

# Children are ready for speech technology - but is the technology ready for them?

Antony Nicol, Chris Casey & Stuart MacFarlane

Department of Computing, University of Central Lancashire  
Preston, Lancashire, PR1 2HE  
England  
anicol@uclan.ac.uk

**Abstract.** This paper explores the feasibility of utilising commercial speech recognition technology as an input mechanism for young children. Children learn to communicate through speech long before learning to read or write hence speech recognition technology can utilise skills they already possess. Several experiments have been conducted to measure the effectiveness of current speech technology when used by adults and young children. The results of the experiments indicate that the children are very willing to use the technology but there are currently severe limitations precluding its use in educational applications. However, the results do show this technology to be very promising for future use because researchers are continually improving recognition accuracy. This paper provides background and guidelines for researchers wishing to embrace this technology.

## 1 Introduction

This paper is the result of several experiments that have been carried out as part of a PhD research project into determining whether modern speech recognition technology can be used as an effective computer interface particularly for young children. Effective speech recognition for young children would be a very valuable tool for teachers to use in the class room and parents to use at home; it would remove the interface barrier between children and the educational potential of the computer. Areas where speech recognition would be of particular value in teaching infant children include: learning colours and shapes; learning the alphabet; learning phonics, creative writing and other areas where the interactivity offered by the computer can provide useful activities that are normally unavailable to children of pre-reading age. Areas of potential benefit for children of all ages include: pronunciation improvement (Russell, 1996); reading tuition (Mostow, 1994) foreign language tuition (Eskenazi, 1996) and other areas that would normally require a human listener to be present to assess performance.

Teachers and parents need to invest a large amount of time on a one-to-one level with children who have not yet mastered the skill of reading. Modern lifestyles limit the one-to-one contact time at home and the Pupil:Teacher ratio of 23:1 in English schools (DES, 2000) limits prolonged individual contact between teachers and pupils.

There is a great deal of research in the field of HCI but the bulk of the effort has concentrated on interface design for adults; “there has been little interest in the design issues that arise when the users are children.” (Crook, 1998). (Oviatt, 2000) has studied children speaking to a computer but research into using commercial speech recognition technology as an interface for young children is very thin on the ground. Some research activity in this area has been detected (Mostow, 1994), (Russell, 1996) but these

research teams use custom built speech recognition software rather than commercially available technology. Some success with older children using commercial technology has been reported (O'Hare, 1999) but much research work has yet to be done in the area of HCI for young children where speech recognition as an input device is a small but very important aspect of this discipline. The high level of research activity in the area of speech recognition technologies and the continual improvement of recognition accuracy coupled with the potential benefits for the education of children justify research in the area of interface design for children using speech.

This paper includes experiments designed to test and optimise speech recognition accuracy through training. Results of accuracy measurements for adults and young children are presented. An overview of methods for customising applications to accept speech as input is provided to help researchers with only rudimentary programming skills to become involved in this exciting technology. A set of guidelines for speech interface designers is proposed to help designers get the best results from current speech recognition technology.

## **2 Speech recognition experiments**

The following experiments were carried out at North-West Lancashire primary school. The subjects in each experiment consisted of three male teachers, three female teachers, six male pupils and six female pupils. The primary school caters for children in the age range of 4 to 11 years consisting of six classes designated as “Year-1” through to “Year-6”. Different teachers are responsible for children in each of these six primary years and all of these teachers were subjects in the experiments. The teacher and two pupils (one male and one female) were selected from each of the six primary school years in order to provide coverage of the whole age range and an even distribution of gender. As these experiments are the first in a series, the teachers were asked to select pupils whom they considered to be the most articulate in their class. The six teachers were asked to compile a list of twenty words and twenty phrases that were representative of national curriculum requirements for their particular year. The use of both single words and phrases enables further experiments to be carried out to determine whether the speech recognition technology performance differs when using single words or phrases. The reason for asking the teachers to choose the vocabulary was to ensure that the words and phrases were suitable for each particular age group of the children. It also removed any potential bias from the experimenter choosing words and phrases that were likely to produce an artificially high accuracy during the recognition accuracy tests discussed later in this paper.

