Performance of Speaker-independent Speech Recognisers for Automatic Recognition of Australian English

Kollengode S. Ananthakrishnan, Ahmad Hashemi-Sakhtsari, Adam Barnes, Serge Bailes
School of Electrical and Information Engineering, University of South Australia, Mawson Lakes
krishnan.ananthakrishnan@unisa.edu.au
ahmad.hashemi-sakhtsari@unisa.edu.au

Abstract

This paper investigates the performance of three speaker-independent speech recognisers (SISRs) that support continuous speech and are currently available for speaker-independent recognition of English. These speech recognisers were tested using a subset of the Australian National Database Of Spoken Language (ANDOSL) for the recognition of digits, sentences and a short paragraph under the "clean" data condition. The recognition accuracy was evaluated by two sets of tests, one set using a limited number of speakers (i.e. 13 speakers) and the other set using 75 of the speakers in the ANDOSL database. This paper describes the preparation taken to compare the recognisers operating under similar conditions. The paper does not reveal the name of the speech recognisers because the study reported here was limited. Nonetheless, the aim of the evaluation study was to ascertain which of the recognisers would be most suitable for applications in a meeting room where a sub-set of the vocabulary is selected as the active vocabulary for control of devices and for issuing commands. The variables investigated in this study were form of speech, vocabulary size, age and gender of the speaker. The contribution of this paper to speech science and technology is in its comparative evaluation of three speaker-independent speech recognisers., and in its use of a database containing utterances with Australian English accent.

1. Introduction

In general, 'speech recognition' refers to the recognition of an uttered number (digit), a word or a phrase occurring in spoken sentences or paragraphs of any natural language. The field of Automatic Speech Recognition (ASR) has seen significant growth in the last decade, mainly due to collaborative research between academia and industries through sponsored projects, resulting in the development of *speaker-dependent* and *speaker-independent* speech recognisers. The majority of these speech recognisers have been mainly developed for the recognition of the English language. Speech recognisers for European and Asian languages are under development, but at the infancy level.

Currently, very few speech recognisers of the English language are available worldwide to the speech research community for research into their integration with various speech technology applications such as Voice Activated Domestic Appliance Systems (VADAS) for the disabled population (Noyes et al., 1989), systems for helping people with dysarthria to control assistive technology by spoken commands (Hawley et al., 2003), automatic transcription of meetings (Sladek et al., 2002, Graham and Sladek, 2004), and command and control situations (Damper et al., 1996) and in Defence applications (Weinstein, 1994).

The existing recognisers for the English language can be classified as *speaker-independent speech recognisers (SISR)* and *speaker-dependent speech recognisers (SDSR).* Speaker-independent speech recognisers do not rely on the individual users of the software to train them with their voice, before they can be used for ASR applications. This makes speaker-independent speech recognition software an attractive choice for real-world applications, as these require minimal user time in training the software. On the other hand, speaker-dependent speech recognisers which need to be trained by each individual user before their integration with any application make them *less flexible.* However, these recognisers support a much *larger vocabulary* and have generally been found to be *more accurate.*

Performance evaluation of speech recognisers is a major area of interest in the field of Speech Processing and the main reason for this interest stems from the intrinsic difficulty of the task itself (Riccio et al., 1993). Furthermore, any performance measurement is meaningless unless the range of variables and experimental conditions under which the results were achieved are clearly defined. A pioneering paper (Lea, 1982) identified over 80 factors that affect the performance of a speech recogniser. An excellent insight into the performance assessment of speech recognisers was provided two decades ago by Pallett (1985). The performance of speech recognisers can vary significantly in terms of their ability to cope with various parameters, notably background noise (Littlefield and Hashemi-Sakhtsari, 2002). More recently Agaiby et al. (1997) investigated the effects of five variables, namely background noise, dialect, speaker vocal characteristics, microphone type and loudness levels that are amongst critical factors affecting the performance of the target application. The problem of realistic and reliable assessment of ASR for telephone applications has also been addressed (Riccio et al., 1993). In this work, however, the performance of the recognisers under investigation was evaluated for noiseless pre-digitised samples of native Australian speakers.

