Speech Recognition Technology for Dysarthric Speech

PETER E ROBERTS CEng, FIEE
Engineering Department, Faculty of Applied Sciences
Lancaster University
Bailrigg, Lancaster, LA1 4YR
UK

Abstract: - The initial results of investigations into the use of current commercial automatic speech recognition (ASR) software by people with speech disability (dysarthria) is presented, together with a brief summary of the history of the development of ASR and its applications for the disabled. Results confirm the viability of dysarthric use, identify areas of further investigation for improved recognition performance and for development of a clinical tool for speech measurement.

Key-Words: - Software, Speech, Recognition, Dysarthria

1 Introduction

Over recent years, the computer technology for automatic speech recognition (ASR) has advanced significantly, with low cost commercial products now available with very good performance for continuous speech recognition.

Research recently started at Lancaster University, UK, in collaboration with the local National Health Service Primary Care Trust, has been investigating the viability of the use of this current ASR by people with dysarthria speech difficulties.

This paper summarises the history of ASR development, and some of the past research into dysarthria usage of ASR. It then describes the new work done to date at Lancaster. Finally key items of further work are identified.

2 Problem Formulation

2.1 Summary History of the Development of Automatic Speech Recognition System (ASR)

There has been considerable research and development of the technologies of automatic speech recognition (ASR) over the past 50 years.

Early developments (1950s) achieved limited recognition of single words, or digits, from a small vocabulary. These implementations relied primarily on the spectral measurements (by analogue filter banks) of effects produced during vowel sounds of the speech. Examples of these early solutions are from USA Bell Laboratories [1] in 1952, and USA MIT Lincoln Laboratories [2] in 1959.

In the 1960s and 70s, research produced significant steps forward for realistic solutions to ASR. They involved the increased use of techniques to reliably determine aspects of speech beyond the vowel frequency characteristics. The detection of start and end of speech events, and the sound characteristics over the period enabled assessment of the consonants, and the ability to compensate for the timing variability of speech, enabled increased recognition performance. Carnegie Mellon University (USA) commenced long and valuable speech recognition work, building on the phoneme tracking research of DR Reddy [3]. Japan and Russia were also now active in the field of ASR.

IBM and AT&T Bell Labs developed techniques for large vocabulary speech recognition [4], and a degree of speaker independence [5]. The processing was now extending to the linguistic aspects of the speech.

In the 1980s and 90s the speech research moved to new concepts in the matching of input speech with stored speech databases. The use of pre-established templates, albeit now complex, was replaced by statistical modelling and matching methods. Hidden Markov model (HMM) techniques [6], [7], enabled the development of a series of viable products which gave continuous speech recognition, with progressively increasing vocabularies.

HMMs, together with neural nets [8], allow effective and rapid "user training" of the inbuilt speech databases, and give a high level of recognition accuracy (95%), and speaker independence. The increasing availability of high performance desktop personal computing has opened the way for an increasing usage of what is now a reasonably effective speech recognition technology.

2.2 Use of ASR with speech disabilities

It is the increasing availability and performance of modern speech recognition computing that justifies further assessment of viability, and of the potential applications in connection with speech disabilities. (Dysarthria).

There has been early research into the use of ASR for the disabled, but the restrictions of technology, in terms of cost and performance, meant limited practical usage. However, even with the early template based solutions, successful recognition of dysarthric speech could be achieved, with closely limited vocabulary and tailoring of the speech model, [9], [10].

The key issues are the capability of the ASR to adapt to the "non-standard" characteristics of the dysarthric speech [11], and the degree to which the recognition can tolerate an increased variability of certain characteristics of the dysarthric speech [12], [13].

It would appear that the newer product technologies with HMM, (and possibly neural nets), give advantages for these issues. The limited research, carried out over recent years, has examined the effectiveness of these newer products.

Work at Northeastern University, in conjunction with Boston Children's hospital, [14], [15], has shown good recognition in trials with 1, and then 10, dysarthric speakers, using Dragon Dictate software. A panel of listeners also scored the intelligibility of the speakers. The work demonstrated the value and effectiveness of the HMM learning/training process where recognition accuracy increased from 30% to 90% over 3 sessions. It was also observed that the poorer speakers displayed more variability in their speech, including that caused by fatigue. This demanded longer ASR "training" sessions, but ultimately similar levels of recognition were achieved to the moderate or mild dysarthria. Further investigation of the causes, and characteristics, of this variability would be useful, especially with larger numbers of speakers.

