The University of Osaka
Institutional Knowledge Archive

Title	放射線科レポート作成における音声認識システムの有用性 : ディクテーション経験による比較
Author(s)	市川, 珠紀; 小泉, 淳; 高原, 太郎 他
Citation	日本医学放射線学会雑誌. 2005, 65(4), p. 384-386
Version Type	VoR
URL	https://hdl.handle.net/11094/16886
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

The University of Osaka

放射線科レポート作成における音声認識システムの有用性：ディクテーション経験による比較

市川　珠紀[1)]　小泉　淳[1)]　高原　太郎[1)]　明神　和紀[1)]　山下　詠子[1)]
那須　政司[1)]　柳町　徳春[1)]　今井　裕[1)]　築根　吉彦[2)]

1）東海大学医学部基盤診療学系画像診断学
2）帝京大学医学部付属溝口病院放射線科

Continuous Speech Recognition System for Radiological Reporting: Comparison with experience of dictation

Tamaki Ichikawa[1)], Jun Koizumi[1)], Taro Takahara[1)], Kazunori Myojin[1)], Eiko Yamashita[1)], Seiji Nasu[1)], Noriharu Yanagimachi[1)], Yutaka Imai[1)], and Yoshihiko Tsukune[2)]

Purpose: To compare rates of accuracy of recognition between experienced dictators and inexperienced ones in using an enrollment-less continuous speech recognition (CSR) system of radiological reporting, and to evaluate the usefulness of the system.
Materials and Methods: Twenty board-certified radiologists were classified into 2 groups: a group of 10 members with more than 6 years' experience of conventional dictation by transcriptionist (group A) and a group of 10 members with no experience of dictation (group B). All radiologists created fresh radiological reports on sets of images using free-style dictation in the reports. We counted errors and total words in the reports individually, and compared the rates of accuracy of word recognition in the two groups. We used a CSR system AmiVoice (Advanced Media, Inc., Tokyo, Japan).
Results: The average rate of accuracy of word recognition was 96.42±1.68% in group A and 95.92±1.15% in group B. There was no significant difference in accuracy rate between the two groups.
Conclusion: The accuracy of word recognition was independent of the experience of dictation, and the enrollment-less CSR system of radiological reporting was considered convenient and useful.

Research Code No.: *220.2*

Key words: *Radiological reporting, Continuous speech recognition system, Dictation*

Received Jul. 25, 2005; revision accepted Aug. 30, 2005

1) Department of Radiology, Tokai University
2) Department of Radiology, Mizonokuchi Hospital, Teikyo University

別刷請求先
〒259-1193　神奈川県伊勢原市望星台
東海大学医学部基盤診療学系画像診断学
市川　珠紀

はじめに

放射線科のレポート作成における音声認識システムの利用は1981年から報告がある[1)]．当初の音声認識システムには，語彙の少なさや初めて使用する際にディクテーター自身の音声を学習させる，いわゆる"ならし"(enrollment)などの制限があり，一般的普及には至っていなかった[1)-3)]．最近の音声認識システムは過去のシステムに比し，さまざまな改良がなされている[4), 5)]．最新の音声認識システムは不特定話者タイプであり，enrollmentが不要である[6)-10)]．この機能はenrollment-lessと呼ばれ，ディクテーターの数語の発語内でenrollmentがなされ，トレーニングの時間が不要である．

高原ら[6)]はenrollment-less音声認識システムにおける正解認識率(以下，認識率)を97.1％と報告している．しかしこの数字は既存の文章の読み上げの結果で[6)]，実際の画像を見ながらのレポート作成では認識率は低下すると思われる．以前，著者ら[7), 8), 10)]はディクテーション経験者における実験で，95％以上の高認識の結果を得た．しかし実験参加者は日常診療でトランスクリプショニスト(通称トランスクライバー)によりレポートを作成し，ディクテーション自体に慣れていた．これが高い認識率を得られた理由のひとつではないかと考えた．

今回われわれは自由発語環境下で読影実験を行い，ディクテーション経験の有無における認識率を比較し，放射線科レポート作成における音声認識システムの有用性を検討する．

対象と方法

20名の放射線科医(放射線科経験6年から21年の放射線専門医)を2群に分ける：A群10名は6年以上のトランスクリプショニストによるディクテーション経験者(うち筆者を含む2名が実験に使用した音声認識システムで日常読影を施行)，B群10名はディクテーション未経験者．全員が単純写真5例，CT 3例，MRI 2例，合計10例の同じ画像を読影し，自由発語環境下で音声認識システムにて読影レポート

