

Speech Recognition Technologies

Author: John Kirriemuir
Email: john@ceangal.com
Web site: <http://www.ceangal.com/>

Last updated: March 30th 2003

Abstract

This report presents an overview of speech recognition technology, software, development and applications. It begins with a description of how such systems work, and the level of accuracy that can be expected. Applications of speech recognition technology in education and beyond are then explored. A brief comparison of the most common systems is presented, as well as notes on the main centres of speech recognition research in the UK educational sector. The report concludes with potential uses of speech recognition in education, probable main uses of the technology in the future, and a selection of key web-based resources.

1. We all speak differently...

For the past 40 years, speech recognition systems have failed to match the predictions and aspirations disseminated through the mass media:

HAL¹: Hello, Frank, can I have a word with you?

POOLE: Yes, Hal, what's up?

HAL: It looks like we have another bad A.O. unit. My FPC shows another impending failure.

POOLE: Gee, that's strange, Hal. We checked the other unit and couldn't find anything wrong with it.

HAL: I know you did, Frank, but I assure you there was an impending failure.

The popular image of speech recognition has been of conversing normally with subservient robots, computers, your house, and other machines.

¹ <http://www.palantir.net/2001/script.html> - excerpt from the script of "2001: A Space Odyssey", where HAL - the computer - converses freely in natural, unrestricted language with the astronauts (shortly before killing most of them).

The reality has been somewhat different, for one main reason; people speak differently:

- across the world there are several thousand active languages
- within those languages, there are wide variations in the syntax and vocabulary used
- people in different regions speak the same language but with different accents
- people speak at different speeds, tones and pitches
- colloquialisms are used; phrases and words can have several meanings

As a crude example, imagine how differently an aristocrat from Surrey, and a docker from Newcastle, would order their favourite drink in a bar.

Consequently, most of the history of speech recognition systems has been in making a trade-off between what the user can say or speak, and what the technology interpret, that is of an acceptably high level of accuracy to the end-user.

The context of speech recognition in this report refers to systems where a person can "speak", via a microphone, to a computer. The computer translates the spoken words into either text (usually for automatic insertion, secretary-style, into items such as word processing documents) or commands to execute functions in the computer e.g. print a document, open a file.

2. How conventional speech recognition systems work

Speech recognition is an alternative to traditional methods of interacting with a computer, such as textual input through a keyboard. An effective system can replace, or reduce the reliability on, standard keyboard and mouse input. This can especially assist the following:

- people who have little keyboard skills or experience, who are slow typists, or do not have the time or resources to develop keyboard skills.
- dyslexic people, or others who have problems with character or word use and manipulation in a textual form.
- people with physical disabilities that affect either their data entry, or ability to read (and therefore check) what they have entered.

A speech recognition system consists of the following:

- a microphone, for the person to speak into.
- speech recognition software.
- a computer to take and interpret the speech.
- a good quality soundcard for input and/or output.

Speech recognition systems used by the general public e.g. phone-based automated timetable information, or ticketing purchasing, can be used immediately – the user makes contact with the system, and speaks in response to commands and questions. However, systems on computers meant for more individual use, such as for personal word processing, usually require a degree of "training" before use. Here, an individual user "trains" the system to understand words or word fragments (see section 2.3); this training is often referred to as "enrolment".

At the heart of the software is the translation part. Most speech recognition software breaks down the spoken words into phonemes, the basic sounds from which syllables and words are built up. These are analysed to see which string of these units best

"fits" an acceptable phoneme string or structure that the software can derive from its dictionary².

It is a common misassumption that such a system can just be used "out of the box" for work purposes. The system has to be trained to recognise factors associated with the user's voice e.g. speed, pitch. Even after this training, the user often has to speak in a clear and partially modified manner in order for his or her spoken words to be both recognised and correctly translated.

Most speech recognition software is configured or designed to be used on a stand-alone computer. However, it is possible to configure some software in order to be used over a network³.

2.2 Discrete and continuous systems

Until the late 1990's, commercial speech recognition systems were "discrete" in nature. This required the user to speak in an unnatural, command-like manner, with a gap between each word (to allow time for the computer to process the input, and to make it clear to the computer when individual words ended).

