COMPUTER-BASED LEARNING: WHAT AUTOMATIC SPEECH RECOGNITION HAS TO OFFER

Helmer Strik, CLST, Radboud University Nijmegen

1 *Introduction*

Teacher-fronted instruction, with a teacher who teaches many students, is a very common teaching setting in many countries. This also applies to language learning, although one could imagine that having a teacher who can devote all his or her attention to only one or a few students would be more beneficial, especially when practicing oral skills. In fact, research has shown that students who receive one-on-one instruction perform as well as the top two percent of students who receive traditional classroom instruction (Bloom 1984). The problem is that, in general, a human tutor for every student is not feasible, because it is too expensive and there are not enough teachers. A possible solution would be to use computer tutors: computer-assisted learning (CAL). Computer-assisted language learning (CALL) has already received considerable attention. Many text-based CALL systems have already been developed, in which information is provided on the computer screen (the output of the computer), and the user can interact by means of a keyboard and a mouse (the input of the computer). But what about speech?

Speech-based applications certainly do exist. A well-known example is the screen reader, a program that reads aloud the text presented on the computer screen. Many websites now offer the possibility of listening to the text. Another example is the reading pen (Figure 1) which can be used to make printed text audible by moving the pen over the text line by line. These and other applications are used by people with reading problems, such as visually disabled or dyslexic people. In these applications use is made of text-to-speech technology, which is also referred to as speech synthesis. More applications with speech synthesis will certainly appear on the market. A recent application is a mobile phone that makes texts audible (Figure 2): a photo is taken of the text, optical character recognition (OCR) technology converts the photo to characters, and speech synthesis converts the characters into speech.

Text-to-speech technology is also used in CALL programs, thus making it possible to practice listening skills. For speaking skills, programs have been developed in which the learner is invited to speak: however, often nothing is done with this speech, no assessment is carried out, or the speech is recorded and the learner has to decide for herself whether her utterance was correct by comparing her utterances to the stored examples. Although self-assessment is better than no assessment at all, it is known from the literature that L2 learners are not always able to detect their own errors. In addition,

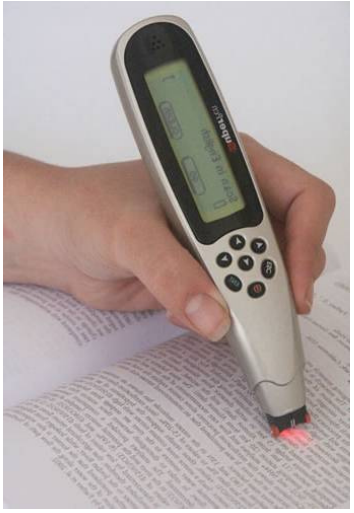


Figure 1: Reading pen

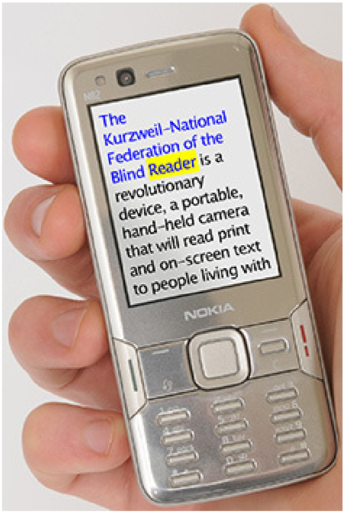

The Kurzweil-National Federation of the Blind Reader is a revolutionary device, a portable, hand-held camera that will read print and on-screen text to people living with
NOKIA

Figure 2: Mobile phone for making texts audible

one should be careful in the way comparison between the learner's output and example utterances is carried out (Neri et al., 2002). In some CALL programs oscillograms (waveforms) or spectrograms of the example (target) and the learner's utterance are shown. In general learners are not clearly instructed about how to interpret these displays and are encouraged to imitate the example utterance. Many learners find such displays difficult to interpret and non informative. Furthermore, imitating examples obviously is not the optimal way to learn languages. A better way of helping learners would be providing corrective feedback on the basis of an automatic analysis of their speech. Teachers do this all the time, but the question is whether computers can do it.