Table　Erroneous words, total words, and accuracy rate of word recognition in groups A and B

group A: No.	EW	TW	AR (%)	group B: No.	EW	TW	AR (%)
1	28	863	96.3	1	34	632	94.6
2	13	641	98	2	13	644	98
3	7	424	98.3	3	21	538	96.1
4	24	769	96.9	4	25	641	95.1
5	54	729	92.6	5	31	633	96.6
6	25	659	96.2	6	29	864	96
7	25	851	97.1	7	23	575	94
8	33	695	95.3	8	53	889	94
9	30	702	95.7	9	16	525	97
10	20	917	97.8	10	27	632	95.7

group A: experience of dictation.
group B: no experience of dictation.
EW: error words, TW: total words, AR: accuracy rate of word recognition

を作成した．レポート内容，言い回し，文章の長さには制限を与えず，読影速度は日常読影と同程度で読影した．使用したenrollment-less音声認識システムは，アドヴァンスト・メディア社製AmiVoice，Radscribe version 5.14(以下AV)．AVの辞書(言語解析モデル)は市販の約3.5Mbのデータを内蔵しており，Windows 2000及びXPにインストールし使用した．発語入力時間は最速の0.1秒とした．語彙数の算出はAVの語彙数の算出法に準じた．認識率の計算は，認識率＝(正解文総語彙数－置換誤り語彙数－脱落誤り語彙数)/正解文総語彙数×100％とした．統計学解析はt-検定を使用し$P < 0.005$を有意とし，ディクテーション経験の有無による有意差を比較した．また総語彙数(レポートの長さ)と認識率の相関を回帰式で検定した．

結　果

誤り語彙数，総語彙数，認識率をTableに示す．平均認識率はA群では96.42 ± 1.68％，B群では95.92 ± 1.15％であり，ディクテーション経験の有無における認識率に有意差はなかった．総語彙数(レポートの長さ)と認識率に相関はなかった．

考　察

近年，放射線科医は多量の読影レポートを短時間で作成しなければならない状況にある．トランスクリプショニストによるディクテーションは大学病院などの大規模な病院では行われているが，経済面やトランスクリプショニストの普及率に問題があり，小規模病院では難しい．

音声認識システムは迅速な多量の読影レポート作成を可能にする手段の一つである．最新の音声認識システムは，enrollment-lessと速い発語入力時間により，レポート作成時間が短縮化され実用可能なレベルである．AVの発語入力時間は最速0.1秒であり，これは従来のトランスクリプショニストによるワードプロセッサー入力より速い[6]．筆者のAVによるレポート作成１例では136文字が30秒で変換され，この数字はワープロ検定１級の変換速度を上回る．

今回の結果ではディクテーション経験の有無における認識率に有意差はなかった．個人差に関して統計学的検討はしなかったが，最も認識率の低かった医師は最も認識率の良かった医師に比し，認識率が約６％低かった．両者の音声波形の比較では差はなかったが，最も認識率の悪かった医師は低い声でぼそぼそ話す特徴があり，低認識率の誘因と思われた．またディクテーションに慣れた医師はレポート画面を見ず，複数の文章を含む長い文章を一呼吸にディクテーションする傾向がある．音声認識システムでは長文のディクテーションはよどみなくはっきりと話さないと誤変換が多く認識率が低下する．一方，ディクテーション未経験者にとって，音声入力や音声ファイルの聞き返しに恥ずかしさがある．手書きや自身によるワードプロセッサー入力に慣れたディクテーション未経験の医師は一文ずつディクテーションする傾向があり，音声が一文ずつ文章に変換される．これがディクテーション未経験者の高認識率の要因と考えられた．一文ずつディクテーションするのであれば誤変換も少なく，誤りを探すのは容易で最後の修正が正確に行える．

今回の実験で得られた96％以上の認識率は実用レベルに達しているわけではない．音声認識システムの欠点として発語入力時間に対し修正時間がかかる点が挙げられる[9]．過去の臨床での報告では，トランスクリプショニストによるディクテーションに比し，音声認識システムの使用では短い文章を入力する傾向があるとされている[10]．今回の実験では総語彙数(レポートの長さ)と認識率に相関はなかった．認識率に差がないが日常臨床では使用頻度が少なくなるのは，修正時間が長く手間がかかる点が原因ではないかと思われる．

認識率の向上と修正時間の短縮に関して下記の解決法が挙げられる．同音異義語や似たような発語の場合には候補

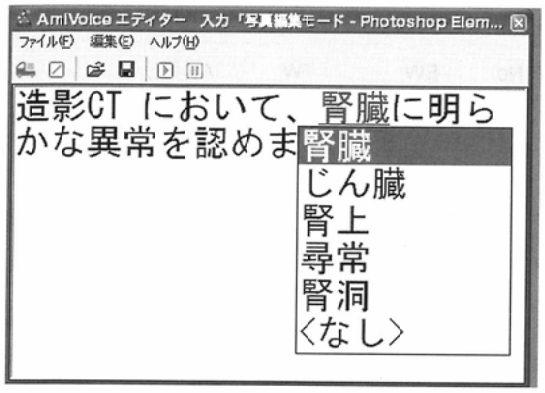

Fig. 1 Examples of words with similar pronunciation. Dictators could choose the correct word in the samples.