### **2.1 Do users want to use this technology? - A pilot study**

#### *2.1.1 Introduction*

An investigation into the acceptance of speech as an input medium was carried out to provide a measure of the likelihood of its adoption. Speech recognition is a relatively new technology which is still under development and is not flawless; (Read, 2001) and (Casey, 1997) describe speech recognition as a “disobedient interface”. Although manufacturers claim recognition accuracies of up to 98%, (Halverson, 1999) argues that even with this level of accuracy “error correction still consumes a significant portion of user effort in text creation tasks”. This pilot study, however, raises the possibility of a more fundamental problem: do users feel comfortable using this technology?

Over a period of several weeks, the teachers and children were observed interacting with a computer using speech as the input mode. The study did not involve speech recognition at this stage and was used to simply observe the users interacting with the computer without the added influence on the user of the errors that would be generated from the speech recognition system. However, the teachers and children were not made aware of this “Wizard of Oz” approach (Kelley, 1983) because the observations required the user to think that the computer was “listening” to them.

A speech capture application was developed to provide the interface for the users. The appropriate forty words and phrases are spoken by the computer and the user simply repeats each one in turn. The spoken input is recorded for use in further experiments.

### *2.1.2 Observations*

The adults found the exercise embarrassing. It was surprising to find how many of the teachers lacked confidence when confronted with the technology. Several apologised in advance for their ignorance of technology and blamed themselves in advance for things that would undoubtedly go wrong. After spending some time reassuring the teachers, the recordings were made. The teachers were asked to speak the same words and phrases as the pupils from Year-5 to enable a direct comparison of speech recognition accuracy of adults and children speaking the same words and phrases. Year-5 pupils were selected because they were the oldest children that would still be available a year later for any follow-up experiments that might be required.

Most of the adults spoke more quietly and with less enthusiasm than they did in conversational speech so after completing their forty words and phrases they were asked to repeat the exercise but this time in total privacy in a room that they knew was well sound proofed. The results from these recordings were noticeably better. This may to some extent be due to the increased familiarity with the system but the prosody of the spoken words and phrases was noticeably more natural and relaxed. The recording of all teachers was not carried out at the same time so they were asked not to discuss the sessions with each other to avoid the feedback from affecting the way the users responded to the system.

The teachers were individually asked to explain how they felt during the two recording sessions. Five of them said they found the exercise embarrassing in the presence of others. They were asked if they felt intimidated in any way by my presence. The general consensus was that they were intimidated at first as they were worried that I would feel that they were incompetent. However, once the simplicity of the application was demonstrated they felt confident with the technology but still felt uncomfortable in speaking to the computer with others present. The teachers were asked whether repeating the exercise in a room on their own was any more comfortable. The general consensus was that it was more comfortable. One teacher however, a male senior teacher and head of Information Technology for the school found the exercise straightforward whether observed or not.

Similar observations of the children showed that they were far less inhibited than the adults and had no problems in speaking naturally to the computer in the presence of an observer. All the children found the recording application very easy to use and even the reception year children (4 years old) were able to make unassisted good quality recordings. They seemed to become absorbed in the task and although I was sat next to

them (but slightly behind) they appeared to be unaware of my presence. In each case I left my seat and walked around for a few seconds and not a single child looked away from the screen to see what I was doing.

### *2.1.3 Conclusion*

The objectives of this exercise were to collect speech data and determine whether users - particularly children - would be willing and able to interact with the computer using speech as the input mode and as such the study was successful. However, the observations were carried out on a small number of subjects and so no definitive conclusions can be made but in the absence of other research findings, it does indicate that there is a possible self-consciousness related problem associated with the use of this technology by adults. (Lewis, 1995), reports that self-consciousness can affect the way we speak and this was observed during the tests. (Anind 97) reports that users phrase their requests and vocabulary differently when they believe they are talking to a computer. However, there were no reported feelings of embarrassment among the participants. There was no evidence of self-consciousness with the children; they were very keen to speak to the computer.