Work during 1996/7 at the University of Toronto, Canada, in conjunction with Bloorview MacMillan medical centre [16], [17], also endorses the effectiveness of the ASR learning with the newer software products, especially for severe dysarthria. (6 speakers: 2mild, 2 moderate, 2 severe). 6 sessions were still displaying ASR learning improvements for these severe dysarthria speakers. Included in the research is a comparison between ASR and a perceptual measure of intelligibility by a panel of 10 listeners. Good consistency is observed (as the ASR learning progresses) between ASR and panel. It was also observed that speech training of the dysarthric speakers also led to good improvements in ASR recognition, [18].

Concerns are raised, especially in the Toronto work, on inconsistencies between the ASR (and panel) intelligibility scoring and the clinical assessment tool results (CAIDS tool, [19].) It is noted that continued further research efforts are required to establish reliable and valid methods for evaluating speech intelligibility in dysarthria. Additional work has also been carried out by Kent et al at Winsconsin-Madison University (USA) on dysarthria speech intelligibility testing, [20], with similar conclusions. In the UK the Frenchay tests, [21], and the Robertson tests, [22], are currently in use.

In the UK, newer work is currently being carried out at Sheffield University and at Frenchay hospital, Bristol.

Given the current maturity, and partially proven effectiveness, of the HMM/neural nets ASR technology for dysarthria speech recognition, it is appropriate to look further into two areas:

- Trials on the viability of a current state-of-art product across a wider range of dysarthric speakers, drawing conclusions on recognition effectiveness and developing guidelines for the decisions on appropriateness of ASR for particular conditions/speakers.

- Analysis of the information available during the speech recognition processing, with the aim of developing a clinical tool for use by speech and language therapists in the assessment of ongoing dysarthria therapy.

3 Problem Solution

3.1 Facility

A facility was established to enable speech recording, usually in the subject's own home surroundings, with subsequent data processing and evaluation in the laboratory.

The portable recording facility was a SONY mini-disc digital recorder, (model MZR30), recording from an Andreas headset/microphone, (model NC61). A monitor headphone set was also available to check on sound quality.

The laboratory facility used a SONY desktop mini-disc digital player/recorder, (model MDSJE510), coupled by analogue and optical digital links to a Creative Labs Audigy Platinum PC soundcard. The PC was Intel P4 with 128MB memory, using Windows98 2nd Ed operating system. Commercial ASR software installed was IBM ViaVoice release8 and Dragon Naturally Speaking v5.

Typically 2 speech recording sessions, each of 1 hour duration, were carried out for each subject. The sessions consisted of reading various brief "set-up" texts appropriate for IBM and Dragon commercial ASR packages, followed by the main "enrolment" text used by the ASR to initially tailor the speech database. Near the beginning, and at the end, of each session, 3 short evaluation texts, ("rainbow", "grandfather", and "north wind"), were also included. This gave a consistent means of assessing speech intelligibility, subjectively and via ASR. It also gave a means of examining any trends of intelligibility through fatigue and across the separate sessions. A period of conversational speech was also recorded in each session for possible later research.

11 subjects were recorded initially, with dysarthric speech subjectively assessed, by professional speech and language therapists, as ranging from mild, through moderate, to severe.

3 control subjects were also recorded, with normal speech.

3.2 Speech recordings and initial data evaluation

Recorded speech has now been fully gathered from 11 patients and 3 control, (each 1 to 3 sessions, giving approximately 30 minutes of reading standard texts plus 15 minutes conversation speech, per person.)

The first stage of evaluation and analysis of this data has been to use some recording as enrolment speech for the IBM and Dragon commercial ASR software packages, and then to use the shorter remaining reference texts to evaluate the recognition performance after this initial enrolment/training.

It was initially found that the Dragon Naturally Speaking v5 software was more flexible than the IBM in its use with the dysarthric speech, especially for the initial enrolment process. It was, therefore, decided to concentrate initial evaluation on the Dragon software.

Table 1 and Chart 1 summarise this initial evaluation analysis.

