dictation and typing in an older people's community mental health team. We compared the time taken to dictate, edit and type letters, and the total number of days required to send them out after seeing the patient, using the two methods. We also correlated the time taken by one doctor to dictate and edit clinic letters with the actual days on which they were dictated.

Results The voice recognition system reduced the time taken to turn around clinic letters but at the cost of increased doctor's time being spent on dictating and editing the letters. We found no increase in efficiency with experience.

Clinical implications The benefits of faster letter production may be outweighed by the effect of the extra time spent by clinicians to the detriment of their other commitments. The narrative form of psychiatry letters may make them less suited to computer transcription than those in other specialties.

Declaration of interest None.

Continuous voice recognition software or speech recognition software (also known as automatic speech recognition, computer speech recognition, speech to text, or just STT) converts spoken words into text. The term 'voice recognition' is used to refer to systems that must be trained to a particular speaker and are commonly used in healthcare settings to replace or improve the efficiency of medical transcribers (usually medical secretaries). Early systems required the speaker to pause between each word, were slow, and had limited vocabulary and high error rates. Recent advances in this field have generated newer systems that understand continuous speech, run on common personal computers and produce more accurate results.[1]

Voice recognition systems comprise a microphone that converts speech to an analogue electrical signal, which is converted to a digital signal by an electronic circuit board within a computer.[2] 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. The acoustic model removes noise and unnecessary information such as changes in volume. The language model then analyses the content of the speech; it compares the combinations of phonemes with the words in its digital dictionary, a huge database of the most common words in the English language. Most of today's packages come with dictionaries containing about 150 000 words. The language model quickly decides which words were said and displays them on the screen. The commercial software package used for this study is widely used in UK healthcare within National Health Service (NHS) trusts and general practices to produce clinical correspondence using digital dictation and transcription (similar software packages are available in other countries). It includes a wide selection of medical terms and comes with a 'training wizard' which learns new words and also adapts itself to the voice of a new user. The software has wide applications in commercial settings, for example in automated telephone messaging services, which have a limited vocabulary for a wide range of users; other applications of voice recognition may have a large vocabulary trained to work best with a small number of users, such as in digital transcription services.

Voice recognition in healthcare

There is a huge pressure in healthcare settings to generate large amounts of documentation in a short time. The use of computerised voice recognition in medicine was first described in radiology in 1981.[3] In early studies the higher error rate of digital transcription when compared with traditional typing was highlighted. The main reasons for the high error rate then were the necessity to speak each word distinctly in a monotonous voice in order for the computer to recognise it. This also resulted in the doctor spending more time in correcting errors as they appeared on the screen in real time. Things have changed over the years in that newer software is able to recognise continuous speech, thereby eliminating the requirement to pause between

words. Since then this technology has been widely used in various fields of medicine, particularly for producing pathology reports which often use standard templates for producing reports in autopsies and gross descriptions.[4] Nevertheless, computer transcription still required more editing time on the part of the dictating clinician when compared with traditional typing methods. However, it was argued that computer transcription was cost-effective in that it minimised the need for employing human transcribers. Overall, it does seem to have had a positive impact in reducing report turnaround times in radiology departments; the study by Rana *et al* highlighted the reduced time taken to produce reports at the expense of increased editing time by the radiologist.[5] In contrast, in emergency departments voice recognition technology seemed to minimise the error rate in documentation at greater workloads when compared with handwritten documentation.[6] Concerns do exist that doctors, in assuming the role of transcriber, reduce their individual productivity by at least 25%.[7] Also, the system has been criticised for producing more errors when compared with a traditional dictation system and professional transcribers.[5]

The use of voice recognition software to produce clinic letters is spreading slowly across other medical specialties. The obvious difference in documentation in psychiatry when compared with such specialties as radiology, pathology and emergency care, where often synoptic reporting (using a structured, preformatted presentation of clinical information) is applied within a set template, is that psychiatric letters frequently are descriptive and narrative, making correspondence a lengthy process. Therefore the inference from previous studies that have claimed increased efficiency in rapid turnover of letters using templates may not be applicable to psychiatric correspondence. Also, psychiatric letters being lengthier than dictation using standard templates may make them more prone to error, thereby increasing the editing time by clinicians. The fact remains, however, that producing psychiatric clinic letters needs to be faster and more cost-efficient. So far, no study has looked at the usefulness of voice recognition software in psychiatry in achieving this.

