THE EFFECTIVENESS OF A COMPUTER ASSISTED PRONUNCIATION TRAINING SYSTEM FOR YOUNG FOREIGN LANGUAGE LEARNERS

Authors MICH, ORNELLA, NERI, AMBRA, GIULIANI, DIEGO

ABSTRACT

In line with the European Union's policy to foster multilingualism in the member countries (Barcelona, 2002), Italy has recently made it compulsory to learn a foreign language from the first year of primary education onwards. Currently, 75% of Italian pupils are learning English as a foreign language (Eurydice, 2005). Within language learning, pronunciation obviously plays an important role because communication cannot take place below a certain level of pronunciation quality, even if grammar and vocabulary have been mastered. For children it is particularly important to start off with correct pronunciation examples to avoid that incorrect forms are acquired and possibly fossilized later on. This need is acute for Italian learners of English, because of the differences in the phonetic/phonological systems of these languages, and because in English the relationship between graphemes and phonemes is not as straightforward as it is in Italian. To help children achieve a correct pronunciation, teachers should offer individualized feedback, but the amount of feedback to the individual pupil is generally small due to time constraints. Some experts believe that a CALL (Computer Assisted Language Learning) system based on ASR (Automatic Speech Recognition) technology and on sound pedagogical guidelines could integrate classroom teaching by providing individualized practice and feedback (Eskenazi, 1999). The question that is often left unanswered is whether this type of systems can indeed offer valuable help towards the improvement of pronunciation skills, especially for children.

To test this hypothesis, a group of researchers at ITC-irst developed PARLING, a CALL system for Italian children learning English, which also includes ASR technology. The design of Parling was based on work with teachers and children and on a study of literature (Mich et al., 2005). The system offers English children stories and games in which the pronunciation of new words can be trained by means of ASR technology. The study presented here aims to investigate whether this system can supplement traditional language teaching by helping young learners acquire the pronunciation of English words at least equally well as if the training were provided by a teacher.

To this aim, we compared a group of pupils receiving teacher-led instruction with a group receiving ASR-based CALL (PARLING). A pretest-posttest design was used to measure the effects of 4 weeks of training. Pre- and post-test consisted of 28 words, selected from those present in the training material so as to cover most English phonemes. Each word was read aloud and recorded by each subject. The recordings were subsequently evaluated by 3 experts. We performed two different analyses: an analysis on global pronunciation skills for all words, and an analysis of the quality of (previously) unknown and difficult words, for which the support offered by the system may be more apparent.

Results show that 1) overall pronunciation quality improved significantly for both groups of pupils, 2) both groups also significantly improved in pronunciation quality with respect to difficult and unknown words, which indicates that the system was equally effective as the teacher in improving the pupils' pronunciation.

PRESENTATION

INTRODUCTION

The use of CALL systems for children is becoming more and more popular (Kawai & Tabain, 2000; Sfakianaki et al., 2001; Bunnel et al., 2000; Eskenazi & Pelton, 2002; Krajka, 2001). The fact that children grow up surrounded by engaging multimedia makes it almost a must to supplement traditional classroom instruction with such media (Purushotma, 2005). Moreover, carefully designed CALL systems can include game-like activities to stimulate task-based learning, thus giving children the impression that they are playing while they are actually also learning, and turning learning into a rewarding and fun experience (Wachowicz & Scott, 1999). As a matter of fact, children seem to enjoy working with CALL tools (Nicol, Casey, & MacFarlane, 2002), and educators are well aware of the importance of motivational factors for learning. Equally important are the pedagogical advantages that CALL systems can offer. First of all, they can provide abundant, realistic, and contextualized input, and opportunity for self-paced practice. In learning pronunciation the availability of these factors is crucial (Leather & James, 1996), especially in a *foreign* language (FL) learning context, i.e. outside the country where the target language is spoken. In this context, exposure to oral examples in the target language is generally limited both quantitatively and qualitatively to the teacher's speech, interaction with native speakers is often not possible, and learning mainly takes place through the written medium, which is likely to lead to orthographic interference in FL pronounciation. At the same time, very little time is available for the teacher to provide individual feedback on pronunciation, while this factor is necessary to make the learner aware of possible serious input-output mismatches that s/he is unlikely to notice alone (Flege, 1995). Moreover, correcting erroneous FL behavior in children is decisive to avoid that wrong habits are learnt, especially at pronunciation level, which may otherwise become fossilized later on.