Proceedings of the 11th Australian International Conference on Speech Science & Technology, ed. Paul Warren & Catherine I. Watson. ISBN 0 9581946 2 9
University of Auckland, New Zealand. December 6-8, 2006.

Accepted after abstract only review

This paper discusses the preliminary results of experiments conducted into three speaker-independent speech recognition software packages. The aim of these experiments was to evaluate the performance of both commercially available and university-associated speaker-independent recognisers for their suitability for applications requiring recognition of Australian English. As only one of the available recognisers had acoustic models specifically for the Australian English, all the recognisers were evaluated for recognition using American English acoustic models. It should however be borne in mind that differences exist between accents of American and Australian English. This is related to first and second formant frequencies in the speech signal (Yan, Q. et al., 2003). The empirical experience of the authors on the use of both American and Australian English acoustic models for a commercially available speaker-dependent speech recogniser, however, has not revealed significant effects on the performance of the recogniser.

This paper is organised as follows: In Section 2, background about each recogniser is given in addition to specifying the software requirements for their operation. This section also briefly describes the acoustic and language models used in these three recognisers. Section 3 gives details of various tests conducted on the three recognisers and includes a discussion on the quantitative and qualitative analysis of the results obtained. Section 4 provides a summary of the outcomes of this paper and concludes with future work needed for the continuing research.

2. Background and requirements of speaker-independent recognisers

This section briefly describes background on the three speaker-independent speech recognisers which were evaluated in this work. To avoid any partial judgments about any of the recognisers, these popular recognisers of the English language were classified into two categories, namely, commercially available recogniser (CAR) and University-associated recognisers (UARs). In this paper, the term CAR is referred to as 'Recogniser A' and the remaining two recognisers under UAR category are referred to as Recogniser B and C respectively.

2.1. Commercially available recogniser (CAR)

Recogniser A has been developed by a leading company engaged in developing automated telephone transactions using voice recognition (e.g. booking a flight). This software package is commercially available under different licenses, including a licence for commercial use, and a developer's licence for research and development.

2.2. University-associated recognisers (UARs)

Recogniser B was developed and released by a consortium of private companies and universities. This recogniser has undergone several version changes with its speed and flexibility having been improved in the recent version. This package is written entirely in Java and requires Java 2 SDK, and Apache Ant installed on the same computer running the Recogniser B. This recogniser has two modes of operation, namely *live mode* and *batch mode.* These modes can be selected via a configuration file. This recogniser is freely available on-line in both source and binary distribution, with the binary version being used for our evaluation purposes.

Recogniser C was developed by a different University, and in this work, the latest test version of this recogniser was employed. This recogniser has a trial version of binary executables that are available freely on-line for non-profit organisations and academic groups. A source code version of Recogniser C is available for a licence fee. The latest binary version of this recogniser was used for our testing. The binary version comes with binary executables that allow the researcher to build and run recognition packages for Recogniser C. The executables include programs that can run *live* and *batch mode* recognition, create acoustic models, and convert spoken words to their phoneme set, given the right training data and commands.

To run Recogniser C on a Windows platform, a C++ compiler is required to be installed on the machine. Furthermore, to run Recogniser C's *live mode* recognition client, Tcl/Tk 8.3 or higher, and Snack version 2.1 or greater is also required to be installed on the computer. Recogniser C's *live mode* application runs in a client-server set up. A configuration file is used to recognise the server. The client then connects to the server and configuration for the client is then selected. Once this is done, the client will send any audio input to the server which will perform recognition and send results back to the client, which then displays the recognised words.

The vocabulary sets that were used to evaluate the performance of the speech recognisers were 1-9 digits, sentences and paragraphs. The variables chosen in evaluating the selected recognisers were form of speech i.e. isolated versus connected or continuous, vocabulary size, as well as age and gender of the speaker.