A possible way to do this is by using automatic speech recognition (ASR) technology. However, up till now relatively little use has been made of ASR in computer-based learning. In the current paper a short overview is presented of some possibilities that ASR has to offer in computer assisted learning (CAL, section 2), computer assisted language learning (CALL, section 3), and finally in literacy training (section 4).

2 *ASR-based CAL*

In speech-driven CAL applications the system first has to recognize the utterances. This can be done by means of automatic speech recognition (ASR) technology. ASR is already used in many applications. In dictation systems ASR is used to convert speech spoken into the microphone into words that appear on the computer screen. In fact, this is exactly what I am doing now in 'writing', or better dictating, this text (by using a dictation system that I have installed on my computer). Sometimes the dictation system makes errors, but these errors can easily be corrected, and even if you consider the time needed to correct errors, dictating for me is more efficient than typing.

ASR is also used for 'command and control'. A well-known example is that on many cellular phones one can speak a name or another short 'command' (like "call home" or "call mother") to dial a number. And with the same program that I am using to dictate this text, I can also 'command and control' my PC. For instance, when I spot an error in the dictated text, I utter commands such as "scratch that" and "correct that", and the PC follows my orders. It is also possible to select text and format it, to select files and open them, to read or send e-mails, etc., all through voice commands. ASR is also present in so-called spoken dialogue systems which can be used to obtain information (over the phone) on different topics such as train time schedules (Strik et al., 1997), and weather (Zue et al. 2000).

ASR is gradually improving, thus opening up possibilities to use it in new applications. One of them is ASR-based tutoring. These systems, which are also referred to as Intelligent Tutoring Systems (ITS), are spoken dialogue systems for learning. For example, ITSPOKE is a system for learning physics (e.g. Litman and Silliman, 2004), and SCoT is about shipboard damage control (Pon-Barry et al., 2004). In these ASR-based tutoring systems the students are thus communicating through speech, while the subject is not necessarily language, it can also be another topic such as physics or math. On the other hand, in ASR-based CALL the subject matter is language and speech.

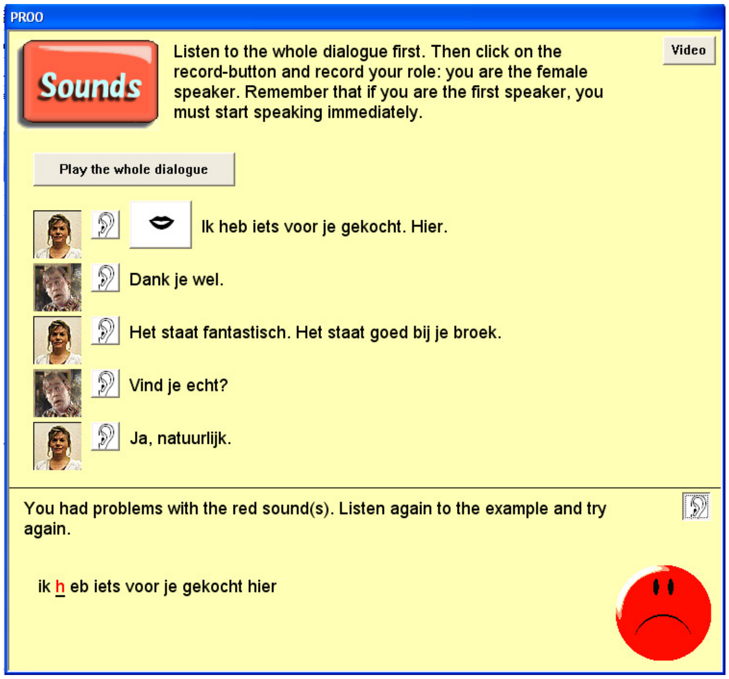
3 ASR-based CALL

ASR-based CALL received increasing attention in the late nineties. In 1998 the 'Speech Technology in Language Learning' (STiLL) workshop was organized in Marholmen (Sweden). This was probably the first time that the CALL and speech communities met, and it was the starting point of a number of STiLL and InSTiL (Integrating Speech Technology in (language) Learning) activities. In 1999 a special issue of CALICO appeared, entitled 'Tutors that Listen', which focused on ASR. It concerned mainly so-called 'discrete ASR', i.e. the recognition of individual words that are uttered with pauses between the words. Obviously, this is not the preferred way of communicating when learning a language. Therefore attention has shifted towards continuous speech. For an overview of the history in this field see e.g. Delcloque (2000) and Eskenazi (to appear).