The most advanced CALL system can nowadays provide automatic feedback on pronunciation quality by means of Automatic Speech Recognition (ASR) technology. This feedback can be limited to rejecting poorly pronounced utterances and accepting 'good' ones, to pinpointing specific errors (e.g. Eskenazi & Pelton, 2002; Bunnell et al., 2000; Tell me More. KIDS, n.d.). However, no study exists, to our knowledge, which examines systematically the actual pedagogical effectiveness of these systems. Since ASR-based CALL systems are known to occasionally generate erroneous feedback which, in turn, may disrupt the learning process, such an assessment is much needed.

To investigate this issue, a CALL system, PARLING, was developed at ITC-irst as an assistive English-FL tool for Italian primary-school-age children. This system, which includes an ASR component, was tested in a school. The remainder of this paper describes the operation of the system, the experiment conducted to test its effectiveness, and the results obtained.

THE CALL SYSTEM CONSIDERED: PARLING

The design of PARLING was based on an analysis of relevant literature and of existing systems with similar purpose. For the latter analysis, a system (Tell me More. Kids, n.d.) was selected by language teachers and by researchers at ITC-irst, which was deemed to meet

most of the requirements set by these experts. 25 ten-year-old children were subsequently asked to use this system in a series of tests to study how they would interact with it and to complete questionnaires on it (Giuliani et al., 2003). The results of this study, together with indications obtained from teachers and from relevant literature, led to the development of PARLING.

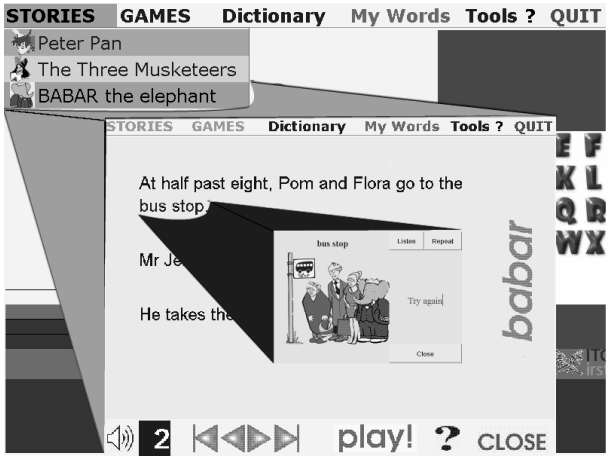

Figure 1: PARLING, the start display, a story page, an active word.

PARLING is a modular system. Each module is composed of a story, an adaptive word game, and a set of *active* words. The system is completed by a visual dictionary, a tool that allows children to create their own dictionary, and a simple help. The initial page of Parling presents the user with a menu to access stories, games, tools to build a new story (only in the teachers' version) or to create a personal dictionary, and a pictorial dictionary (see Fig. 1).

The stories consist of simplified versions of famous children's stories. Each time a page is loaded, its corresponding audio is played back. Some words in these stories can be activated: when the user clicks on them, a window appears showing the meaning of the given word (see Figure 1). The user can hear its pronunciation and try recording the word herself. The system analyses the recording in real time by means of ASR technology and responds with a message telling whether the word was pronounced correctly or not and eventually prompting the child to repeat the incorrect utterance. Each story comes with a different game meant to help children memorize the words in that story. The game dynamically adapts its content to the user's personal work path.

The dictionary includes a tool with which children can add new words, select images for them, and record the corresponding audio. All operations performed by a user are logged. This way, a teacher can always monitor the child's work and progress.

METHOD

To measure and compare possible improvements in global pronunciation quality after four weeks of traditional training and of training with PARLING, two groups of Italian children were studied before and after the training. The control group received instruction in the

form of traditional, teacher-led classes. The experimental group worked with PARLING during individual sessions.

Subjects

The subjets were all Italian native speakers aged 11 attending the same public school and sharing the same curriculum. They belonged to two different groups, but they were following the same type of classes and had the same English teacher. At the time of the experiment, they all had had 4 years of English-FL classes. Group C, i.e. the control group, was composed of 15 children, while group E, i.e. the experimental group, included 13 children.

Training

Group C followed four teacher-led (British) English-FL classes of 60 minutes each. During each session, the teacher read an excerpt from a simplified, English version of the Grimms' children story *Hansel and Gretel*. This story was chosen because Italian children of 11 are generally familiar with it, which could help our participants to more easily understand the English version they were presented for this study. Children in this group were given a printed version of the story. The teacher also discussed some words in the story, explaining their meaning and prompting the children to repeat them aloud after providing the correct pronunciation. At the end of each training session, each child also completed a printed word game based on words extracted from the excerpt that had been read in that session.