2.3. Language and acoustic Models of SISRs

A *language model* contains a list of words (dictionary) and gives a probability for each word in the list appearing in the training data used.

An N-gram language model, where N can be uni, bi, tri etc, contains the probability of individual words appearing in speech as well as the probability of a word sequence containing N words appearing. Recognisers A and B support both bigram and trigram Language Models. Recogniser C supports unigram, bigram, trigram and quadgram Language Models.

N.B. All language models were set up as tri-gram language models and had a vocabulary size of 2000 words.

3. Testing strategy

3.1. Testing preparation

In order to conduct fair comparative evaluation studies, the speech recognisers were set-up to work under as similar operating conditions as possible. In particular attention was paid to uniformity in selecting acoustic and language models. Their performance in terms of percentage recognition accuracy was investigated using five different tests. All experiments used pre-digitised audio files from the ANDOSL database in '.wav' format. These tests were designed to observe which recogniser achieved the highest recognition accuracy. Four trials were conducted for every test for consistency and, as

Proceedings of the 11th Australian International Conference on Speech Science & Technology, ed. Paul Warren & Catherine I. Watson. ISBN 0 9581946 2 9
University of Auckland, New Zealand. December 6-8, 2006.

Accepted after abstract only review

expected in playing back the same recorded data, only minor variation was observed in the recognition accuracies among trials. The average of four trials was taken as the *recognition accuracy* (percentage of spoken words correctly recognised by the recognisers and as reported by the American National Institute of Standards and Technology's SClite package). SClite is a program that has been designed for the purpose of scoring in speech recognition programs. Given a reference text and the hypothesis (recognised) text, SClite will compare the two and produces a report with statistics describing the number of correct and incorrect words and provides the types of errors encountered in the form of insertion, rejection and substitution errors.

3.2. Speech database and testing methodology

The research carried out in this paper used the ANDOSL database. This database is available in nine CDs under a licence from the Australian Speech Science and Technology Association (ASSTA). Several tests were performed to determine the recognition accuracy of selected speaker-independent speech recognisers with all the tests employing only a subset of this large database. The strategy of testing consisted of two parts. The main focus in the first phase of this project was to evaluate the behaviour of these recognisers only for "clean" data with no noise added. Based on the results of recognition accuracies for noise-free data, the second phase was planned thereafter to investigate these recognisers for noisy data. This paper reports the preliminary results obtained in the first phase of the project only since these were completed in time for reporting.

3.3. Experimental results and discussion

In this work, *two sets* of experiments were carried out on three recognisers. The number of speakers in the first set was 13, whereas for the second set, the number of speakers was increased to 75. The age group of speakers varied from 18-46+ years.

In the first set, two tests were carried out. The first was to determine the recognition accuracies (% of correct words recognised) of *digits 1-9* in ascending, descending & random orders. This was to avoid *rhythmic Patterning* that affects pronunciation of words. The second test was to determine the accuracies for *sentences*. In the sentences test, ten sentences were chosen from a list of 200 phonetically rich sentences from the ANDOSL database. These sentences from 12 speakers (9 males, 3 females) were used as input to only recogniser A. Limitations in time and human resources did not extend the testing to the other two recognisers.

In the second set of experiments, three tests were carried out. The first one was conducted to study the behaviour of all three recognisers for 10 sentences from 75 speakers, limiting the size of the vocabulary to only 148 words. In the second test, to study the effect of the vocabulary size on the recognition accuracy, the size of the grammar file was increased to 2000 words, whilst keeping the number of speakers constant. Finally, a short paragraph test was included to study the performance of all three recognisers with continuous speech. The details of the test are as follows:

3.3.1. First Set-Test 1: Digits 1-9 in ascending, descending and random order:

Table 1 shows the recognition accuracy results of the three recognisers for 13 speakers (9 males, 4 females). This table also gives the age group of the speakers. Each entry in the table is the average value of 4 independent repeated trials for every speaker. For a small set of speakers, the balance in the age group was achieved by selecting nearly equal numbers of speakers from each category (4–Young; 4-Matured; 5-Elderly).