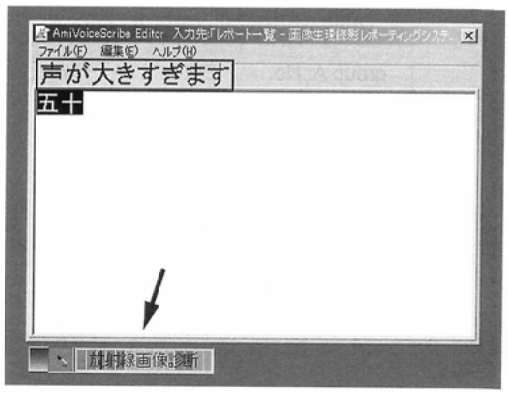

Fig. 2 Appropriate volume zone (arrow) and attention sentence when speaker's voice is too loud for recognition are shown in the report window.

を選択できる(Fig. 1). 今回の実験では使用しなかったが，使用すればさらに正確なレポートを作成できる．また実際のディクテーションでは，レポート画面であいまいな発語と思われる部分が赤字で表示され，読影者に注意を促し，修正が正確にできるよう工夫されている．至適音量はレポート画面に示され，声が大きすぎるときには忠告の文章が表示される(Fig. 2). また長文の定型文を登録しておくなどユーザーの用途に対応し改定や学習を行っている．

トランスクリプショニストによるレポートの認識率は99.8%であり[2), 5)]，音声認識システムの認識率は及ばず，トランスクリプショニストによるレポート作成に取って代わるものではない．音声認識システムでは助詞の選択("は"または"が")が不適切な点や個人の話し癖，文頭や文末の認識の誤り等の問題が解決されず，最終的なレポートの完成には人為的操作が必要である．しかし音声認識システムは忙しい放射線科において，多量のレポート作成のための解決法の一つと思われる．

結　論

自由発語環境下で，音声認識システムによるレポート作成を行い，ディクテーション経験の有無における認識率を比較した．平均認識率はディクテーション経験群では96.42 ± 1.68%，ディクテーション未経験群では95.92 ± 1.15%であり，ディクテーション経験の有無における認識率に有意差はなかった．音声認識システムは忙しい放射線科において，多量のレポート作成のために有用で簡便な方法であると思われた．

謝　辞

多忙な日常業務の中で実験に協力していただいた，若林ゆかり先生，田中　修先生，松浦克彦先生，小林泰之先生，佐藤雅史先生，梶原景子先生，鶴田晴子先生，山田丈士先生，三角茂樹先生，大川伸一先生，小南公人先生，篠原真木子先生，小林由香先生，下山田和裕先生，田中淳司先生に深謝いたします．

文　献

1) Leeming BW, Poster D, Jackson JD, et al: Computerized radiologic reporting with voice date-entry. Radiology 138: 585–588, 1981
2) Rosenthal DI, Chew FS, Dupuy DE, et al: Computer-based speech recognition as a replacement for medical transcription. AJR 170: 23–25, 1998
3) Makhoul J, Schiwartz R: State of art in continuous speech recognition. Proc Natl Acad Sci USA 92: 9956–9963, 1995
4) Ramaswamy MR, Chaljub G, Esch O, et al: Continuous speech recognition in MR imaging reporting: advantages, disadvantages, and impact. AJR 174: 617–622, 2000
5) Papps JN, Durhan NC, Biseet GS, et al: Practical experience with a voice recognition dictation system in a busy academic radiology department. Radiology 217 (P): 401, 2001
6) 高原太郎，中島美佳，似鳥俊明，他：連続音声認識による読影レポート作成の試み．日本医放会誌 62：23–36，2002
7) Ichikawa T, Takenaka H, Fujisawa H, et al: Enrollment-less continuous speech recognition for radiological report. Radiology 225 (P): 678, 2002
8) 市川珠紀，藤澤英文，浮洲龍太郎，他：音声認識を用いた放射線科レポート作成の実際．Rad Fan 2: 58–60, 2003
9) 小野木雄三：連続音声認識を利用したレポートシステム．新医療 28：109–111，2001
10) 北之園高志，櫛橋民生，武中泰樹，他：完全フィルムレス，ペーパーレス病院における全画像検査のレポーティングシステム．Radioligy Frontier 6 (31): 199–204, 2003