Children in group E had four individual CALL training sessions in the school's language lab, each lasting 30 minutes,[1] during which they worked with PARLING. During each session, they listened to part of the story and read it on the screen, and they repeated some of the words presented in that excerpt. They also played a word game which only included words from the story excerpt of that session. For this game, children had to pronounce and record the words proposed by the game. If the spoken utterance was rejected by the ASR module, the child had to repeat the word at least one more time.

In this way the training was considered approximately comparable for both groups, the only difference being that instruction was imparted by a teacher in the case of C, and by a computer in the case of E.

Testing

In order to be able to evaluate and compare the participants' possible improvements in pronunciation quality, we had all children read and record a set of 28 isolated words

[1] The limited amount of time for these sessions was due to practical constraints in the use of the language lab.

before and after the training, which were subsequently scored by three experts. Read speech was chosen as elicitation material for comparability purposes across subjects. The words were taken from the simplified version of the story presented during the training, and were chosen so as to cover the most frequent British English phonemes. These words varied with respect to length, articulatory difficulty, and lexical frequency. For some of them (e.g. *woodcutter, breadcrumbs*) it was assumed that they were entirely new to the participants before the training.

For the recording sessions, a dedicated tool was used which presented one word at a time on the screen, prompted the child to read it aloud and simulataneously recorded it. If the child felt that s/he had not pronounced the word correctly, s/he was allowed to repeat it and record it as often as s/he wished. The recordings took place with head-mounted microphones and the speech was sampled at 16 KHz.

Ratings

The recordings were scored independently by three native speakers of British English who were working in Italy as English-FL teachers. Each rater was asked to provide a score of global pronunciation quality for all utterances on a 10-point scale. Two different ratings were requested: for one, single words in separate audiofiles had to be scored individually, whereas for the other one, all words recorded by one participant were concatenated in a larger audiofile in order to have one score per speaker. Duplicates of 32 audiofiles were later presented for another scoring session, to allow for the calculation of intra-rater reliability.

Raters were allowed to complete the task in several sessions. To help them familirize with the scoring scale, examples of spoken words of 'poor' pronunciaiton quality were provided at the beginning of a rating sessions, together with words produced by a native speaker of English, i.e. examples of good pronunciation quality. The single-word audiofiles were presented in 28 blocks, with each block containing the same word uttered by all subjects at both testing conditions. The audiofiles in each block were presented in random order.

RESULTS

Reliability of ratings

The raters' scores were first analyzed to determine intra-rater and inter-rater reliability. The computation of inter-rater reliability was based on 1568 scores from each rater (28 participants x 28 words x 2 testing conditions). A Cronbach's alpha coefficient of .872 was obtained, which can be considered satisfactory. Intra-rater reliability was calculated on the basis of 32 duplicates for each rater. The intra-rater reliability coefficients for the three raters ranged from .757 to .859, indicating, again, high reliability.

Global pronunciation quality

Before analysing possible improvements in the two groups, we examined the relationship between the *speaker scores* and the *single-word scores*, by comparing the former with the mean of the single-word scores for each speaker. In other words, in one case we had one score assigned by three raters to one participant, in the other case we had one score per participant that was obtained by averaging all *single-word scores* for that participant. It

might have been possible that these scores differed if, for instance, few seriously mispronounced words located at the end of a concatenated audiofile affected more severely the *speaker scores* than the mean of the *single-word scores*. We found a strong, positive correlation between the two types of scores (r=.884, p<.01). For group C, the results were: r=.904, p<0.01, for group E: r=.850, p<0.01. We therefore assumed that the *speakers scores* were a good reflection of overall quality of the words selected.

We then carried out a *t*-test on the participants' *speaker scores prior* to the training, which showed that pronunciation quality in the two groups was not significantly different at pretest (t=.321, p=.754). We thus proceeded to analyze the participants' *speakers scores* to assess possible improvements in pronunciation quality after the training and whether there were any differences between control and experimental group in this respect. An ANOVA with repeated measures with Test time (pre, post) as within-subjects factor and Training group (C, E) as between-subjects factor indicated a main effect for Test time, with F(1,26)=78.818, p<.05. The overall mean score at posttest (M=6.67, SD=1.26) was significantly higher than at pretest (M=4.49, SD=1.34). No significant effect was found for Training group, nor was there a significant Test x Training interaction. These results indicate that both groups improved in pronunciation quality, and that their improvements were comparable (see also Figure 2).