Of the three speaker-independent speech recognisers studied and evaluated, commercially available recognizer (CAR) scored 96.55% for digits 1-9 in *ascending order*, 82.91% for digits 1-9 in *descending order* and 74.38% for digits in *random order*. Around 90% of all errors were caused by *digit 4* being substituted with "for" and *digit 2* being substituted with "to". The recognisers categorised as UARs scored 57 % for *digits 1-9* in *ascending order*, 46-55% for *digits 1-9* in *descending order*, 51%-53% for *digits 1-9* in *random order,* for the same set of speakers.

3.3.2. First Set - Test 2: Ten Sentences from 12 speakers

This test was performed for all 10 sentences (S001-S010) for 12 speakers (9 males, 3 females) for Recogniser A. Table 2 shows the recognition accuracies for the speakers averaged over 4 trials for this recogniser only for one sentence S001, which produced the maximum accuracy of 39.83%. The figures obtained for other sentences (S002-S010) were below this value and this recogniser produced an overall average accuracy of only 19.77 % for all sentences by all speakers. Sentences from the ANDOSL database which were used as input to these three recognisers are listed below in the order of ANDOSL sentence ID, along with the sentence itself and the number of words in each sentence within parenthesis.

S001: *The price range is smaller than any of us expected (10 words)*
S002: *They asked if I wanted to come along on the barge trip (12 words)*
S003: *Amongst her friends she was considered beautiful (7 words)*
S004: *The smell of the freshly ground coffee never fails to entice me into the shop (15 words)*
S005: *I'm often perplexed by rapid advances in state of the art technology (12 words)*
S006: *John could lend him the latest draft of his work (10 words)*
S007: *From forty love the score was now deuce and the crowd grew tense (13 words)*
S008: *The Presbyterian minister managed to curb the drinking habits of the loitering youths (13 words)*
S009: *The bulb blew when he switched on the light (9 words)*
S010: *It is futile to offer any further resistance (8 words)*

Proceedings of the 11th Australian International Conference on Speech Science & Technology, ed. Paul Warren & Catherine I. Watson. ISBN 0 9581946 2 9
University of Auckland, New Zealand. December 6-8, 2006.

Accepted after abstract only review

Table 1: Recognition accuracy for Digits (1-9)

			Recognition Accuracy (%)								
Spkr-Id/ ANDOSL	**Gender**	**Age Group#**	**CAR**			**UARs**					
			Recogniser A			**Recogniser B**			**Recogniser C**		
			Asc	**Des**	**Ran**	**Asc**	**Des**	**Ran**	**Asc**	**Des**	**Ran**
s015	Male	E	93.20	84.40	77.80	57.80	66.70	42.20	63.10	52.50	37.80
s045	Male	M	90.90	77.80	77.80	57.80	44.40	57.80	60.80	66.40	59.00
s075	Female	Y	100.00	86.70	77.80	57.80	55.60	62.20	61.00	68.20	67.90
s088	Female	M	100.00	88.90	71.10	55.60	66.70	51.10	53.20	62.60	49.40
s094	Male	M	97.80	88.90	77.80	51.10	33.30	53.30	51.30	47.60	56.90
s099	Female	E	100.00	88.90	77.80	66.70	44.40	46.70	67.40	58.70	52.40
s109	Female	E	100.00	82.20	66.70	46.70	44.40	44.40	42.40	47.10	30.30
s110	Male	E	73.30	51.10	60.00	55.60	22.20	55.60	55.70	46.70	44.10
s111	Male	Y	100.00	77.80	68.90	75.60	22.20	68.90	78.50	68.70	70.40
s115	Male	Y	100.00	88.90	77.80	53.30	44.40	55.60	46.80	43.00	42.80
s119	Male	M	100.00	88.90	77.80	40.00	33.30	51.10	41.80	43.80	64.80
s128	Male	Y	100.00	84.40	77.80	64.40	66.70	44.40	59.70	57.40	37.90
s134	Male	E	100.00	88.90	77.80	64.40	55.60	51.10	62.10	62.60	55.80
% correct words			**96.55**	**82.91**	**74.38**	**57.45**	**46.15**	**52.65**	**57.22**	**55.79**	**51.50**

#Young: (18-30)-Y; Matured: (31-45)-M; Elderly: (46+)-E; Asc-Ascending; Des-Descending; Ran-random.