All 3 control recordings give a high level of recognition accuracy, (80-95%), even after the minimum enrolment, which endorses the understanding that the current new versions of SR software are very effective.

Of the 11 dysarthric speakers, the 2 with mild condition and 3 of the 6 with moderate condition were able to successfully set up the ASR software, in its standard "out of the box" configuration. (ie They were able to complete the initial session training the system with their voice - enrolment.) The remaining 3 with moderate condition and all of the severe could not complete the enrolment process.

Subsequent tests of the recognition accuracy, for those who had completed enrolment, showed variation across the patients of 30% to 80%. The poorest recognition occurred with the worst dysarthria.

The lower end of this performance would probably not be considered viable for satisfactory use of the systems, because of the frustration such a high error level would cause. (Although patients with severe physical disabilities have acknowledged that they may be able to tolerate relatively poor recognition performance because it could be their only practical way of using computer facilities without assistance.) Further research (and possible developments) as identified in paras 4.2.1, 4.2.2 are intended to follow on from this initial evaluation.

As noted previously, 6 of the speakers were not able to complete the standard enrolment, (ie they could not complete the initial speech system training process for IBM or Dragon package.) These results could be expected, since 3 of the subjects were assessed as severe, and 2 moderate/severe. However, 1 of the 6 (P9) was subjectively assessed as moderate

dysarthria, but showed a strong nasal characteristic to the speech. Therapist and researcher found it generally easy to understand the speech, but the SR package failed completely, (it was not possible to progress with the enrolment process at all.) This specific condition is considered appropriate for further investigation (paras 4.2.4), since it could identify some means of generally improving the enrolment process for dysarthrics, as well as helping this specific category of patient.

It is also intended to examine further (para 4.2.5) the characteristics of the lower quality speech observed in the 5 successful speakers, to investigate any parameters that can be isolated and improved.

The final area of further analysis currently identified (para 4.2.6) is in connection with the speech parameters of possible value for a clinical tool.

An additional test/control was also incorporated, by attempting recognition of all the speakers by a configuration set up by one of the controls. This gave relatively poor recognition for most speakers, as anticipated, recognising that it was effectively the wrong speaker using the system. This endorsed the relevance of the enrolment process.

4 Conclusion

4.1 Viability established

The initial work has established that the current ASR technologies, as released in 2 commercial packages, are viable for use by mild and by some moderate levels of dysarthric speech. Guidelines will be produced to aid this usage, and to introduce certain specific features (see 4.2.1).

There are, however, certain characteristics of speech, especially with moderate and severe dysarthria, that prevent the enrolment process or result in poor recognition performance. In particular the clarity of inter-word gaps, and the consistent start up and ending of the words appeared very relevant These, and other characteristics, are to be investigated further, (see 4.2.2, 4.2.4, 4.2.5).

Discussions with speech and language therapist professionals, during the work to date, have endorsed the value in investigating further the possibility of a clinical tool for the measurement of certain speech characteristics, and of overall intelligibility. There is also considerable value in a tool for use by patients themselves to support their speech therapy, (see para 4.2.6).

Patients and carers have also commented on the difficulties in disabled use of internet and e-mail facilities. Para 4.2.3 identifies investigation in this area.

4.2 The following areas have been identified for further investigation as part of this ongoing research at Lancaster University:

4.2.1 Introduction of frequently used key words and short-cuts/special commands, to enable more straightforward use by dysarthrics.

4.2.2 Improvement of enrolment viability by simpler reduced vocabulary texts, (but with a consequential reduction of recognition capability.)

4.2.3 Review ways of optimising or modifying the use of SR tools to give more convenient access to the internet and to e-mails.

4.2.4 Closer assessment of characteristics of the highly nasal speech (P9), to identify possible ways forward to enable enrolment and use. This could also lead to more general observations on this specific condition, and what can be done to enable viable use of SR systems.

4.2.5 Closer assessment of characteristics of the poorer performing recognition speech to identify relevant characteristics that are limiting the recognition, and to investigate possible solutions for improvement.

4.2.6 Further analysis of the effects investigated in sections 4.2.4 and 4.2.5 above, in connection with any speech parameters of possible value for a clinical and/or patient tool.

