

Parling, a CALL system for children

Ornella Mich

ITC-irst, Centro per la Ricerca
Scientifica e Tecnologica
I-38050 Povo (Trento), Italy
mich@itc.it

Diego Giuliani

ITC-irst, Centro per la
Ricerca Scientifica e
Tecnologica
I-38050 Povo (Trento),
Italy
giuliani@itc.it

Matteo Gerosa

ITC-irst, Centro per la
Ricerca Scientifica e
Tecnologica
I-38050 Povo (Trento), Italy
University of Trento,
International Graduate School
I-38050 Povo (Trento), Italy
gerosa@itc.it

Abstract

The aim of this paper is to introduce Parling, a multimodal system developed for supporting learning of English as a foreign language. It is devoted to 8-11 year-old Italian children learning English at primary school. Parling features a set of instructional games that change their vocabulary content dynamically based on the user's learning needs. Moreover, Parling proposes a key feature to involve more young learners: the Make-Your-Story tool, that enables the user to create herself a new story. The technological core of Parling is a speech recognizer that allows implementing automatic pronunciation assessment. The acoustic models are trained by exploiting speech data collected from native speakers of British English in the age range of the target system users. In this initial version of Parling a simple binary response, i.e. an accept/reject response, is generated for each input utterance. A more detailed feedback will be developed in future Parling versions.

1 Introduction

Many interesting CALL (Computer-Aided Language Learning) applications for supporting learning of a foreign language have been developed in recent years. The most appealing ones are based on the use of the state-of-the-art automatic speech recognition and pronunciation scoring technology. Most of these kinds of systems are devoted to adults, while much less effort was devoted to children.

This paper presents Parling (Parla inglese - Speak English), a CALL system for children under development at ITC-irst, whose technological core is a speech recognizer. Parling is devoted to Italian children, aged 8 to 11, who learn English as a Foreign Language (EFL)¹.

□

□ EFL (English as a Foreign Language) is referred to studying English in non-English-speaking countries whereas ESL (English as a Second Language) is referred

Children's English literature is used in Parling to offer young users a way of learning English without stress. They can choose from a list of well-known stories, appropriate for their age and English level. Then, they can listen to the story and read it, play entertaining games with individualized content, try word pronunciations and explore a pictorial dictionary. To allow users to be more active during the learning process, Parling offers also the Make-Your-Story tool that enables them to create a new story by themselves.

The decision of proposing to children good literature written in the foreign language they study is based on the fact that research demonstrated that thematically-based bilingual programs are very successful in language acquisition (Gianelli, 1991). Although the children's own environment (family, school ...) are good sources of theme units, children's literature offers a more complete medium for learning a new language. Indeed, thematic literature contains repetitive patterns that reinforce user's vocabulary and structures. Carefully chosen children's literature allows children to develop their receptive language in an entertaining, meaningful context and naturally invites them to repeat many of the predictable words and phrases. Moreover, literature provides background knowledge and cultural information (Ghosn, 1997), helping teachers to build emotional, social, and intellectual responses to the natural language of engaging stories (Smallwood, 2002).

As research demonstrated, reading aloud to children produces positive effects at any age. Jim Trelease, author of *The Read-Aloud Handbook* (Trelease, 1995), tells us that reading aloud to children "stimulates their interest, their emotional development, their imagination, and their language". Language skills are fostered when children listen to the same stories over and over again until they develop an *almost*

□

to studying English as a non-native speaker in a country where English is spoken.

unconscious familiarity with literary elements and story structure. Their vocabulary grows larger by the story. But reading aloud at school is a highly time-consuming activity; teachers cannot dedicate too much time to it. Systems like Parling can mitigate this problem. After having installed Parling on their own computer at home, children can listen to the story as many times as they like.

Language learning is hard work. Effort is required at every moment and must be maintained over a long period of time. Games help and encourage many learners to sustain their interest and work. Through playing games, students can learn English without being aware they are studying. Vocabulary acquisition is increasingly viewed as crucial to language acquisition. The use of games is considered highly important for presenting and revising vocabulary (Giuliani et al., 2003; Uberman, 1998). For these reasons a lot of games have been inserted into Parling, games that use only words coming from the story.

A brief overview of Parling structure will be given in the next section, followed by the description of the speech recognition technology. At the end, our future work will be presented.

2 System Description

Two important paradigms underline Parling design: the learning-by-playing and the learning-by-doing paradigms (Anzai and Simon, 1979). In synthesis, there is a story as a leading thread, several games aimed at helping students memorize the vocabulary of the story and a tool that allows children write and record new stories.