Most applications that use ASR, e.g. the ones mentioned in the previous section, concern native speech. However, in CALL applications ASR has to deal with non-native speech. Non-native speech is much more challenging for ASR than native speech. In general, non-native speech shows more variation than native speech and is likely to contain non-native speech sounds and more disfluencies such as filled pauses, repairs, restarts and repetitions. ASR is not flawless, not for native speech, and certainly not for non-native speech. Note that this is not even the case for (speech recognition by) humans. Native listeners may have a hard time in understanding what non-native speakers are trying to say. In a normal conversation this is not necessarily a problem: we do not have to recognize each word perfectly to understand the message, and if we do not understand something, we can always make this clear through verbal or non-verbal (e.g. raising the eyebrows) communication. Furthermore, we use a lot of extra knowledge (general world knowledge, context of the conversation, speaker-specific information, etc.) to recognize utterances. Out of context many native utterances are even difficult to understand, especially those from 'unfamiliar voices' (i.e. spoken by persons the listener does not know). It is extremely difficult to include these aspects in an ASR-based CALL system.

Even though ASR for non-natives is certainly not perfect yet, it can be usefully employed in CALL applications if one takes its possibilities and limitations into account. This is what we did in our research. We have carried out research on the assessment of oral proficiency, pronunciation error detection, and pronunciation training; and currently we are involved in research on pronunciation, morphology, and syntax training for L2 speaking. Some of these projects are briefly mentioned below.

From 1996-1999 we were involved in the project 'Automatic Testing of Oral Proficiency' (ATOP, http://lands.let.ru.nl/~strik/research/ATOP.html). Methods for testing oral proficiency were proposed, implemented and tested, and the resulting scores correlated well with human judgments. Especially oral fluency could be assessed well by means of automatically calculated temporal measures. Speech technology can thus also be applied to carry out assessments at several levels (for fluency, but also for pronunciation, morphology, syntax, etc.; see below).



Listen to the whole dialogue first. Then click on the record-button and record your role: you are the female speaker. Remember that if you are the first speaker, you must start speaking immediately.

Ik heb iets voor je gekocht. Hier.

Dank je wel.

Het staat fantastisch. Het staat goed bij je broek.

Vind je echt?

Ja, natuurlijk.

You had problems with the red sound(s). Listen again to the example and try again.

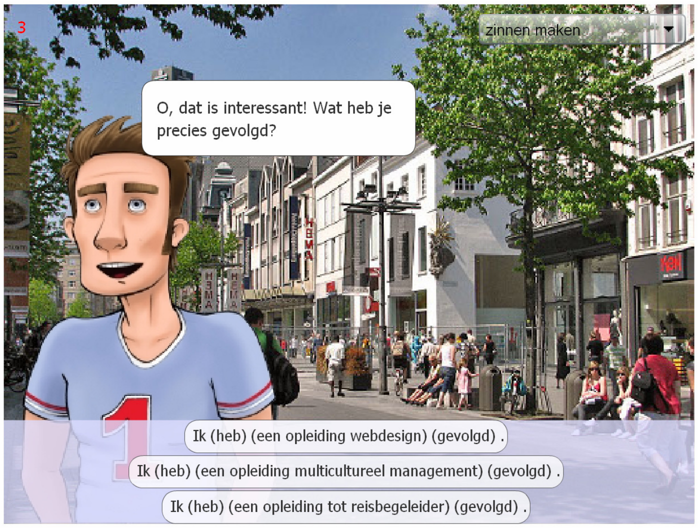
ik h eb iets voor je gekocht hier

Figure 3: Screen shot of the Dutch-CAPT system.