The results of this test suggest that even though speech recognition system developers concentrate on optimising the technology for adult speakers, adults may in fact be less at ease with the technology than children who appear to be natural users. However, if the children of today regularly speak to the computer then it is possible that they, as adults of tomorrow, will be more willing and able to use technology they have grown up with.

## **2.2 An experiment to evaluate the effectiveness of speech recognition as an HCI input device**

### *2.2.1 Introduction*

The underlying speech technology responsible for converting speech into text is often referred to as a speech recognition engine. This experiment reports on the accuracy measurements of Microsoft's Windows Highly Intelligent Speech Recognition engine (WHISPER) when used with adults and young children. This experiment was carried out to determine whether current speech recognition technology is actually useable as an input method; adults and children were observed using the technology and the recognition accuracy was measured for both adults and children to enable a comparison of recognition accuracies for users of various ages. An equal number of male and female users were involved in order to determine whether the recognition accuracy was gender specific as (Casey, 1997) found that recognition accuracy tended to be higher for boys than girls.

### *2.2.2 Process*

The users were asked to speak the appropriate forty words and phrases into a custom built speech recognition system using Microsoft's SAPI 4 speech recognition engine as the application's core. The teachers were again asked to speak the words and phrases prepared for the Year-5 children so direct comparison of recognition accuracy of the two age groups could be made.

These tests were carried out using an untrained speech recognition engine. This means that the recognition engine tries to recognise speech using the generic acoustic model provided by the manufacturer. Speech recognition systems generally provide two modes

of operation: dictation mode and command and control mode. In dictation mode the user's speech is considered to be made up from any of the words in the recognizer's vocabulary (typically 10,000 to 50,000 words in a modern system) whereas command and control mode expects a set of words and phrases from a limited and predefined set; in this case, the forty words and phrases are all that the speech recogniser has in its vocabulary. Command and control mode has been used for this experiment as it enables higher recognition accuracy and faster recognition as the system is only looking for a match on a fixed set of forty words and phrases as individual utterances and does not check to see if any of the recognised words in any phrase is out of context. Command and control mode however, requires a grammar to be provided which defines the structure of the phrases to be recognised. One grammar is required for each set of 40 words and phrases.

Recognition accuracy can be measured in different ways. During the recognition of an utterance, the engine generates several hypotheses of what the result could be and attached to each of these potential results is a confidence rating. After processing, the text string with the highest confidence value wins and if the confidence level is greater than or equal to a predefined threshold, the text string is output from the engine as a recognition result. If the confidence level of the best contender is less than the threshold, the engine reports a recognition failure. Therefore the result returned from the engine for a given utterance could be one of three values: the correct text string; an incorrect text string; recognition failure. As there are three results, the recognition accuracy could be measured in terms of an accuracy measurement or an error rate. Percentage recognition accuracy  $A$  and error rate  $E$  of 'n' utterances is determined by:

$$A = \frac{\textit{Correct}}{n} \cdot 100\% \qquad E = \frac{\textit{Incorrect}}{n} \cdot 100\%$$

The difference in these two measurements is in how the failed recognitions are treated. Educational applications need a low error rate whilst maintaining an acceptable accuracy level as failed recognitions will simply require the user to repeat the utterance. Incorrect recognitions produce misleading or educationally unsound results. For example, it is far better for the system to fail to recognise an utterance than to allow the possibility of the child being asked to read the word "Dog" and the system reporting a correct result when the child utters "Log".

Using the "out of the box" default settings for voice profile and recognition accuracy threshold settings, the results from each person were recorded. The number of correct and incorrect recognitions was recorded to enable both accuracy and error rates to be calculated. The group accuracy and error rate results were calculated using an arithmetic average of the group members' results and are presented in Table 1.

### 2.2.3 Results

The following table contains the results of 40 utterances from each user. The teachers and Year-5 children spoke the same 40 words and phrases. Words and phrases for the children in years 1 to 6 were different as each set contained words and phrases chosen by the teacher that conform to National Curriculum requirements for children of that age.