Not surprisingly, these speech recognition systems were more oriented towards command-based applications. Entering and editing a large document, such as a book chapter, was very time consuming.

The rapid increase in desktop computer processing power and better algorithms led to the development of "continuous" systems. These allowed the user to speak at near-normal speed, and still (after sufficient training) obtain a high accuracy rate. Contemporary continuous systems can give an accuracy of 90%-95%, depending on training (though this still means that one in ten or twenty words can be expected to be incorrect).

2.3 Training the system

The amount of enrolment required depends on the software used, the processing power of the computer, and the desired final accuracy of the system. Older versions of the software/computer can require a considerable amount of training; however, enrolment times have fallen drastically (partially due to the rapid increase in standard processing power) in recent years.

For example, an observer noted that one package from 1999 *"requires a demanding enrolment procedure when a slower computer is being used. It requires the user to read up to 100 paragraphs of an adult level text such as 'Alice in Wonderland' or '3001: The Final Odyssey'."*

However, the next year, an upgraded version of the package *"may take less than 10 minutes, after installation, to set up and enrol...Nevertheless, achieving good results in a short time demands a computer with a suitably high specification, e.g. Pentium*

² <http://www.generation5.org/aisolutions/howrworks.shtml> - a brief but readable overview, from a technical perspective, on how a speech recognition system works.

³ <http://www.becta.org.uk/technology/speechrecog/docs/kirkhallam.pdf> - a report from Kirkhallam community school on the use of speech recognition software, which details the problems encountered in configuring the software to work over a network.

II or Pentium III with at least 64 Megabytes of RAM (128 Mb is better), a good quality soundcard and microphone."⁴ Some speech recognition software manufacturers claim an enrolment time of as short as 5 minutes⁵, though in practise this may adversely affect the accuracy of the system in use.

Most speech recognition systems allow the user to undertake further enrolment procedures, or retraining of specific misidentified words, if needed. Contemporary systems also allow you to train against your own text or documents, therefore creating your own specialised "vocabulary" for the system to recognise. Multiple language support is now commonplace, with systems such as SpeechPearl supporting over 40 languages⁶.

2.4 Reducing extraneous factors

There are a number of methods for increasing the accuracy and ease of use of speech recognition systems

1. using a high-performance computer. If using contemporary speech recognition software, the computer will usually need to contain a fast processor and a large amount of RAM in order to work efficiently. Though software packaging often states that it will run on "64MB of RAM", this has often found to be inadequate, resulting in a much longer training time⁷. 256MB of RAM minimum is more preferable, which can cause problems in schools, colleges and universities that use older computers.
2. using a good quality microphone. Microphones with "Active Noise Reduction" or "Active Noise Cancellation" can reduce the amount of background noise that can "confuse" the software.
3. installing a good sound card. Soundcards that come installed in computers are often of variable quality. For serious use of speech recognition systems, high quality duplex (input and output) sound cards are recommended by a number of projects and speech recognition websites.
4. working in a sound-free environment. Most contemporary speech recognition software can eliminate background noise to a certain extent. However, educational settings, where there can be 20 or 30 people in the same classroom or laboratory are obviously not ideal. Setting up the computer in a place where there is a minimum of background noise, such as in a separate room or a cubicle, can reduce instances where the software picks up and translates other speech.
5. using the right operating system. Some speech recognition systems are designed to run on only one, or a narrow set, of operating systems⁸. This can cause problems in educational establishments with a mixture of operating systems, and can restrict opportunities for the pupil or student to use a system they are familiar with and have "trained" at home on their own computer.

⁴ <http://www.worc.ac.uk/services/egalopps/VRProject/voicerec.htm> - a review of the use of voice recognition software at University College Worcester.

⁵ <http://www.voicepower.co.uk/faq.htm#faq9> - recommended enrolment time for Dragon NaturallySpeaking v6.

⁶ <http://www.connectweb.co.uk/public/products/report/philips.html> - SpeechPearl press release.