In the Dutch-CAPT (Computer Assisted Pronunciation Training) project (Cucchiarini, Neri, and Strik, to appear; http://lands.let.ru.nl/~strik/research/Dutch-CAPT/) a pronunciation training system was developed that gives feedback on segmental errors (see Figure 3). In this system we try to simulate a communicative setting. Users first look at a video. Depending on their gender they than play the male or female part, and the system plays the part of the other interlocutor. The system recognizes the spoken utterances, and after the conversation the feedback is presented. For instance, in Figure 3, the feedback is that the first sound of the second word was not pronounced correctly. The user can then listen to the example utterance and a recording of what was uttered by the user, and try again. In addition to these 'dialogue-type exercises' there are also some other exercises, e.g. with minimal pairs (short versus long vowels, etc.). For language learners who used this system the decrease in the number of pronunciation errors addressed in the training was substantially larger, compared to a control group that did not use our system (Cucchiarini, Neri, and Strik, to appear). The language learners only used our system four times (once a week, for 30 to 60 minutes). More intensive training will probably lead to larger improvements.

In the Dutch-CAPT system we gave immediate feedback on pronunciation errors (see Figure 3). In order to do so we developed algorithms for pronunciation error detection. Using a more generic ASR-based technique the algorithms correctly detect the pronunciation errors in about 80-90% of the cases (Cucchiarini, Neri, and Strik, to appear). However, if we focus on certain errors and make use of specific acoustic-phonetic information for each of these errors, the performance of pronunciation error detection can even be improved (Strik, Truong, de Wet, Cucchiarini, 2007; to appear).

Currently we are involved in a project called DISCO: Development and Integration of Speech technology into Courseware for language learning (Cucchiarini, van Doremalen, and Strik, 2008; http://lands.let.ru.nl/~strik/research/DISCO/). In this project we will develop a program for oral proficiency training for Dutch as a second language (DSL). The application optimizes learning through interaction in realistic communication situations and provides intelligent feedback on various aspects of DSL speaking, viz. pronunciation, morphology and syntax. In Figure 4, a syntax exercise is shown: "zinnen maken" ['making sentences']. The prompt is "O, dat is interessant! Wat heb je precies gevolgd?" ['Oh, that is interesting! What have you exactly followed?']. The prompt is shown on the screen, and made audible (recorded utterances). The learner can choose from some possible answers: "Ik (heb) (een opleiding X) (gevolgd)." ['I (have) (a course X) (followed).']. In the final application the word groups within brackets will be presented in a random order. The learner then has to put it in the correct syntactical order, and utter it. The order of the words within the brackets should remain fixed, the complete word groups within brackets can be swapped. Using ASR the user gets feedback on the spoken utterance, and the dialogue then proceeds depending on the chosen answer. The complete dialogue is structured as a tree with various branches, and thus the dialogue could evolve differently the next time the learner enters the exercise. We will develop dialogues on several topics, each of them structured as a branching tree. Within the team of the DISCO project we have people with various kinds of expertise, e.g. on DSL teaching, acquisition, and related pedagogical aspects, on designing language courseware (keeping in mind pedagogical goals and personal goals of the learners), on developing speech technology for language learning, and on software integration. This project is carried out within the framework of the Stevin Program (http://taalunieversum.org/taal/technologie/stevin/).



O, dat is interessant! Wat heb je precies gevolgd?

Ik (heb) (een opleiding webdesign) (gevolgd) .

Ik (heb) (een opleiding multicultureel management) (gevolgd) .

Ik (heb) (een opleiding tot reisbegeleider) (gevolgd) .

Figure 4: Screen shot of the DISCO system.

Furthermore, a new project, 'corrective feedback and the Acquisition of Syntax in Oral Proficiency' (ASOP; http://lands.let.ru.nl/~strik/research/ASOP.html), will start soon. In this project we will compare and test different types of (corrective) feedback, in order to study the effectiveness of different feedback forms.