User group	Accuracy		Error rate	
	Mean ( $\bar{x}$ )	STD ( $\sigma$ )	Mean ( $\bar{x}$ )	STD ( $\sigma$ )
Male adults	87%	4	10%	5
Female adults	79%	4	15%	0
Male children	2%	4	52%	8
Female Children	1%	2	54%	6

**Table 1. Group accuracy measurements using an untrained SR engine**

Most of the teachers initially spoke to the computer in an unnatural manner which could be likened to that of a native speaker trying to communicate with a foreign tourist; the words in the phrases were spoken in isolation with each word emphasized and spoken more loudly than normal. Modern speech recognition systems use continuous speech recognition techniques that look for the merging of the end of one word into the start of another so do not perform as well if spoken to like a foreigner; this speech style better matches isolated speech recognition (Alleva F, 1997). Users should speak naturally to the system but a little more slowly than normal as there is an immense amount of processing involved in recognising speech. After explaining this, the teachers were able to achieve the accuracy results illustrated in Table 1. The teachers requested that this phase of the experiment should be unobserved.

The children spoke to the system confidently. However, they were simply repeating words and phrases that they had just heard; when used in an educational application such as a reading tutor, they are more likely to speak far less fluently as they will be using unheard text. This potential problem and the “sounding out” of unfamiliar words are just two of the many issues to deal with when designing an interface to an application such as a reading tutor.

#### 2.2.4 Conclusions

The object of this experiment was to check the accuracy levels against the manufacturers’ claims and to determine the extent of the problems that would be encountered by children using a technology that had been developed primarily for adults. The accuracy and error rate results for the adults are considerably better when compared to those of the children. The results for the adults are not as high as the typical 95% measured during independent tests (ZDNet, 1999) but were comparable. They were probably slightly lower than the reviewed packages as the teachers have a noticeable regional dialect. Error rates between 10% and 15% for the adults is high for general use and training is recommended to attempt to improve these figures by speaker adaptation which attempts to correct the differences between the speaker’s voice and that of the engine’s default voice profile.

This experiment has shown that it is pointless to use current speech recognition for young children in its default form as the accuracy and error rates are too poor. A further experiment was conducted to determine whether children training the engine will improve the recognition accuracy and error rates or whether there is a fundamental limitation of the technology when using speech from children.

## **2.3 An experiment to determine the effectiveness of training a speech recognition engine**

### *2.3.1 Introduction*

There are two uses of the term “training” when discussing speech recognition. One definition used by (Coulton, 2000) reports recognition accuracy improvements after training the user to speak to the system in a manner that will improve recognition accuracy. The other definition refers to the system attempting to adapt to the user’s voice. Some systems insist that initial adaptation training is carried out by a new user and refer to this process as “enrolment”. (Gandhi, 2002) discusses these and other recommendations for improving recognition accuracy for children.

When training a speech recognition system through adaptation, the system presents the user with a piece of text to recite. The engine “listens” to the user’s speech in order to determine the differences in the way the user speaks compared to the engine vendor’s default speech models and attempts to modify the speech model to better represent the user’s way of speaking. The more training a user undertakes, the better the recognition accuracy and error rates should become. This phase tends to be quite time consuming and a single session typically requires between 10 and 30 minutes of continuous speech depending on the engine used. Although time consuming, it is a simple procedure for an adult. However, it is a serious problem for a young child as training assumes that the user can read. Although the older children can usually read, their reading ability is representative of their age and unfortunately much of the training text contains advanced and complex sentence constructs. Observations of children attempting to read the training text highlight this problem – especially the younger children. The training text must be read fluently if the engine is to modify the user’s voice profile appropriately.