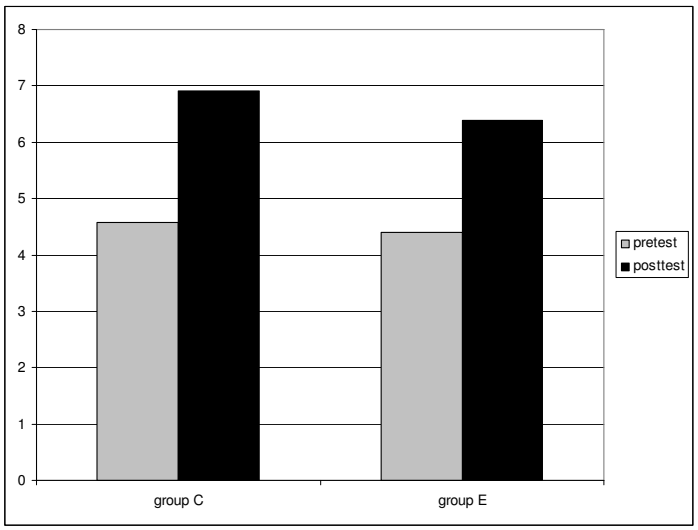

Figure 2: Group *Speaker scores'* means for C and E before and after the training.

In order to gain more insight into the effectiveness of the training provided, we also examined the scores by looking at specific types of words. We asked the English-FL teacher to indicate which of the 28 words might have been more difficult to pronounce for the children, and which were unlikely to have been known to the children before the training. We retained the easy/known words (n=19) and the difficult/unknown words (n=5). For

these words, we combined *single-word z-scores* into one mean per participant, thus obtaining two scores per participant (one for the easy/known words, and one for the difficult/unknown words). We then submitted these scores to an ANOVA with repeated measures involving Test time (pre, post) and Word type (easy-known, difficult-unknown) as within-subjects factor, and Training group (C, E) as between-subjects factor. This analysis revealed a significant effect for Test, with $F(1,26) = 192.176$, $p<.01$. A significant, main effect was also found for Word type ($F(1,26) = 18.805$, $p<.01$) with the mean scores of the easy/known words ($M=5.44$, $SD=0.93$)being significantly higher than those of the difficult/unkown words ($M=4.32$, $SD=1.01$). An Test x Word was also found, with $F(1,26) = 47.620$, $p<.01$, with the pretest mean of difficult/unknown words ($M=3.06$, $SD=1.10$) being significantly lower than those of the easy/known words ($M=5.11$, $SD=1.08$), while at posttest, the means of the two types of words are not significantly different ($M= 5.59$, $SD=0.93$, $M= 5.74$, $SD=.79$, respectively). In other words, pronunciation quality of difficult/unknown words improved significantly after the training, whereas pronunciation of easy/ known words did not. Since no significant Test x Word x Training was found, we concluded that the improvements of the two groups were not significantly different for the two types of words (see Figure3).

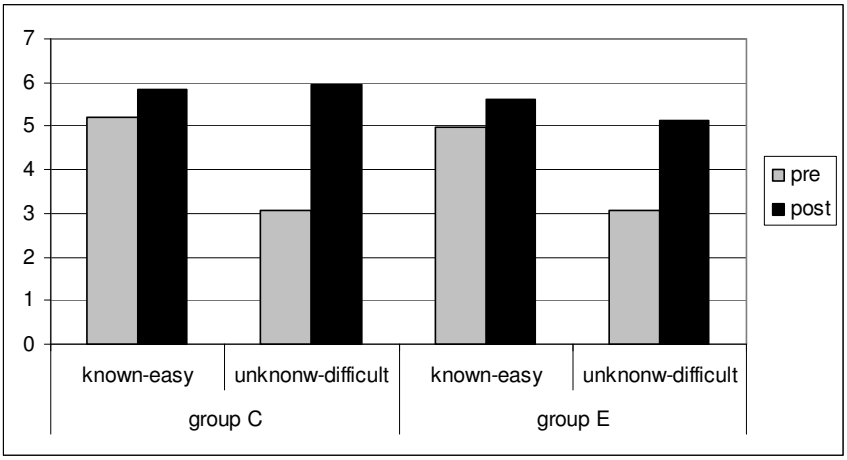

Figure 3: Group means for C and E before and after the training for known-easy words and for unknown- difficult words. The values were obtained by averaging the corresponding *single-word* z-scores.

Conclusions

The results of the analysis on *speaker scores* show that the children who trained with PARLING were able to make improvements in pronunciation quality that are comparable to those made by the children who received teacher-led instruction. This finding is all the more positive considering that the children training with PARLING could only train for 30 minutes per session, while the children in the control group had 60 minutes each time, and that the feedback provided on pronunciation in PARLING was a simple reject/accept response. The analysis carried out on the word-specific scores further indicated that the children in the two groups also made comparable improvements in pronunciation quality of words which they did not know before the training, and which were particularly difficult to pronounce. These positive results might be explained by the fact that the children using

PARLING enjoyed the computer's 'undivided attention' for the 30 minutes of training, while the children training with the teacher were not addressed individually during the 60-minute lesson.