Table 2: Recognition Accuracy (%) for 12 speakers-Single Sentence (S001)

Speaker Id/ ANDOSL	**Gender**	**Age# Group**	**CAR**
			Recogniser A
s015	Male	E	54.00
s045	Male	M	30.00
s075	Female	Y	40.00
s088	Female	M	28.00
s094	Male	M	42.00
s109	Female	E	34.00
s110	Male	E	46.00
s111	Male	Y	40.00
s115	Male	Y	32.00
s119	Male	M	56.00
s128	Male	Y	34.00
s134	Male	E	42.00
correct Words (%)			**39.83**

3.3.3. Second Set-Test 1: Recognition Accuracy for 75 speakers-10 sentences for a limited vocabulary (148 words)

Table 3 summarises the recognition accuracy for 10 sentences from 75 speakers. The number of speakers for this test represents nearly 70% of the native Australian speakers from ANDOSL database (75/108 speakers). The total number of speakers selected for the *sentence test* includes a gender balance of *37 male speakers* and *38 female speakers*. Speakers were selected based on their *age* and *gender*. The number of speakers chosen for each gender was reasonably balanced. This would be a representative sample of the population. The male category included 13 speakers from '*General'* category, 12 speakers from *'Broad'* category and 12 speakers from *'Cultivated'* category. The diversity in age-group was realized by selecting 12 Young, 15 Matured and 10 Elderly speakers.

Table 3: Recognition accuracy for 10 sentences (S001-S010-ANDOSL) -75 speakers

Sentence Id ANDOSL	**Recognition Accuracy (%)**		
	CAR	**UARs**	
	A	**B**	**C**
S001	73.90	93.40	85.30
S002	68.90	90.20	83.20
S003	76.40	96.20	82.30
S004	77.50	92.60	93.30
S005	65.00	92.20	26.00
S006	76.00	90.90	83.10
S007	80.00	91.30	83.10
S008	80.00	90.30	91.00
S009	74.10	77.10	88.40
S010	72.30	91.10	15.90
% correct words (Ave)	**74.41**	**90.53**	**73.16**

In the case of female speakers, the distribution was 18 from *'General'*, 7 from *'Broad'* and 13 from *'Cultivated'*. The results shown in the table are for 4 trials (i.e. each sentence was repeated 4 times for every speaker). In this category, the age-group distribution was 15 Young, 12 Mature and 11 Elderly.

In this test, the size of the vocabulary *was limited to 148 words* and this test was designed to determine which recogniser gives the highest recognition accuracy for all 10 sentences of the limited vocabulary for 75 speakers. This subset of vocabulary was chosen from a larger data set containing 2000 words. As can be seen from Table 3, for all

Proceedings of the 11th Australian International Conference on Speech Science & Technology, ed. Paul Warren & Catherine I. Watson. ISBN 0 9581946 2 9
University of Auckland, New Zealand. December 6-8, 2006. Copyright, Australian Speech Science & Technology Association Inc.

Accepted after abstract only review

sentences, Recogniser B gave the highest accuracy (90.53%). The maximum accuracy of 96.20% was obtained for sentence S003. The lowest recognition accuracy of 15.90% was achieved by the Recogniser C for the sentence S010. The performance of Recogniser C was better than Recogniser A for all sentences except for sentence S005.