Chart 1 : Recognition %

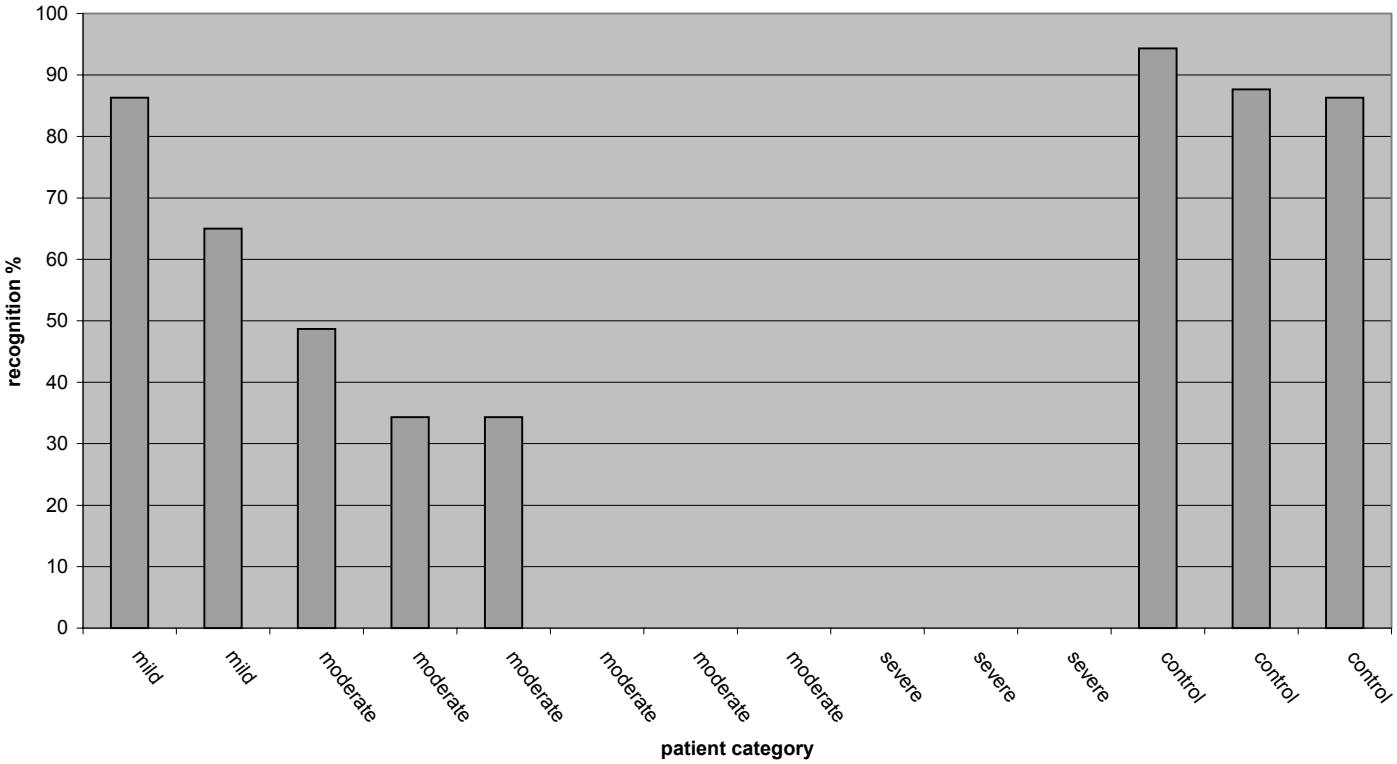

Table 1 : Summary data

	Enrol?		1st analysis tests			
patientID	**y/n**	**cat**	**R**	**GF**	**NW**	**control**
1	y	mild	87	84	88	41
7	y	mild	65	69	61	42
9	n	mod	no scoring yet because enrolment not achievable			
2	y	mod	40	30	33	17
5	y	mod	54	47	45	18
4	y	mod	34	27	42	15
6	n	mod	no scoring yet because enrolment not achievable			
3	n	s		"	"	
8	n	mod		"	"	
11	n	s		"	"	
10	n	s		"	"	
C1	y	-	86	83	90	85
C2	y	-	95	97	91	36
C3	y	-	87	90	86	12

ID: subject identification no., **texts % scores**: R: rainbow, GF: grandfather passage, NW: north wind
Enrol y/n: able to carry out enrolment process **con**: control % score
cat: categorisation of patient dysarthria- mild, moderate, severe.