⁷ <http://www.ace-centre.org.uk/download/VoiceRec.doc> - Voice Recognition - Getting Specific. Section 12 outlines the problems of using speech recognition systems on PCs with a small amount of RAM.

⁸ <http://www.becta.org.uk/technology/speechrecog/docs/kingsbfinal.pdf> - in this particular case, a speech recognition system was found (after some delay) to work under Windows 95 but not Windows 98, resulting in the school having to use/install an *earlier* version of an operating system on its PCs.

3. Accuracy

It is notoriously difficult to measure the accuracy of speech recognition systems, as there are so many technical and human factors involved. Several experiments have attempted to compare speech recognition with other kinds of data entry, such as mouse, keyboard and handwriting recognition. Many of these experiments do not reach an overall conclusion concerning "which system is better"⁹.

One of the problems with the wider take-up of speech recognition is that the level of accuracy attained by a user does not match that stated on the software packaging¹⁰. This can be for all manner of reasons, such as the machine specification being insufficient, or (more often) the level of training undertaken by the reader.

Independent reviews of speech recognition systems indicate a score of around 95% accuracy being possible with an increasing number of systems. For example, in tests¹¹ involving dictating a newspaper story, email message and business letter, Dragon NaturallySpeaking 6.0 scored 95% accuracy, ViaVoice scored 92% accuracy and NaturallySpeaking 5.0 scored only 85% accuracy.

Speech recognition systems increasingly offer specialist vocabulary building systems. This step is particularly helpful when subject and user specific words and acronyms are likely to be used, such as specialist vocabulary from university subjects¹². Anecdotal evidence from various web sites points to a reduction in the *error rate*, by using specialist vocabulary, of usually around a third.

4. Contemporary uses of speech recognition

4.1 Speech recognition in the education sector

A literature review revealed that there has been more published on the use of speech recognition in UK schools than in colleges or universities over the last decade. Much of this use was as part of short-term or experimental investigative projects. Various comparative reports noted that speech recognition is but one of many methods of data entry for people with physical or learning difficulties. For example, Braille input/output devices, touch screen systems and trackballs have all been used successfully in the classroom¹³.

However, over the last few years, there has been a surge in interest from within HE and FE regarding the use of speech recognition technologies. This is due to:

⁹ <http://www.uclan.ac.uk/facs/destech/compute/staff/read/Publish/read.pdf> - *Measuring the usability of text input methods for children* - A comparison of speech recognition and three other methods of data entry. While the research revealed interesting observations about how the systems were used and reacted to input, the authors (as with several similar comparative papers in this field) shied away from an overall "this is best" conclusion.

¹⁰ <http://www.vhml.org/workshops/HF2002/papers/broughton/broughton.pdf> - *Measuring the accuracy of commercial automated speech recognition systems during conversational speech* - for a number of stated factors, the accuracy in the experiment did not match that quoted by the software.

¹¹ <http://reviews.zdnet.co.uk/review/43/2/1605.html> - review of Dragon NaturallySpeaking Preferred 6.0

¹² <http://www.uow.edu.au/student/services/uow/VRSoftware.html> - a university-authored guide to the training process involved in using the speech recognition software it provides.

¹³ http://www.ataccess.org/faq/E_technology.html - the Alliance for Technology Access has a Frequently Answered Questions page which describes some of these alternative access technologies.

- the increase in performance and accuracy of such systems
- a greater incorporation of students and learners with some disability into the educational sector
- increasing legal and more obligations to provide accessible¹⁴ ICT systems

For example, the University of Exeter has a disability policy, statement and dedicated resource centre. The University library has an "IT special needs zone"¹⁵, where students can use one of a number of speech recognition systems to write essays or carry out other computer-based work. Meanwhile, the University of Glasgow provides a range of disability-oriented ICT software to students, including a speech recognition software package¹⁶.

An informal survey of various university web sites indicated that an increasing number offer access to speech recognition software. However, in most cases this was offered not in a department or research group, or even in a laboratory, but in the university library or closely associated central IT centre. Details on what staff support the student would receive in either using the package, or the enrolment procedure, were infrequently found.