3.3.4. Second- Set-Test 2: Comparison of recognition accuracies of three recognisers for limited (148) and large (2000) vocabularies

This test was carried out to study the effect of the *size of vocabularies* on the recognition accuracy of a recogniser. At the time of writing, only testing of Recogniser A has been completed. Table 4 gives the comparison of the recognition accuracies for a large vocabulary consisting of 2000 words and a limited vocabulary of only 148 words for 75 speakers. In testing the limited vocaburay, the 10 sentences were fed as input to this recogniser. The 148-word vocabulary consisted of all words appearing in the test sentences, and only those words; the 2000-word vocabulary was a superset of the 148-word vocabulary chosen randomly from the dictionary in the ANDOSL database. In the latter case, although the same 10 sentences were chosen as input to the recogniser, it had to choose from many more words in the dictionary over and above those contained in the limited sub-set. Hence, the recogniser was attempting to match the words in the input utterances from potentially 2000 words, which contains many more similar sounding words.

Table 4: Comparison of recognition accuracy (%) for Limited Vs Large Vocabulary Grammar files (75speakers-10 sentences) Recogniser A (CAR)

Sentence Id-ANDOSL	**Recognition Accuracy (%) Limited Vocab**	**Recognition Accuracy (%) Large Vocab**
S001	73.90	41.70
S002	68.90	43.10
S003	76.40	58.50
S004	77.50	45.40
S005	65.00	39.30
S006	76.00	47.50
S007	80.00	50.30
S008	80.00	49.30
S009	74.10	43.60
S010	72.30	54.70
Average % correct words	74.41	47.34

As predicted and demonstrated in the results of Table 4, the performance for the larger vocabulary was poor, when compared with the limited vocabulary size. This is clearly attributed to the number of words in the dictionary and the grammar files, where the recogniser was attempting to match with the test pattern.

3.3.5. Second Set - Test 3: a short paragraph (29 words) test.

The following paragraph consisting of 29 words was used as input to three recognisers A, B and C in order to evaluate the effect of continuous speech on the percentage recognition accuracy.

"He jerked round in an instant to face his assailant. He emphasised his strengths while concealing his weaknesses. They launched into battle with all the forces they could muster"..

The average recognition accuracy achieved for 12 speakers is given in Table 5. The results shown in the table are averaged over 4 trials. Recogniser B belonging to UAR category performed well in comparison with the Recogniser A. The recognition accuracy of Recogniser C was far below that of the other two recognisers, indicating that possibly the language models implemented need further enhancement.

Table 5: Recognition Accuracy for short paragraph (limited vocabulary)

Recogniser A	Recogniser B	Recogniser C
79.03 %	83.5%	54.20%

4. Discussion and Conclusions

The primary variables under investigation in this study were form of speech, i.e digits, sentences and a paragraph in order to test sound structures like vowels and consonants.

The performance of three speaker-independent speech recognisers (SISRs) was evaluated for the recognition of *digits (numbers 1-9)*, *sentences* and *a short paragraph* under "'clean" data conditions using a subset of the ANDOSL database for Australian English. For *digits,* the recognition accuracy of CAR was better than that of UARs for a smaller data set, regardless of the order in which the numbers were fed to the recogniser.

In the case of Sentences test for limited vocabulary size, the recognition accuracy of Recogniser B (UAR) was better (90.53%) than CAR (74.41%) for larger datasets i.e. with larger number of speakers. However, the performance of Recogniser A and C were comparable. It is expected to observe the effect of the vocabulary size on the recognition accuracy of recogniser A, with the performance diminishing significantly when the size was approximately 13 times larger. This is due to the recogniser having to match each word in an input utterance to many more candidate words, with many more similar sounding words..

For short paragraphs, although the performance of Recogniser B appears to be better than the other two

Proceedings of the 11th Australian International Conference on Speech Science & Technology, ed. Paul Warren & Catherine I. Watson. ISBN 0 9581946 2 9
University of Auckland, New Zealand. December 6-8, 2006.

Accepted after abstract only review

recognisers, no definite conclusion can be drawn without conducting further research.This will require paragraphs with utterances longer than one minute duration to clearly observe the robustness of language models for the recognisers.