### *2.3.2 Process*

The adults and Year-5 children were asked to train three speech recognition engines: IBM Viavoice – UK English, Dragon Naturally Speaking (preferred edition) and Microsoft’s WHISPER. This involves speaking a set piece of text which is provided by the speech recognition engine’s training software. The text is long and complex in content (i.e. difficult for some Year-6 children to read). The IBM Viavoice system was difficult for the adults to train as it continually reported recognition errors. It was found to be impossible for the children to train so training was stopped prematurely and the engine was no longer used. The Dragon systems’ engine provided training text at various levels of reading difficulty. As the training text only provided approximately 10 minutes of training, a second training piece was selected to increase the training time to better match that provided by the Microsoft engine training session. When the children met words they did not recognise, I whispered the words to them. The Microsoft speech recogniser provides a training interface that enables the user to specify their gender and whether they are a child or an adult. Training the Microsoft engine was very similar to training the Dragon engine and both used a simple interface. If the children could read all the text confidently they could have managed the training themselves. However, the children from years 1 to 4 would not have been able to read the text and time constraints did not permit the assisted training needed by the younger children.

### 2.3.3 Results

After training the engines, the recognition exercise of reciting the forty words and phrases was repeated with the teachers and Year-5 children. The results were averaged and are recorded in Table 2.

	Accuracy				Error Rate			
	Dragon		Microsoft		Dragon		Microsoft	
User group	$\bar{x}$	$\sigma$	$\bar{x}$	$\sigma$	$\bar{x}$	$\sigma$	$\bar{x}$	$\sigma$
Male adults	98%	3	97%	3	1%	2	1%	2
Female adults	97%	4	97%	6	1%	2	2%	3
Male child	75%	0	73%	0	8%	0	10%	0
Female Child	76%	0	74%	0	8%	0	10%	0

**Table 2.** Accuracy and error rate measurements across groups of users after SR training

### 2.3.4 Conclusions

There is a very marked improvement in the recognition accuracy and error rates after training. The results in table 2 are consistent with those cited by (Oviatt 2000) which estimate that the error rate for children is 2 to 5 times greater than for adults because children's speech contain much greater irregularities (dysfluencies or idiosyncratic lexical content) than adult speech. To combat this, a children's speech interface must be carefully limited, guiding the child to use the appropriate responses without appearing to be as tightly constrained as it actually is (Oviatt 2000).

These results are encouraging but a fundamental problem with the development of a speech interface for young children is highlighted: the engine training requires the user to be able to read. These results highlight the need for research into automating the training procedure. If it were possible for the application to utter a word or phrase, capture the speaker's utterance when they repeat the word or phrase and pass the textual and acoustic data to the engine for training, the whole process could be automated. Unfortunately, the technology used in these experiments does not currently support user-defined training text. However, this limitation should not stop work in this area. As long as the engines can be trained by reading the training text to a young child in order that they can repeat the text, further meaningful research can be carried out into designing speech enabled user interfaces for young children. It is likely that speech recognition vendors will eventually provide the facility for developers to specify their own training text under program control or provide default speech profiles for children as part of the speech recognition package.

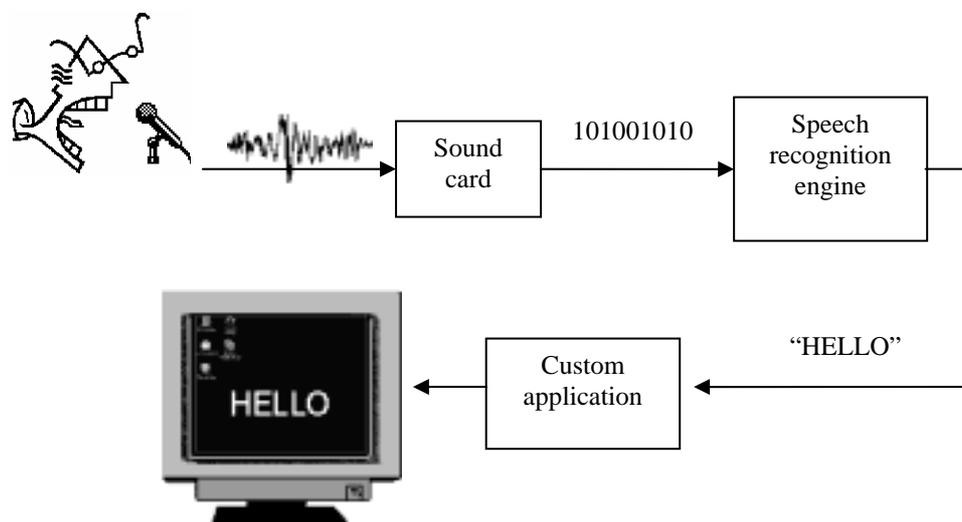
## 3 Creating a speech interface – an overview

It was once the case that to customise an application for speech recognition, the developer required a high level of programming skills. However, modern systems enable applications to be developed using less technical programming languages such as Visual Basic and C# (Microsoft, 2001) so enables this technology to a much wider range of developers. Only rudimentary programming skills are required to develop

simple speech recognition facilities to be used in custom built applications such as educational systems or prototypes to investigate the technology and its use.