The objective of our study was to evaluate the efficiency of commercially available voice recognition software in generating psychiatric clinic letters when compared with the traditional dictation/typing method.

Method

The Huntingdon old age psychiatry team works mainly with out-patient clinics and on average 150 letters are typed and sent every month. The majority of these letters are dictated by doctors and the remainder by psychiatric nurses and mental health social workers. All clinicians use the generic template which is adapted for the care programme approach used in all assessments and reviews of patients.

Training in software use

Two doctors – a medical consultant and a specialist registrar – participated in the trial. The voice recognition software tested was designed for use in professional and healthcare organisations where dictation is used for document production. The doctors were given a 1 h orientating lecture on the features of the software program (Dragon Naturally Speaking, with a psychiatric lexicon), as well as an opportunity to examine the instruction manual provided by the software developer. They were then given a brief, personalised training session on use of the headset microphone and the relevant software. This helped in training the personal computer (Pentium 4, 256 MB, Dell, www.dell.com) in the phraseology and intonation of each doctor. The software incorporated the dictation into Microsoft Word, Office 2003 on a Windows XP platform. This is the standard software used predominantly in most computers within NHS settings.

Voice recognition procedure

When the doctor speaks into the microphone the dictation appears as a Word document, which the speaker then corrects for any errors. At the same time the computer is trained to recognise certain recurring words and phrases. The doctor then emails the corrected text to the administrative assistant using Microsoft Outlook. The assistant again edits and pastes the text so that it fits the electronic template used for all medical correspondence within the trust. The letter is then available on the electronic medical record (clinical document library, known as the CDL) throughout the trust hospitals, and the paper copy is posted to the people concerned on the same day.

Dictation and typing procedure

In the conventional dictation process, letters are dictated into a handheld, portable dictating machine using micro-format tapes. An experienced medical transcriber types the dictation onto the medical template used for all correspondence. This is emailed to the doctor, who corrects it and emails it back to the administrative staff, who then make it available on the electronic medical record. The paper copy is posted to the people concerned on the same day.

Comparison

We compared 47 letters prepared using the dictation method in February and March 2011 against 63 letters prepared by voice recognition technology during March, April and May 2011. The letters were of various lengths, pertaining to the assessment needs of individual patients, and were dictated on random days. This, we believe, simulates the real-life situation that one might expect in common clinical settings. We analysed the total time, average dictation time and correction time required to produce one letter by each method. The doctors timed the steps in each process with a built-in computer clock, rounding to the nearest minute. The administrative staff timed the typing and editing time, also using a built-in computer clock. Results were entered into a Microsoft Excel spreadsheet. Means and standard deviations were calculated using Excel; the Mann–Whitney *U*-test was used to compare the respective medians in the two groups using

SPSS version 16.0 on Windows XP. We used non-parametric testing because the data were not normally distributed and had a high variance.

Results

Forty-seven letters produced by the traditional dictation method were compared with 62 letters using voice recognition software (dictated during April 2011 by the consultant and during March, April and May 2011 by the registrar). The average time taken for each step of the dictation/typing process and the average total time required for each letter are shown in Table 1. The typing was done by the administrative staff and the doctor edited the letter presented to him in paper format. The average total time for each letter is a representation of combined time spent by the doctor and the administrator. The average time taken for the voice recognition software dictation process and the average total time spent on each letter are shown in Table 2. In contrast to Table 1, there are two editing times: one taken by the doctor in real time as the spoken words appear on the computer as text, and the editing time by the administrator who makes any additional corrections and adjusts the text to the template. The average time taken to send a clinic letter from the day the patient was seen was shorter using the voice recognition technology than using the dictation/typing method.