The gender of speakers was listed since pitch and vocal tract length are gender-specific in adults.

The age of speaker was listed since voice quality varies at different life stages.

Speaker-independent speech recognisers currently available on the market are far from being adequate to perform accurate recognition with large sentences and long paragraphs. However, our results have shown that for numbers,, simple commands and small phrases, the accuracy of Recogniser A and B are moderately high and if configured correctly, these recognisers will be useful for applications requiring a small well-defined vocabularies, such as those that may be used for control of devices in an intelligent room.

Future research work, in the second phase of the project, will focus on investigating the inadequacies of these recognisers and developing additional language models as the front-end to enhance the recognition accuracy for digits, words in sentences or phrases and paragraphs of Australian English under adverse (noisy) conditions.

5. Acknowledgements

The authors would like to thank Prof. Robin King, Pro-Vice Chancellor, DITEE, University of South Australia (UniSA) for providing the ANDOSL database and extending his support for collaborative research between UniSA and the Defence Science and Technology Organisation (DSTO), Adelaide. Command Decision Environments in Command and Control Division of DSTO allocated the funds for this research through the research agreement, 2006/1039612.

6. References

Agaiby, H., Fyfe, C., McGlinchey, S. and Moir, T.J. (1997). *Commercial speech recognisers performance under adverse conditions, a survey*. In: Robust speech recognition using unknown communication channels, ESCA-NATO Tutorial and Research Workshop, 163-166.

Damper, R.I., Tranchant, M.A. and Lewis, S.M. (1996) Speech versus keying in command and control: effect of concurrent tasking. *International Journal of Human-Computer Studies*, 45 (3) 337-348.

Graham, S. & Sladek, J. (2004) *An Automatic Transcriber of Meetings Utilising Speech recognition Technology, DSTO Client Report (DSTO-CR-0355)*. Command and Control Division, Information Sciences Laboratory, DSTO Edinburgh.

Hawley, M., Enderby, P., Green, P., Brownsell, S., Hatzis A., Parker, M., Carmichael, J., Cunningham, M., ONeill, P. and Palmer, R (2003). STARDUST*;* Speech Training And Recognition for Dysarthric Users of Assistive Technology*, Association for the Advancement of Assistive Technology in Europe* (AAATE), 959-964.

Johnston, R. D. (1996). Are Speech Recognisers Still 9% Accurate, o ha The Time Come to Repeal 'Hyde's Law'?. *BT Technical Journal,* 14 (1), 165-17.

Lea, W.A. (1982) What causes speech recognizers to make mistakes? In: *Proc. ICASSP,* 2030-2033.

Littlefield, J., and Hashemi-Sakhtsari, A. (2002). *The Effects of Background Noise on the Performance of an Automatic Speech Recogniser.* DSTO-RR-0248.

Noyes, J.M., Haigh, R., and Starr, A.F. (1989) Automatic speech recognition for disabled people. *Appl. Ergon,* 20 (4),293-298.

Pallett, D.S. (1985) Performance assessment of automatic speech recognisers. *Joumal of Research of the National Bureau of Standards* (USA), 90(5), 1-17.

Riccio, A., and Ceglie, F. (1993) Automatic Performance assessment of speech recognisers through telephone quality speech databases. *IEEE Conference on Telecommunications,*.23–27.

Sladek, J., Zschorn, A. & Hashemi-Sakhtsari, A. (2002), Speech-to-Text Transcription in Support of Pervasive Computing, in B. Ainsley and B. Thomas (Eds.) *Proc. of Asia Pacific Forum on Pervasive Computing 2002*, Vol. 25. Adelaide: Australian Computer Society.

Weinstein, C.J. (1994) Military and government applications of human-machine communication by voice. In David B Roe and Jay G Wilpon (Eds.) *Voice communication between human and machines*. Washington D.C.: National Academy Press

Yan Q., and Vaseghi S. (2003). Analysis, Modelling and Synthesis of formants of British, American and Australian accents, *ICASSP*, I-712-715.

Accepted after abstract only review