Finally, we are involved in a demonstration project with the Dutch title 'Alfabetisering met een luisterende computer' (learning to read and write with a listening computer; http://lands.let.ru.nl/~strik/research/ST-AAP.html). In this project we will 'add speech technology' to an existing literacy training course called 'Alfabetisering Anderstaligen Plan' (AAP; http://www.alfabetiseren.nl/). In the current version of the course no evaluation is carried out on spoken utterances. Evaluation of the spoken utterances will be made possible in the ST-AAP project by developing suitable speech recognition technology. This project, like the DISCO project, is carried out within the framework of the Stevin Program.

4 *ASR-based literacy training*

Research on reading tutors already started long ago, e.g. the Listen system by Mostow et al. (1994) and the STAR system by Russel et al. (1996); and still receives attention, e.g. within the Flemish SPACE project (Duchateau, et al., to appear). Note that in reading tutors it is not the computer that reads, it is the learner that reads. The text that has to be read is displayed on the computer screen, and the computer tracks the learner (by means of ASR technology) who is reading what is visualized on the screen (e.g. a prompt jumping from word to word). Optionally, the computer can give additional feedback, e.g. if it notices that the learner has problems reading some parts of the text it can provide reading (pronunciation) instructions.

Interesting work has also been carried out in the Foundations to Literacy reading program (see, e.g., Wise et al., 2008). For instance, exercises have been developed to teach phonics (see Figure 5) and fluency, by means of interactive books (see Figure 6). Marni, the virtual person on the top-right, acts as the teacher: natural voice has been recorded and can be made audible, accompanied by synchronized lip movements and facial expressions. In the phonics exercise (Figure 5) the learners look at the screen and listen, and respond by means of keyboard and mouse; and in the interactive books (Figure 6) the learners read the text aloud, and the system tracks them by means of ASR.

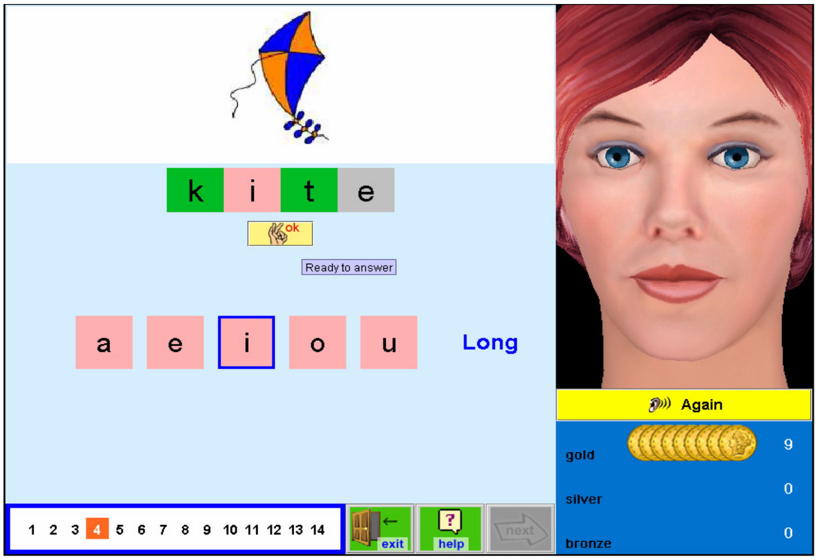


Figure 5: Training phonics

So far, most of the research carried out concerned children, but there is also is a need for literacy training programs for adults. What are the possibilities regarding speech technology for literacy training? First of all, useful listening exercises could be made by using text-to-speech technology. Many good quality text-to-speech programs are available nowadays; the speech is intelligible but does not always sound natural, i.e. one can hear that it is not a human voice but a computer voice. Another option is to use recorded speech, which obviously is more costly and less flexible. Text-to-speech technology is more flexible because it can easily be applied to new texts and because it is easier to change aspects of the produced speech, such as speech rate, pitch, and voice types (e.g. different female and male voices). Given the quality of the current state-of-the-art text-to-speech technology, an interesting question is for which applications (goals) text-to-speech technology can be applied within the context of literacy training.