Typical modern speech recognition systems such as IBM Viavoice, Dragon dictate etc., are speech aware applications that have been specifically designed to replace the keyboard for text input and replace the mouse for control actions such as accessing menu options. So how is this technology incorporated into educational applications in order to provide a speech interface?

The core of the typical dictation applications is the actual speech recognition engine. The recognition engine handles the conversion of speech into text and passes the textual results to the user's application to be used as required. Figure 1 illustrates a typical system where the speech recognition engine is effectively used as a "black box". Analogue speech from a microphone is converted into a digital form using a standard PC sound card. The speech recognition engine converts the digital data into text and passes it to the user's application to be used or displayed as required.



**Figure 1.** Typical speech enabled application

When each of the systems such as IBM Viavoice or Dragon dictate are installed on the PC, part of the installation process installs the manufacturer's speech recognition engine. Developers can use the engines to provide speech recognition facilities within their own applications with minimal programming effort. Details and comprehensive help files can usually be downloaded free of charge from the manufacturer's web site (Microsoft, 2001), (Lernout&Hauspie, 2002), (IBM, 2001).

#### **4 Overall conclusions**

The experiments have shown that it is worth pursuing the use of speech recognition as a user interface for children. In the short term, the recognition accuracy is high enough and the error rate low enough for its use in further research. However, it is unlikely to be useable in a commercial educational application until the problem with training the

recognition engine can be overcome or at least simplified to the point where it becomes practical.

This research exercise has raised many more wide ranging research questions that need to be answered before the technology can be used in schools as part of an educational tool. For example: how low does the error rate need to be before an application is considered to be educationally sound; how well does the system perform in a typical classroom environment; how well will training compensate for strong regional dialects or a speech impediment; how long will a child tolerate wearing a headset; how well will children tolerate failed recognition; how is the application going to provide an intuitive interface with useful feedback? Can speech recognition be used effectively in a classroom environment?

There are many areas of research that could benefit from the use of speech recognition technology and the results from these experiments should encourage fellow researchers in diverse areas such as computer assisted learning, education, literacy, child psychology and human computer interaction to consider the use of this evolving natural input medium.

## 5 Acknowledgements

Much of this work was carried out at Longton Primary school, Lancashire, England. I thank the teaching staff and head teacher Mr. Michael Dickinson for their help and cooperation. This research work is carried out in collaboration with Vektor Ltd. (<http://www.vektor.com>).

## 6 References

- Alleva F, Huang X.D., Hwang, M., Jiang Li. (1997). *Can Continuous Speech Recognizers Recognize Continuous Isolated Speech?* Paper presented at the EuroSpeech '97, Rhodes Greece.
- Anind K. Dey, Lara D. Catledge, Gregory D. Abowd & Colin Potts (1997). *Developing Voice-only Applications in the Absence of Speech Recognition Technology*. Technical Report GIT-GVU-97-06 Graphics, Visualization & Usability Center, Georgia Institute of Technology, Atlanta, GA 30332-0280 USA.  
Available: <http://www.cc.gatech.edu/fce/savoir/pubs/savoir.html> [2002, April 15]
- Casey, C, Snape, L, MacFarlane, S, Robertson, I. (1997). *Using Speech in Multimedia Applications*. Paper presented at the Teaching Company Directorate Conference on Multimedia.
- Coulton, David. (2000). *Can Voice Recognition Work for Computer Users? The Effecte of Training and Voice Commands*. Paper presented at the Proceedings of Human Computer Interaction 2000.
- Crook, C. (1998). Children as computer users: the case of collaborative learning. *Computers Educ.*, 30(3/4), 237 - 247.
- DES. (2000). *Statistical Bulletin: Class Sizes and Pupil:Teacher Ratios in Schools in England*. Department for education and skills. Available: <http://www.dfes.gov.uk/statistics/DB/SBU/b0222/030-t1.htm> [2002, March 29].
- Eskenazi, M. (1996). *Detection of Foreign Speakers' Pronunciation Errors for Second Language Training - Preliminary Results*. Paper presented at the CSLP 96 Fourth International Conference on Spoken Language Processing.
- Gandhi, Parmod. (2002). *Dragon NaturallySpeaking Complete, CD version*. The Literacy Centre. Available: <http://www.the-literacy-center.com/dns-c-cd.htm> [2002, April 13].
- Halverson, C, Horn, D, Karat, C, Karat, J. (1999). The Beauty of Errors: Patterns of Error Correction in Desktop Speech Systems. *Human-Computer Interaction - INTERACT '99*, 113 - 140.