4.2 Speech recognition outwith the education sector

Possibly the most widespread application of speech technology in contemporary life has been its incorporation into telephone-based information retrieval systems. This is almost a natural development, as telephones take speech input (though in a passive, "passing it on" manner) anyway.

At the most basic level, some mobile phones offer the facility to select a phone number from the in-phone directory by saying the name associated with it e.g. "David", "Husband", "Mother"; the phone then dials the stored number automatically.

At a more useful level, speech recognition is increasingly used in automated telephone-based interactive services. For example, it is possible to check the weather forecast, the price of a stock market share, or book a flight using an increasing number of these services¹⁷. There are advantages to this for both the customer (no waiting for a human operator) and the service supplier (less staff required, can operate 24/7).

However, such systems still tend to heavily guide the customer through a range of options, as interpreting free and natural instructions is beyond the capabilities and knowledge banks of contemporary systems. In addition, security considerations are an impediment to development of some areas; for example, there are concerns over who can buy airline tickets, or transfer money between bank accounts, without human intervention or examination, using such a system. In view of this, there is a

¹⁴ http://www.jisc.ac.uk/index.cfm?name=topic_accessibility - the JISC Advice and Guidance card on accessibility provides further information on ICT accessibility requirements and relevant resources.

¹⁵ <http://www.ex.ac.uk/~SJMacwil/lib/zone.html> - University of Exeter library IT special needs zone.

¹⁶ <http://www.gla.ac.uk/services/tls/sls/specialneeds/software.htm> - accessibility software available to students at the University of Glasgow.

¹⁷ <http://www.molo.co.za/applications.htm> - brief details of a few automatic telephone-based information retrieval systems.

related area of research and development that is examining voice verification systems (see the Cave project, Appendix A.2).

Speech recognition technology is being tentatively used, and researched, in the car industry¹⁸. This is not surprising, as contemporary cars are heavily marketed according to technical innovations and features. Development is based in four areas:

1. hands-free use of mobile phone handsets in the car e.g. "Dial office"
2. speech instructions to navigation systems e.g. GPS-connected digital maps: "How far is it to the motorway junction?"
3. in-car system interaction e.g. "Turn on the radio to the travel reports channel."
4. in-car steering systems

The last of these, not surprisingly, is the one least developed due to obvious safety consideration. Though there are advantages for speech recognition in cars (such as being able to keep both hands on the steering wheel while telling the "car" to do something), there are considerable obstacles in terms of in-car noise, and vocal interference from passengers who are inches/feet away.

5. Contemporary speech recognition systems

There are a surprisingly large number of speech recognition systems (of different kinds) that are commercially available. The TechDis Accessibility Database (see appendix A2) contains details of 12 such products. The first two in the list below appear to be the most widely used in UK university and colleges.

5.1 Dragon NaturallySpeaking Preferred 6.0

Adequate training for this package was found by reviewers to take around 20 minutes, or 10 minutes with follow-up practice. With NS, the user can "speak" into any open window, such as a Word 2002 document. The software also allows the user to "surf the web hands-free" - in a web page, if the user says the first few words of a link then the browser will go to the linked page. However, several reviewers indicated that the software requires a computer with a considerable amount of processing power to produce excellent results.

As of January 2003, the package was available for roughly £124, with licence deals available for multiple user use. A search of various websites indicated that versions of this package were the most frequently available speech recognition software in UK universities.

5.2 IBM ViaVoice 10.0

Training for this system takes between 20 and 40 minutes (depending on reviewer). Version 10 is very new, and according to some reviewers gives accuracy scores of over 96%, making it currently possibly the best package in terms of accuracy. Speech input, still resulting in a high level of accuracy, is possible up to a rate of 160 words per minute. However, the software requires a considerable amount of

¹⁸ http://www.eng.nus.edu.sg/EResnews/0602/sf/sf_9.html - a few of the issues pertaining to the use of speech recognition in cars.

processing power to work very accurately, and over 0.5 of a gigabyte of hard drive space.