References:

[1] KH Davies, R Biddulph, S Balashek "Automatic recognition of spoken digits" *Jour Acoustic Soc Amer*, 24(6), 637-642, 1952

[2] JW Forgie, CD Forgie, "Results obtained from a vowel recogniser computer program",*Jour Acoustic Soc Amer,* 31(11), 1480-1489, 1959

[3] DR Reddy, "An approach to computer speech recognition by direct analysis of the speech wave", *Tech report C549, Computer Science Dept, Stanford Univ, 1996*

[4] F Jelinek, LR Bahl, RL Mercer, "Design of a linguistic statistical decoder for the recognition of continuous speech" *IEEE trans, Information theory ,* IT-21, 250-256, 1975

[5] LR Rabiner, SE Levinson, AE Rosenberg, JG Wilpen, "Speaker independent recognition of isolated words using clustering techniques" *IEEE trans, acoust, speech, signal proc,* ASSP-27, 336-349, Aug 1979

[6] LR Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition" *Proc IEEE,* 77(2), 257-286, Feb 1989

[7] F Jelinek, "Statistical methods for speech recognition", ISBN 0-262-10066-5

[8] A Weibel, T Haragawa, G Hinton, K Shikano, K Lang, "Phoneme recognition using time delay neural networks", *IEEE trans, acoust, speech, signal proc,* 37, 393-404, 1989

[9] WW Ahmed, "Computer recognition of cerebral palsy speech", P*roc speech tech conf*, 205-209, 1985

[10] M Fried-Oken, "Voice recognition device as a computer interface for motor and speech impaired people", *Arch Phys Med Rehabil,* 66(10), 678-681, 1985

[11] F Chen, A Kostov, "Optimisation of dysarthric speech recognition", *Int conf IEEE engineering in medecine & biology v4*, 1436-1439, 1997

[12] N Thomas-Stonell, A-L Kotler, HA Lepper, PC Doyle, "Computerised speech recognition: influence of intelligibility and perceptual consistency on recognition accuracy", *Augmentative & Alternative Communication*, v14, 51-56, Mar 1998

[13] B Blaney, J Wilson, "Acoustic variability in dysarthria and computer speech recognition", C*linical linguistic & phonetic*, ISSN 0269-9206, 14(40), 307-327, 2000

[14] LJ Ferrier, N Jarrell, T Carpenter, "A case study of a dysarthric speaker using Dragon Dictate voice recognition software", *Jour comp users in speech & Hearing*, v8, 33-52, 1991

[15] LJ Ferrier, HC Shane, HF Ballard, T Carpenter, A Benoit, "Dysarthric speakers intelligibility and speech characteristics in relation to computer speech recognition", *Augmentative & Alternative Communication*, v11, 165-174, 1975

[16] PC Doyle, HA Lepper, A Kotler, N Thomas-Stonell, C Oneil, M Dylke, K Rolls, "Dysarthric speech: a comparison of computerised speech recognition and listener intelligibility", *Jour rehabilitation research & dev.,* 34(3), 309-316, Jul1997

[17] N Thomas-Stonell, A Kotler, HA Lepper, PC Doyle, "Computerised speech recognition: influence of intelligibility and perceptual consistency on recognition accuracy", *Augmentative & Alternative Communication*, v14, 51-56, Mar 1998

[18] A Kottler, N Thomas-Stonell, "Effects of speech training on the accuracy of speech recognition for an individual with speech impairment", *Augmentative & Alternative Communication*, v 13(2), 71-80, June 1997

[19] KM Yorkston, DR Beukelman, C Traynor "Computerised assessment of intelligibility of dysarthric speech (CAIDS)", *ISBN 0-88120-2215, CC Pubs, PO 23699 Tigard Oregan USA, 1984*

[20] RD Kent, G Weismer, JF Kent, JC Rosenbek, "Toward phonetic intelligibility testing in dysarthria", *Jour speech & hearing disorders*, 54, 482-499, 1989

[21] PM Enderby, "Frenchay Dysarthria Assessment", ISBN 0-933014-82-1

[22] S Robertson, B Tanner, F Young "Dysarthria Sourcebook" ISBN 0863880711