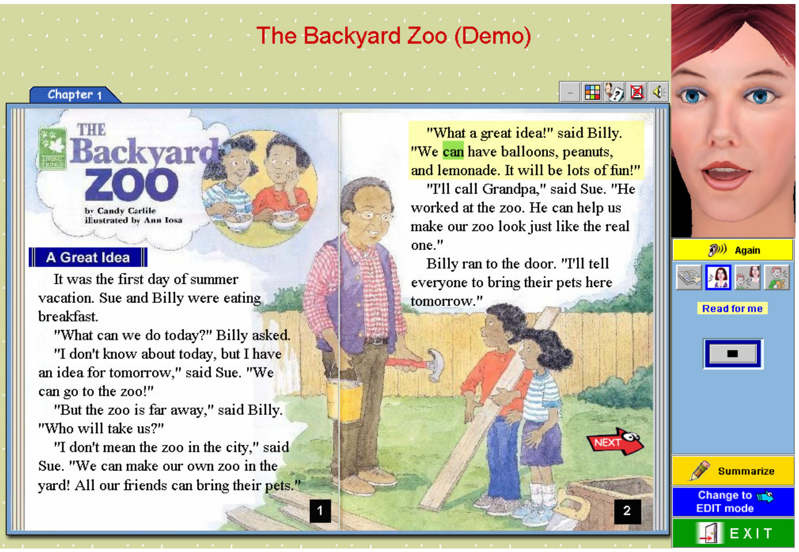
Speech technology can also be used for oral proficiency training and assessment. In general, ASR technology can be applied to recognize sounds, words, and utterances. In addition, it is possible to calculate a measure (a so-called confidence measure) that quantifies how well the recognized unit (sound, word or utterance) matches the speech signal. This combination of recognizing different speech units (sounds, words and utterances) and calculating confidence measures for these speech units can be used creatively to develop many exercises for oral proficiency training. Some examples are provided below.

In the ideal case, the software should be adaptive in several ways: the exercises should adapt to the level of the learner, and the errors made by the learner (e.g. the speech rate could be gradually increased; and if someone frequently makes errors regarding vowel length and much less on the voiced-voiceless distinction, that person should be presented many exercises on vowel length and much less on voicing); the user interface should adapt to the preferences of the learner (e.g. someone could have preferences for the amount of feedback (much or little), the type of feedback (at meta level or not), the timing of the feedback (immediate or delayed), etc.). In addition also the ASR should be adaptive. Every speaker, native and non-native, has its peculiarities, and if the speech recognizer can tune in to 'the voice' of that person the performance can be increased substantially (i.e. the number of errors made by ASR can be substantially reduced). There are several ways to do this. The first option is so called enrollment: at the start the user reads some text that is presented on the screen, and the recorded speech material is used to retrain the speech recognizer. This is a kind of adaptation in advance, before actually using the system. Another option is to adapt during use. This can be carried out in a supervised way, in which the user has to confirm whether an utterance was recognized correctly or not, or in an unsupervised way, in which the system has to determine automatically which utterances should be used for a adaptation.

In order to get a better idea of the possibilities of speech technology, and more specifically automatic speech recognition (ASR), for literacy training, some examples are presented here. Practicing phonics (sound letter combinations) has already been briefly mentioned above. One option is that the learner listens to speech sounds produced by the system, and enters what the corresponding graphemes are (Figure 5). Another option is that the learner reads the graphemes presented on the screen, pronounces the corresponding speech sounds, the system checks by means of ASR whether the correct sounds have been uttered, and provides feedback.

Such procedures can be carried out for single speech sounds, but similar procedures can be carried out for combinations of speech sounds, whole words, and even utterances. In all the examples in this paragraph, ASR is employed to analyze what has been said by the learner, and then to provide feedback on it. For instance, the learner reads a complete word on the screen, e.g. "kat" [cat], or listens to the word, and then utters the phonemes one by one (in this example, the three phonemes corresponding to the three graphemes in the word "kat"). It could also be done the other way around: the system pronounces the individual phonemes, and the learner has to utter the complete word. Another possibility is that the student repeats utterances orally. The utterances could consist of single words, combinations of words, complete sentences, or even a couple of sentences (depending on the level of the student). These utterances could be presented visually (i.e. text on the screen), auditorily (i.e. speech generated by the system), or a combination of both (as done in the Dutch-CAPT system, see Figure 3). In all these exercises, the pace could be gradually increased in order to make the exercise more challenging, and to enhance automatization.