As of January 2003, the package was available for around £80.

5.3 Keystone SpeechMaster 5

This package is a combination of Dragon NaturallySpeaking and Keystone ScreenSpeaker. The added functionality provided by the latter provides various spelling aids and word recognition support, making it suitable for people with Dyslexia.

However, as of October 2002 the package costs £346, presenting a significant barrier to its use in school situations.

5.4 Microsoft Office XP (inbuilt)

Speech recognition is included within the latest version of the popular Office range of software. This allows the user to both enter commands, such as "open file", and to dictate text straight into an application. Added functionality enables the translation of text between a small number of languages.

A small portion (four paragraphs) of this report was entered using this system. This proved frustrating at first, as the author became very conscious (and distracted) by the process of entering content, as opposed to concentrating on the content itself> However, accuracy improved significantly from the first paragraph to the fourth.

6. Speech recognition research in the UK education sector

6.1 Areas of research

There are a number of research groups in the UK who are active in speech recognition, speech-to-text and/or text-to-speech research (though far less in the development of commercial products). However, most of these research groups focus more on the technical side of developing and researching speech recognition systems¹⁹. There is unfortunately a lack of formal analytical research into how effective speech recognition systems have been in UK HE/FE situations and environments, such as those mentioned in section 4.1.

Such research would have the benefits of:

- proving or showing how effective such systems were
- enabling fund-holders in similar educational institutions to determine whether to purchase such systems
- publicising the use of such systems in educational institutions

As opposed to HE/FE, there has been more research into the use and application of speech recognition in school classrooms.

¹⁹ Prominent research groups with technical leanings include the Speech Research Group at Cambridge University (<http://svr-www.eng.cam.ac.uk/research/speech/>), the Speech and Hearing Research Group at Sheffield University (<http://www.dcs.shef.ac.uk/research/groups/spandh/>) and the Centre for Speech Technology Research (<http://www.cstr.ed.ac.uk/>) at the University of Edinburgh.

One key research project²⁰ experimented with the use of speech recognition systems by people with special educational needs. This illustrated a variety of educational benefits, including:

- people could see their inaccurate pronunciation of words in textual form, as the wrong word appeared on the screen. This encouraged the correct pronunciation of the desired word.
- people had to decide whether to accept a word, or choose one from a list provided by the software, thus forcing them to read and consider vocabulary they had used.
- the speech of the people in the study improved, as to use the software efficiently involved correct pronunciation at a regular speed of delivery.

6.2 VoiceXML

One area of note that UK HE researchers have become involved in is in the development and progress of VoiceXML²¹. This XML language is increasingly used in information systems where enquiries can be passed and resolved, from the verbal output of the user, through a VoiceXML interpreter. This would, as described elsewhere in this paper, allow companies and services to provide unstaffed and immediate information-based systems that can be accessed through the telephone and other devices. A large number of companies (most, but not all, in telecommunications and IT) are involved in this initiative as well as researchers from many countries.

Highly related to these developments is the work involved in producing an official speech recognition grammar specification. This would allow content and system developers to produce systems that can input and output content that can be recognised by VoiceXML and other interpreters. The first version of this is publicly available²²; development is ongoing by the overseeing working group into refinement.

7. Potential uses in education

Contact with a number of practitioners and researchers in the field of speech recognition led to some interesting speculation regarding the feasible use of this technology in education.

No.	Application	Problems and likelihood
1	Teaching students of foreign languages to pronounce vocabulary correctly.	Unlikely in near future on a large scale, due to the software training currently involved.
2	Teaching overseas students to pronounce English correctly.	Ditto.
3	Making notes of observations during scientific experiments, so the	Likely, and is probably already used in individual circumstances.

²⁰ <http://www.becta.org.uk/technology/speechrecog/docs/finalreport.pdf> - the final report (June 2000) from a experimental project to see how effective speech recognition technologies could be to people with special educational needs.

²¹ <http://www.w3.org/Voice/Guide/> - an excellent overview of VoiceXML.