The time required (in days) to send the letter out after seeing the patient and the total amount of time spent on each letter using the dictation/typing method were compared with the times using the voice recognition software with the Mann–Whitney *U*-test (Table 3). We used data collected from the letters dictated by both consultant and specialist registrar for this purpose. The null hypothesis was that there was no significant difference in the total days to send the letter and the total time spent on each letter between the two methods. The test determined that there was a statistically significant difference between the two groups to a level of $P \leqslant 0.001$. To investigate possible improvements with experience using voice recognition technology, we analysed data for letters dictated by the registrar only, as he had used the system for 3 months, whereas the consultant used it for only a month (he had to discontinue as there was a service reorganisation in the interim). Therefore, the time taken by the registrar to dictate and edit the clinic letters using the voice recognition technology was plotted against the actual days of dictation to see whether any correlation existed between the time taken per letter and the day on which it was dictated (Fig. 1). We assumed that there would be a reduction in time taken

Table 1 Dictation/typing: time to produce one letter

	Time per letter
Dictation time, minute: mean	7.1
Typing time, minute: mean	15.3
Editing time, minute: mean	3.0
Total time, minute	
Mean	25.9
Median	21.3
Time to send letter/post on CDL, days	
Mean	10.7
Median (s.d.)	9.0 (5.3)

CDL, clinical document library.

Table 2 Voice recognition software: time to produce one letter

	Time per letter
Doctor	
Dictation time, minute: mean	9.4
Editing time, minute: mean	7.8
Administrator	
Editing time, minute: mean	15.1
Total time, minute	
Mean	41.5
Median	35.2
Time to send letter/post on CDL, days	
Mean	7.2
Median (s.d.)	7.0 (2.9)

CDL, clinical document library.

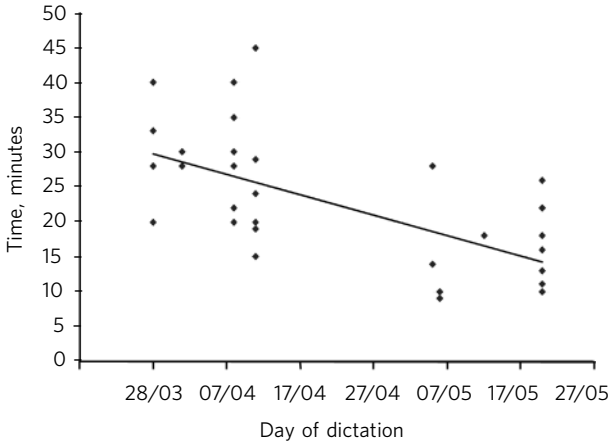

Fig 1 Voice recognition software: relation between time to produce the letter and day of dictation.

Table 3 Comparison of voice recognition software method with dictation/typing

	Voice recognition software $n = 62$	Dictation/typing $n = 47$	*U*	Z^a	*P* (one-tailed)	*P* (two-tailed)
Mean time to send letter	7.2	10.7	2002	3.3	0.0004	0.0008
Mean time spent on each letter	41.5	25.8	2017	3.4	0.0003	0.0006

a. Normal approximate *Z* value.

to produce a letter during the later days, as a result of learning to use the software and with experience gained over the months. The scatter plot shows that the values are widely distributed around the trend line, demonstrating no real correlation between the time taken to dictate and edit the letter and the day in which it was done. However, the average times taken by the doctor to dictate and edit one letter were 28.43 min, 22.95 min and 16.2 min in March, April and May respectively. We feel that if the study was to be continued we might see a statistically significant correlation between the time taken to produce letters with voice recognition method and the number of days gone by.

Discussion

Computerised voice recognition technologies over the years have become more powerful and widely used in various healthcare specialties. Although several studies have explored the use of this technology in various medical fields, none studied its use in psychiatry. In previous studies the major reported benefit of this technology was the decrease in turnaround time of clinical correspondence and reports.[8–11] This result was replicated in our study, in that there was a significant reduction in the average time taken to send out clinic letters. This resulted in better overall satisfaction among clinicians and administrative staff. Quicker turnaround time for letters also meant that general practitioners were informed of the care plan sooner, thereby enhancing overall patient care. Correction of the reports by the clinician immediately after dictation is another advantage over human transcription, where the letter is not available for editing for some days.[8] It is reasonable to believe that immediate real-time editing should reduce errors such as medication dose errors, compared with editing several days after dictating the letter.