Besides text-to-speech and ASR (speech to text) technology, other kinds of technology might also be useful for literacy training. For instance avatars, virtual persons (of which often only the head is shown) in the form of two- or three-dimensional models, that are often present in computer games. Avatars could be used, either as the interlocutors in a communicative setting (see, e.g., Figure 4), or as a virtual teacher that gives instruction and feedback (like Marni in Figures 5 and 6). For providing instructions and feedback on pronunciation a special kind of avatar may be useful (which is often called a virtual talking head) for which it is possible to show the positions and movements of the articulators. In the future it might also be possible to make creative use of the many resources present on the Internet, such as youtube movies. In many cases it might also be useful to include gaming aspects. Playing games is often stimulating and motivating, not only for children, but also for adults.



It was the first day of summer vacation. Sue and Billy were eating breakfast.

"What can we do today?" Billy asked.

"I don't know about today, but I have an idea for tomorrow," said Sue. "We can go to the zoo!"

"But the zoo is far away," said Billy. "Who will take us?"

"I don't mean the zoo in the city," said Sue. "We can make our own zoo in the yard! All our friends can bring their pets."

"What a great idea!" said Billy. "We can have balloons, peanuts, and lemonade. It will be lots of fun!"

"I'll call Grandpa," said Sue. "He worked at the zoo. He can help us make our zoo look just like the real one."

Billy ran to the door. "I'll tell everyone to bring their pets here tomorrow."

Figure 6: Interactive book

5 Discussion

At the moment the number of text-based CALL systems is much larger than the number of speech-based systems. Furthermore, ASR is used in only a small fraction of the speech-based systems. Given the current state of ASR technology, it is possible to usefully apply it in CALL applications. Using speech has many advantages. First of all, for many of us it is more natural, easier, and faster than using the keyboard or mouse. The speech signal also contains extra information, such as prosody (which is related to word stress, sentence accent, etc.), which can provide information on the emotional state of the speaker, or the confidence with which the utterances were spoken.

Important aspects of ASR-based CALL systems are that they can provide immediate feedback on errors in the spoken utterances, which is important for successful training. These computer programs can be available 24 hours a day, making it possible to practice more intensively than is usually possible with human teachers. These systems can also be available everywhere. Using them in a more private setting could be less stressful, and could stimulate learners to be more talkative. After all, it is known that many learners are reluctant to speak when others are present and that practicing is important to improve speaking proficiency (Swain & Lapkin, 1995).

ASR is not flawless, as was already mentioned above. Still ASR-based tutoring is possible if one takes account of what is possible with current technology, and what is not. ASR-based tutoring can be useful in cases where the subject matter is not speech, for instance for physics and mathematics (see the examples in section 2). Obviously, ASR technology becomes even more important when one has to practice oral skills, e.g. if one wants to learn another language, or for people with communicative disabilities. In the near future, the number of ASR-based tutoring systems will certainly increase, especially the number of ASR-based CALL systems.

Owing to the increasing mobility of workers around the world, the demand for language lessons is growing steadily in many host countries. In several cases the demand clearly outstrips the supply and immigrants have to wait months before being enrolled in a language course. A compounding problem is that many immigrant workers simply do not have the time to attend language courses. Such situations call for innovative solutions that can make language learning more effective, more personalized, less expensive and less time consuming. ASR-based CALL systems seem to constitute a viable and appealing alternative, or complement, to teacher-fronted lessons. ASR-based CALL is certainly interesting for non-literate learners for whom speech is even more important for communicating; especially for teaching non-literates ASR offers many promising possibilities.

References

Bloom, B. (1984). The 2-sigma problem: The search for methods of group instruction as effective as one-to-one tutoring, *Educational Researcher* 13: 4-16.