²² <http://www.w3.org/TR/speech-grammar/> - speech recognition grammar specification v1.0.

	scientist/research can focus on the observation without needing to view the monitor or keyboard. Similar to how a coroner verbally records notes during an autopsy.	Noise from the experiment, the researcher need to rapidly record some observations, and a vocabulary that understands the scientific terms present some issues.
4	Enabling students who are physically handicapped and unable to use a keyboard to enter text verbally.	Used already, though becoming increasingly widespread.
5	Enabling people with textual interpretive problems e.g. Dyslexia, to enter text verbally.	Used already, though becoming increasingly widespread.
6	Restrictive access on a high security computer, where a keyboard or other input device may be used by hackers.	Interest from a number of people, though a lack of "proof of concept" research hinders further development. Unlikely to be available in the near future.
7	Narrative-oriented research, where transcripts are automatically generated. This would remove the time to manually generate the transcript, and human error.	Likely in the near future. Current speech recognition technology places unacceptable compromise between accuracy and inhibiting the interviewee. Quicker and easier training systems for the interviewee will help, as will increases in portable computing processing power.
8	Capturing the speech of a lecturer or tutor.	Unlikely on a large scale, due to vocabulary, training and interpretive issues. In addition, filming of the lecture results in audio and visual content combined which may be more useful. However, see the "Liberated Learning Project" in A.2.
9	Using a speech recognition system in an examination.	Very likely. Technically, this is possible, and within current UK examination guidelines this appears to be acceptable ²³ .

8. Future applications outside education

There are a number of scenarios where speech recognition is either being delivered, developed for, researched or seriously discussed. As with many contemporary technologies, such as the Internet, online payment systems and mobile phone functionality, development is at least partially driven by the trio of often perceived evils that are "games, gambling and girls (pornography)". Though these applications are outside the educational sphere, it is important to remember that many ICT

²³ <http://www.ace-centre.org.uk/html/resources/vrpart2/vrec17.html> - some of the issues involved in using voice recognition system during a school examination.

innovations, incorporated into academia over the last decade, were developed in other sectors.

8.1 Computer and video games

Speech input has been used in a limited number of computer and video games, on a variety of PC and console-based platforms, over the past decade. For example, the game Seaman²⁴ involved growing and controlling strange half-man half fish characters in a virtual aquarium. A microphone, sold with the game, allowed the player to issue one of a pre-determined list of command words and questions to the fish. The accuracy of interpretation, in use, seemed variable; during gaming sessions colleagues with strong accents had to speak in an exaggerated and slower manner in order for the game to understand their commands.

Microphone-based games are available for two of the three main video game consoles (Playstation 2 and Xbox). However, these games primarily use speech in an online player to player manner, rather than spoken words being interpreted electronically. For example, a MotoGP for the Xbox allows online players to ride against each other in a motorbike race simulation, and speak (via microphone headset) to the nearest players (bikers) in the race. There is currently interest, but less development, of video games that interpret speech.

8.2 Gambling

Online gambling has become a major industry in the last four years (to the degree that it has effected changes in gambling taxation laws in the UK and other countries). Speech recognition has application in games such as online poker (multiplayer), where vocal commands can be both heard by the other players, and are (where appropriate) interpreted by the host computer in order to deal more cards, adjust the money staked and so forth.

8.3 Precision surgery

Developments in keyhole and micro surgery have clearly shown that an approach of as little invasive or non-essential surgery as possible increases success rates and patient recovery times. There is occasional speculation in various medical journals regarding the use of speech recognition in precision surgery, where a procedure is partially or totally carried out by automated means.

For example, in removing a tumour or blockage without damaging surrounding tissue, a command could be given to make an incision of a precise and small length e.g. 2 millimetres. However, the legal implications of such technology are a formidable barrier to significant developments in this area. If speech was incorrectly interpreted and e.g. a limb was accidentally sliced off, who would be liable - the surgeon, the surgery system developers, or the speech recognition software developers?

8.4 Domestic applications

²⁴ <http://www.womengamers.com/revprev/sim/seaman.html> - review of Seaman (2000) for the Sega Dreamcast.