At the beginning, Parling presents to the user a menu of stories, a list of games, tools to build a new story and a box displaying the alphabet letters, which is the entry to a visual dictionary.

2.1 The story

A list of stories is available to the users. Original versions of famous stories, as *Peter Pan* or *The Three Musketeers*, are proposed in a simplified manner: few short phrases with simple grammatical structures, subdivided in pages. After choosing a story, the child can freely go back and forth through its pages by using a navigation bar. Each time a page is loaded, its corresponding audio is played back. There is also a background music playing all the time.

When the user clicks on one of the anchor words, a window appears that explains the meaning of the words (see Figure 1). Then, by clicking the appropriate buttons the user can hear the pronunciation of the word and she can also try to utter the word herself. The system

will respond with a message telling whether the word was pronounced correctly or not.

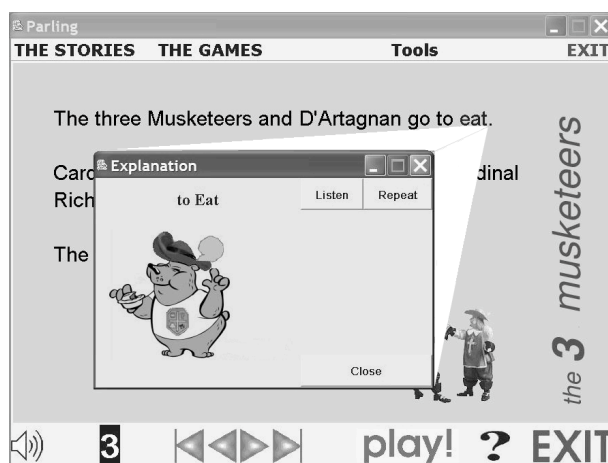


Figure 1. Parling: a story page and an explanation window.

2.2 Adaptive games

Experts in education affirm that learners, and *children* learners in particular, must be active subjects; only in this way the learning process becomes really effective (Piaget, 1983). Following this theory and in order to effectively engage children with the learning of a new language, each story in Parling is associated with a different type of game. For example, *Peter Pan* has been associated with the *Hidden Words* game whereas *The Three Musketeers* with the *Memory* game (see Figure 2).

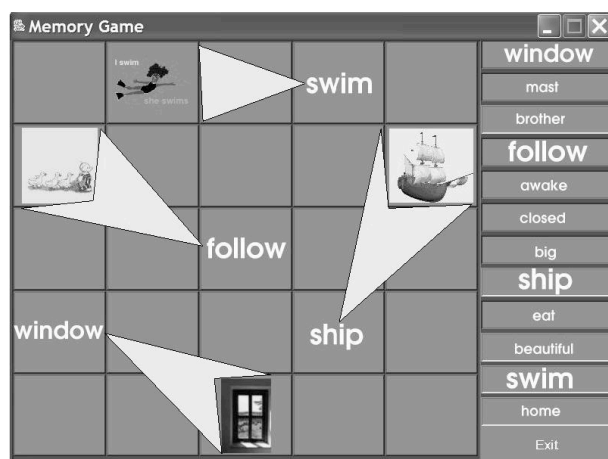


Figure 2. Parling: the Memory game

A play! button on the story page allows the user to start the associate game, each story a different one to help user in memorizing the story vocabulary. The list of words that the game considers is initially extracted randomly from all the words in the story. But this list is dynamically changed every time the user activates a word during the story flowing: the

chosen word is added to the list, while an old one is deleted. Moreover, the system changes randomly the initial word organization each time the game is initialized..

2.3 The *Make-Your-Story* tool

The *Make-Your-Story* tool enables the user, a child as well as a teacher, to create herself a new tale inserting text and images and recording the corresponding audio. This tool is a simple text editor with the additional function of recording short pieces of audio. When the new tale is completed, it is automatically inserted into the system story list. In this way, the user can go back and forth also through their own stories.

2.4 The visual dictionary

The dictionary is one of the most useful tools to enrich one's own vocabulary in a foreign language. In order to provide an effective and easy-to-use tool, we have developed a dictionary based on the truism that a picture is worth a thousand words. Parling Visual Dictionary presents a display of all the letters of the alphabet (see Figure 3).