Some studies have reported improved efficiency in using voice recognition if English was the first language of the user, with use of a pre-programmed template and with increased experience with voice recognition.[12] Both the doctors who participated in our study had English as their first language and so the data were not compared between them. The doctors also used a common format to dictate letters, with subheadings such as 'diagnosis', 'presenting complaint', 'past history' and 'medications'. In this study we did not have provisions to incorporate the CDL template in the voice recognition software and we recommend this should be considered in future in order to improve the efficacy of voice recognition. However, the use of standardised templates seems beneficial when there are frequently repeated words and phrases, as occurs in radiology and pathology departments. Whether such templates would be useful in psychiatry, where letters are narrative and subject to individual variation, is the million-dollar question. Also, due to lack of literature in this area, it remains unclear whether use of such standardised templates in psychiatry is a more efficient way of producing clinic letters with the conventional dictation/typing methods. More evidence is needed to weigh the benefit of such suggestions.

The average time taken by the specialist registrar to dictate and edit one letter using the voice recognition technology was found to decrease with increasing experience; however, this reduction did not seem significant on the scatter plot, possibly owing to the variable complexity and length of individual letters in psychiatry in comparison with previous studies that have shown such an improvement in efficiency with experience.[12] Again, such studies are about letters that follow a consistent set template. We were unable to compare the data between the registrar and the consultant for the reasons stated above, but the general consensus among the doctors was that the more experienced the doctor was at dictating using a conventional dictaphone, the easier it was to adapt to the voice recognition technology. Also, a standard template might be useful in minimising potential differences in the dictations produced by junior and senior doctors. We suggest that future studies consider this fact and any comparison among doctors at various levels of training would prove beneficial.

Disadvantages

The major disadvantage in using the voice recognition technology as shown in this and previous studies is the increased burden of editing time on users. In our study the total time spent on each letter by both doctor and administrator was significantly higher for computer transcription in comparison with human transcription. This could be due to reported lower accuracy rates for the former when compared with the latter.[4,13] Lower accuracy rates entail added editing time to correct the errors. Pezzullo *et al* found that 90% of all voice recognition dictations contained errors prior to sign-off, whereas only 10% of transcribed reports contained errors.[14] We found that the doctors on average spent an additional 6.4 min per letter using voice recognition compared with dictation/typing, whereas the administrative staff spent on average 0.2 min less using voice recognition than dictation/typing. This places a considerable burden on the doctors while at the same time not being very advantageous in saving the secretary's time. It is concerning that the extra time spent on editing the letters might compromise other commitments that doctors have, such as research, teaching and patient care. The extra time spent by the doctors on each letter, if extrapolated to the whole year, based on the average number of letters dictated per year, adds significantly to the cost of setting up and maintaining the software. Any savings made using the voice recognition technology would be undermined by the lost productivity of the doctors. Such findings of increased aggregate costs using voice recognition software have been reported in previous studies.[14] Perhaps a blended system whereby an experienced transcriber corrects a dictation originally transcribed by the software might prove a solution to this.

The secretary, apart from transcribing, also performs high-level functions such as formatting and grammar checks, which the computer transcription does not do efficiently.[13] This is particularly relevant in psychiatry owing to the free text style of the letters, in contrast to synoptic reporting where the new technology has proved beneficial. Individual variations among users may influence the accuracy of the computer transcription, which may be reduced when used by people with speech problems such as

stammering, and also some individuals might find it easier to detect errors on paper than on the computer screen. This means users might have to formulate their entire report before beginning to speak, rather than 'speaking while thinking', which partly explains the longer time taken to dictate using voice recognition.

The strength of the study is that it replicated real-life practice by including more than one person who dictated letters just as in the routine clinical settings. The limitation of the study is that both doctors spoke English as their native language and had no speech difficulties. An evaluation of a wide range of speakers would have been more meaningful and this provides scope for future studies.