It should nevertheless be pointed out that these results only pertain to the short-term effects of the training, since no delayed posttest was carried out. Besides, the sample size considered here is relatively small. However, the findings from this study give us reason to believe that CALL training that includes ASR is effective in helping children to improve global pronunciation skills. In turn, this suggests that such systems could be used to integrate traditional instruction, for instance to alleviate typical problems due to time constraints or to particularly unfavourable teacher/student ratios, or to help children that are lagging behind by offering them an engaging and more private form of training.

REFERENCES

Bunnel, H.T., Yarrington, D.M., & Poliknoff, J.B. (2000). STAR: Articulation training for young children. In Proceedings of ICSLP 2000, vol. 4 (pp. 85-88).

Eskenazi, M., & Pelton, G. (2002). Pinpointing pronunciation errors in children's speech: examining the role of the speech recognizer. Proposed to the Pronunciation Modeling and Lexicon Adaptation for Spoken Language Technology Workshop, Colorado.

Eskenazi, M. (1999). Using automatic speech processing for foreign language pronunciation tutoring: Some issues and a prototype. *Language Learning & Technology* 2(2), 62-76.

Eurydice (2005) Key Data on teaching languages at school in Europe. Available: http://www.eurydice.org/Documents/KDLANG/2005/EN/FrameSet.htm

Flege, J.E. (1995) Second language speech learning: Theory, findings, and problems. In W. Strange (Eds.), Speech perception and linguistic experience. (pp. 233-277). Baltimore: York Press.

Giuliani, D., Mich, O., & Nardon, M. (2003). A Study on the Use of a Voice Interactive System for Teaching English to Italian Children, in V. Devedzic, J.M. Spector, D.G. Sampson & Kinshuk (Eds.), Proceedings of The 3rd IEEE International Conference on Advanced Learning Technologies (pp. 376-377).

Kawai, G., & Tabain, M. (2000). Automatically detecting mispronunciations in the speech of hearing-impaired children. In Proceedings of InSTIL 2000, 39-48.

Krajka, J. (2001). English for Kids, CALICO Software review, [On-line]. Available: http://calico.org/CALICO_Review/review/englishkids00.htm

Leather, J., & James, A. (1996). Second language speech. In W.C. Ritchie (Eds.), Handbook of Second Language Acquisition (pp. 269-316). San Diego, CA: Academic Press.

Mich, O., Neri, A., & Giuliani, D. (2005). Testing a System that Supports Children in Learning English as a FL: Implications for a More Effective Design. In Proceedings of INTERACT 2005 - WS7, Child Computer Interaction: Methodological Research, Rome, Italy. Available: http://www.uclan.ac.uk/facs/destech/compute/staff/read/Publish/ChiCi/interact/MichNeirGiuliani.pdf

Nicol, A., C. Casey, et al. (2002). Children are Ready for Speech Technology - but is the Technology Ready for Them? In Proceedings of Interaction Design and Children 2002, Eindhoven,The Netherlands.

Purushotma, R. (2005). Commentary: You're not Studying, you're just …. Language Learning & Technology, 9, 80-96.

Sfakianaki, A., Roach, P., Vicsi, K., Csatari, F., Oster, A.-M., Kacic, Z., et al. (2001). SPECO: Computer-based Phonetic Training for Children. In J.A. Maidment & E. Estebas-Vilaplana (Eds.), Proceedings of the Phonetics Teaching & Learning Conference, London: Department of Phonetics and Linguistics, UCL.

Tell Me More. KIDS. By Auralog. Review. [Online]. Available: http://esl.about.com/library/weekly/aa093001a.htm

Wachowicz, K.A., & Scott, B. (1999). Software That Listens: It's Not a Question of Whether, It's a Question of How. CALICO Journal, 16(3), 253-276.

BIODATA

Ornella Mich received a degree in Electronic Engineering from the University of Padova, Italy in 1987. Then she worked in the design group for radar electronics for jet fighter development in FIAR, Milano. Then she taught Electronics and for one year in a high school. Since October 1989 Mich has been working at ITC-Irst. Field of specialization: computer vision (automatic face recognition, image retrieval by content, video processing), multimedia systems, children-computer interaction. Currently she is working on the design, implementation and evaluation of a CALL system for children, based on automatic speech recognition technology.

CONTACT

Ornella Mich
ITC-irst
Via Sommarive, 18Street
POVO - Tn
Italy

mich@itc.it
http://tev.itc.it/people/mich/