There is, inevitable, interest in the use of speech recognition in domestic appliances such as ovens, refrigerators, dishwashers and washing machines. One school of thought is that, like the use of speech recognition in cars, this can reduce the number of parts and therefore the cost of production of the machine. However, removal of the normal buttons and controls would present problems for people who, for physical or learning reasons, cannot use speech recognition systems.

8.5 Wearable computers

Perhaps the most futuristic application is in the use and functionality of wearable computers²⁵ i.e. unobtrusive devices that you can wear like a watch, or are even embedded in your clothes. These would allow people to go about their everyday lives, but still store information (thoughts, notes, to-do lists) verbally, or communicate via email, phone or videophone, through wearable devices. Crucially, this would be done without having to interact with the device, or even remember that it is there; the user would just speak, the device would know what to do with the speech, and would carry out the appropriate task.

The rapid miniaturisation of computing devices, the rapid rise in processing power, and advances in mobile wireless technologies, are making these devices more feasible. There are still significant problems, such as background noise and the idiosyncrasies of an individual's language, to overcome. However, it is speculated that reliable versions of such devices will become commercially available during this decade.

Appendix A: Relevant resources

A.1 Mailing lists

Fonetiks

<http://www.jiscmail.ac.uk/lists/FONETIKS.html>

Mainly formal announcements concerning conferences and posts in speech-related research.

INSTIL (Integrating Speech Technology in Language Learning)

<http://www.jiscmail.ac.uk/lists/INSTIL.html>

A mostly quiet list, with an occasional mixture of speech-related postings.

Phonet

<http://www.jiscmail.ac.uk/lists/PHONET.html>

This list is for those interested in techniques and materials for the teaching of phonetics, with a special focus on teaching phonetics to present and future speech therapists.

Voice

<http://www.jiscmail.ac.uk/lists/VOICE.html>

A list for voice and speech practitioners working in HE and the creative industries.

²⁵ http://www.furui.cs.titech.ac.jp/publication/2000/icassp2000_3735.pdf - Speech recognition technology in the ubiquitous/wearable computing environment. An analysis of the potential, and current problems, of wearable computers that could be operated through speech recognition technology.

A.2 Interesting projects and services

CAVE

<http://www.kpn-telecom.nl/cave/>

CAVE was a 2 year long project, funded by the European Commission, to develop and test speaker recognition and verification systems for use in telephone applications like calling-card or financial services.

CORETEX: Improving Core Speech Recognition Technology

<http://coretex.itc.it/>

This four-partner ongoing European project aims to improve core speech recognition technologies, which are central to the most important applications involving voice technology, e.g. multimedia information access and automatic services over the telephone network.

Liberated Learning Project

<http://www.canberra.edu.au/pathways/papers/bain.pdf>

An attempt by a university to translate lectures, in real-time, into text so that students can both see and hear what the lecturer is saying, and collect a hard copy transcription at the end.

Multimedia Document Retrieval

<http://svr-www.eng.cam.ac.uk/research/projects/mdr/>

A completed project which aimed to provide speech-oriented indexing of audio and visual material. Trials of the system in TREC evaluation 9 provided promising results, indicating that speech-oriented indexing systems have the potential to provide acceptable levels of recall and precision.

SCRIBE

<http://www.infj.ulst.ac.uk/scribe/>

The aim of the SCRIBE Project is to provide training, advice, and support in the use of voice technology for computer users with disabilities in order to enhance their access to education and to employment. This is achieved by delivering short courses of up to two days duration in the use of relevant software. The training is provided free of charge

Speaking to write

<http://www.edc.org/spk2wrt/>

A completed project that examined the use of speech recognition software by US schoolchildren with disabilities.

TechDis Accessibility Database

<http://www.niad.sussex.ac.uk/>

This web-based database contains details of ICT hardware and software products which can be used to assist accessibility.

VOICE project

<http://voice.jrc.it/>

This investigating the use of speech recognition systems in conversation, conferences, television broadcasts and telephone calls. It is developing prototypes of user friendly interfaces allowing an easier use of commercial products in translating the spoken voice into computer screen messages and subtitles.