Cucchiarini, C., van Doremalen, J., & Strik, H. (2008). DISCO: Development and Integration of Speech technology into Courseware for language learning. *Proceedings of Interspeech*-2008 (pp. 2791-2794). Brisbane, Australia. http://lands.let.ru.nl/~strik/publications/a144-DISCO-IS08.pdf

Cucchiarini, C., Neri, A., and Strik, H. (to appear). Oral proficiency training in Dutch L2: The contribution of ASR-based corrective feedback. Accepted for publication in *Speech Communication.* http://lands.let.ru.nl/~strik/publications/a149-DutchCAPT-SpeCom.doc

Delcloque, P., (2000). *History of CALL.* http://www.ict4lt.org/en/History_of_CALL.pdf

Duchateau, J., On Kong, Y., Cleuren, L., Latacz, L., Roelens, J., Samir, A., Demuynck, K., Ghesquière, P., Verhelst, W., & Van hamme, H. (to appear). Developing a reading tutor: design and evaluation of dedicated speech recognition and synthesis modules. Accepted for publication in *Speech Communication.*

Eskenazi, M. (to appear). An overview of spoken language technology for education. Accepted for publication in *Speech Communication.*

Litman, D., & Silliman, S. (2004). ITSPOKE: An intelligent tutoring spoken dialogue system. *Companion Proceedings of the Human Language Technology Conference*: 4th Meeting of the North American Chapter of the Association for Computational Linguistics (HLT/NAACL), Boston, MA.
http://www.cs.pitt.edu/~litman/demo-final.pdf

Mostow, J., Roth, S., Hauptmann, A.G., & Kane, M. (1994). A prototype reading coach that listens. *Proceedings of the twelfth national conference on artificial intelligence* (AAAI-94) (pp. 785-792). American Association for Artificial Intelligence, Seattle, WA, August 1994.

Neri, A., Cucchiarini, C., Strik, H., & Boves, L. (2002). The pedagogy-technology interface in computer assisted pronunciation training. *Computer Assisted Language Learning*, 15, 5, 441-467.
http://lands.let.ru.nl/~strik/publications/a99.pdf

Pon-Barry, H., Clark, B., Bratt, E., Schultz, K., & Peters, S. (2004). Evaluating the effectiveness of SCoT: a Spoken Conversational Tutor. In Mostow, J. & Tedesco, P. (Eds.), *ITS 2004 Workshop on Dialog-based Intelligent Tutoring Systems*, 23-32.

Russell, M., Brown, C., Skilling, A., Series, R., Wallace, J., Bonham, B., and Barker, P. (1996). Applications of automatic speech recognition to speech and language development in young children. *Proceedings of the International Conference on Spoken Language Processing,* ICSLP'96, Philadelphia, USA, 3-6 October 1996.

Strik, H., Russel, A., van den Heuvel, H., Cucchiarini, C., Boves, L. (1997). A spoken dialog system for the Dutch public transport information service. Int. Journal of Speech Technology, Vol. 2, No. 2, pp. 121-131.
http://lands.let.ru.nl/~strik/publications/a31.pdf

Strik, H., Truong, K., de Wet, F., & Cucchiarini, C. (2007). Comparing classifiers for pronunciation error detection. *Proceedings of Interspeech*-2007, Antwerp, 1837-1840.
http://lands.let.ru.nl/~strik/publications/a133-PronErrD-IS07.pdf

Strik, H., Truong, K., de Wet, F., & Cucchiarini, C. (to appear). Comparing different approaches for automatic pronunciation error detection. Accepted for publication in *Speech Communication.*
http://lands.let.ru.nl/~strik/publications/a150-PED-SpeCom.doc

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics* 16, 371-391.

Wise, B., Cole, R., Van Vuuren, S., Schwartz, S., Snyder, L., Ngampatipatpong, N., Tuantranont, J., & Pellom, B. (2008). Learning to read with a virtual tutor. In C. Kinzer, & L. Verhoeven (Eds). *Interactive literacy education: facilitating literacy environments through technology* (pp. 31-75). NJ: Lawrence Erlbaum, Taylor and Francis Group.

Zue, V., Seneff, S., Glass, J., Polifroni, J., Pao, C., Hazen, T.J., & Hetherington, L., "JUPITER" (2000). A telephone-based conversational interface for weather information. *IEEE Transactions on Speech and Audio Processing*, vol. 8, 1.