- IBM. (2001). *ViaVoice SDK for Windows*. Available: [http://www-3.ibm.com/software/speech/dev/sdk\\_windows.html](http://www-3.ibm.com/software/speech/dev/sdk_windows.html) [2002, March 26].
- Kelley, J.F. (1983) *Natural Language and computers: Six empirical steps for writing an easy-to-use computer application*. Available: <http://www.musicman.net/oz.html> [2002, April 15]
- Lernout&Hauspie. (2002). *Dragon naturally speaking SDK download*. Available: <http://www.lhsl.com/naturallyspeaking/developers/download.asp> [2002, March 26].
- Lewis, Michael. (1995, January - February 1995). Self-Conscious Emotions. *American Scientist*.
- Microsoft. (2001). *Microsoft Speech SDK 5.1*. Available: <http://www.microsoft.com/speech/> [2002, March 26].
- Microsoft. (2002). *Platform SDK agent: Speech recognition*. Available: [http://www.msdn.microsoft.com/library/default.asp?url=/library/en-us/msagent/guidlin\\_2qzy.asp](http://www.msdn.microsoft.com/library/default.asp?url=/library/en-us/msagent/guidlin_2qzy.asp) [2002, April 15].
- Mostow, J, Roth, S, Hauptmann, A, Kane, M. (1994). *A prototype reading coach that listens*. Paper presented at the Twelfth National Conference on Artificial Intelligence (AAAI'94), Seattle.
- Najjar, L, Ockerman, J, and Thompson, J. (1998). *User Interface design guidelines for wearable computer speech recognition applications*. In IEEE VRAIS98. Georgia Tech,. Available: <http://www.hitl.washington.edu/people/grof/VRAIS98/home.html>
- O'Hare, E, McTear, M. (1999). Speech recognition in the secondary school classroom: an exploratory study. *Computers & Education*, 33(1), 27 - 45.
- Oviatt, S.L., (2000) *Talking To Thimble Jellies: Children's Conversational Speech with Animated Characters*, In B. Yuan, T. Huang & X. Tang (Eds.), *Proceedings of the International Conference on Spoken Language Processing (ICSLP'2000)*, Vol. 3, (pp. 877-880). Beijing, China: Chinese Friendship Publishers. Available: <http://www.cse.ogi.edu/CHCC/Publications/icslp-00067.pdf> [2002, April 15]
- Read, J. (2001). *Describing Disobedient Interfaces*. Paper presented at the Computing@UCLan, UCLAN, Preston, Lancashire.
- Russell, M, Brown, C, Skilling, A, Series, R, Wallace, J, Bonham, B, Barker, P. (1996). *Applications of Automatic Speech Recognition to Speech and Language Development in Young Children*. Paper presented at the ICSLP 96 Fourth International Conference on Spoken Language Processing.
- ZDNet. (1999). *Speech Recognition* (PC Magazine, Nov 5th 1999). ZDNet independent reviews. Available: <http://www.zdnet.com/products/stories/reviews/0,4161,2388289,00.html> [2002, April 1].