Future directions

Although voice recognition software has technically advanced over the years, it is used with benefit in few medical specialties. It is definitely advantageous in reducing the turnaround time of the clinical correspondence and reports, but at the cost of increasing the editing time by clinicians. The main reason for the increased time taken could be the reported high error rate in computer transcription. Although voice recognition technology is meant to represent a cost-effective way of replacing traditional transcribers, the additional medical time in editing the dictation might minimise any savings in costs. Some studies have reported improved efficiency of the voice recognition software with experience, but this was not reflected in our study. Computer transcription is said to be more efficient when used in conjunction with pre-programmed templates, the use of which is yet to be evaluated in psychiatry. As the system is a form of artificial intelligence which learns from repetitive words and phrases, it might not be as useful in psychiatry (where letters are narrative) as in radiology or pathology (where letters are predominantly synoptic). We recommend future studies to evaluate the benefits of digital dictation and computer transcription to improve the efficiency in clinical correspondence. Such studies will be of value if factors such as first language of the speaker and level of training are taken into account, and if they evaluate the role of a template incorporated in the voice recognition software.

Acknowledgement

We sincerely acknowledge all the support received from the doctors and administrative staff in the Huntingdon Older People's Community Mental Health Team for collecting all the data for this study.

About the authors

Malarvizhi Babu Sandilyan is a specialist registrar in old age psychiatry and **Jonathan Darley** is a consultant psychiatrist in old age psychiatry, both at the Newtown Centre, Huntingdon.

References

1 Alwang G, Stinson C. Speech recognition: finding its voice. *PC Mag* 1998; **17**: 191–8.

2 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.

3 Leeming BW, Porter D, Jackson JD, Bleich HL, Simon M. Computerized radiology reporting with voice data-entry. *Radiology* 1981; **138**: 585–8.

4 Al-Aynati MM, Chorneyko KA. Comparison of voice automated transcription and human transcription in generating pathology reports. *Arch Pathol Lab Med* 2003; **127**: 721–5.

5 Rana DS, Hurst G, Shepstone L, Pilling J, Cockburn J, Crawford M. Voice recognition for radiology reporting: is it good enough? *Clin Radiol* 2005; **60**: 1205–12.

6 Zick R, Olsen J. Voice recognition software versus a traditional transcription service for physician charting in the ED. *Am J Emerg Med* 2001; **19**: 295–8.

7 Hayt DB, Alexander S. The pros and cons of implementing PACS and speech recognition systems. *J Digit Imaging* 2001; **14**: 149–57.

8 Ramaswamy MR, Chaljub G, Esch O, Fanning DD, van Sonnenberg E. Continuous speech recognition in MR imaging reporting: advantages, disadvantages, and impact. *AJR Am J Roentgenol* 2000; **174**: 617–22.

9 Chapman WW, Aronsky D, Fiszman M, Haug PJ. Contribution of a speech recognition system to a computerized pneumonia guideline in the emergency department. *Proc AMIA Symp* 2000; 131–5.

10 Kauppinen T, Koivikko MP, Ahovuo J. Improvement of report workflow and productivity using speech recognition: a follow up study. *J Digit Imaging* 2008; **21**: 378–82.

11 Lemme PJ, Morin RL. The implementation of speech recognition in an electronic radiology practice. *J Digit Imaging* 2000; **13** (suppl 1): 153–4.

12 Bhan SN, Coblentz CL, Norman GR, Ali SH. Effect of voice recognition on radiologist reporting time. *Can Assoc Radiol J* 2008; **59**: 203–9.

13 Schwartz LH, Kijewski P, Hertogen H, Roossin PS, Castellino RA. Voice recognition in radiology reporting. *AJR Am J Roentgenol* 1997; **169**: 27–9.

14 Pezzullo JA, Tung GA, Rogg JM, Davis LM, Brody JM, Mayo-Smith WW. Voice recognition dictation: radiologist as transcriptionist. *J Digit Imaging* 2008; **21**: 384–9.