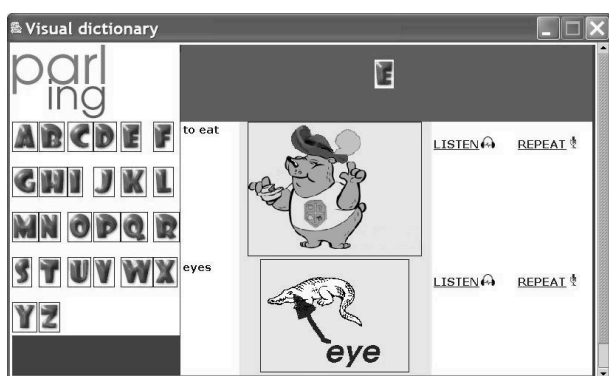


Figure 3. Parling: the visual dictionary.

The user can browse through the dictionary by clicking one of the letters. Each inserted word is presented as text, an auto-explanatory picture, and its reference pronunciation (LISTEN icon).

The user can test her English pronunciation with the REPEAT functionality: a click on the REPEAT icon opens a small window with a traffic light; as soon as the light becomes green the user can utter the word; the system will then respond with a positive or a negative message according to the correctness of the user's pronunciation.

3 Speech Recognition Technology

Characteristics of children's speech vary rapidly as a function of age due to the anatomical and physiological changes occurring during child's growth and because children become more skilled in sound coarticulation with

age (Lee et al., 1999; Narayanan and Potamianos, 2002). Therefore, for reliable recognition of children speech age-dependent acoustic models are usually adopted (Narayanan and Potamianos, 2002; Giuliani and Gerosa, 2003). These acoustic models are trained with speech data collected from speakers in the same age range of the target speakers.

In this work, in order to improve acoustic modeling for the target users of the system, two databases of native and non-native English speech, were used. Both databases were developed within the EC funded project PF-Star (<http://pfstar.itc.it>). Speech data in the first corpus were collected by Birmingham University from native English speakers aged between 7 and 12. Speech data in the second corpus were collected by ITC-irst from Italian learners of English aged between 10 and 11. Children were attending the fifth grade of primary school and had been studying English since the first grade. Each child was asked to read a set of English words and short phrases.

Firstly, a baseline speech recognizer for British English was trained on the corpus of native speech and applied to recognize native and non-native speech. Recognition accuracy achieved for Italian children was drastically lower than that achieved for native English children of the same age. Non-native speech is characterized in fact by pronunciation errors and accented pronunciation that dependent on the level of proficiency in the target language as well as on the cross-language interference between the mother language of the speaker and the target language. By assuming systematic differences between native and non-native speech and by using a small amount of non-native speech from a group of Italian learners of English, acoustic models were adapted to this particular category of speakers. Adaptation of context-dependent hidden Markov models (HMMs) showed to really improve recognition performance on non-native speech. Detailed description of the experiments carried out in order to improve acoustic modeling for non-native speakers is reported in a companion paper in these proceedings (Gerosa and Giuliani, 2004).

At present, manual annotation at phonetic and word level of the non-native speech available is ongoing. This work has been doing with the purpose of identifying the most common pronunciation errors made by Italian learners of English in this age range and with this level of proficiency in English. By exploiting non-native speech and corresponding annotations, we aim at developing a module able to detect most common pronunciation errors in a reliable manner and to produce a suitable feedback to the user.

For preliminary usability tests of the system, we implemented a simple module that, given an input utterance, provided a simple accept/reject response. This was obtained by performing time alignment of the input utterance with the sequence of HMMs corresponding to canonical pronunciation of the uttered text and comparing the obtained likelihood with that achieved by a phone loop network having a suitable weight associated to each phone loop arc.

4 Future Work

In a preceding study (Giuliani et al., 2003) we found that the way of reporting to the user the system response on his pronunciation is really important. Teachers argued that the pronunciation assessment functionality of the system is really useful only if it is reliable and a pertinent comprehensible feedback is given to the user. Being conscious that erroneous feedback has negative effects on the acquisition of L2 pronunciation (Neri et al., 2002), a specific research will focus on effective ways of giving feedback on the pronunciation. Rather than an accept/reject response, we will investigate on how to detect most common pronunciation errors and to produce proper feedback to the user. In addition, recognition accuracy achieved on non-native speech is encouraging and suggests the use of speech recognition technology in voice-enabled interfaces especially designed for engaging young users and fostering their oral production in the target language.

References

- Y. Anzai and H. Simon. 1979. *The Theory of Learning by Doing*, Psychological Review 86:124-140.
- M. Gerosa and D. Giuliani. 2004. *Preliminary Investigations in Automatic Recognition of English Sentences Uttered by Italian Children*. Proceedings of the InStill Workshop, Venice.
- I.K. Ghosn. 1997. *ESL with children's literature. The way whole language worked in one kindergarten class*. Forum 35(3).
- M. Gianelli. 1991. *Thematic units: Creating an environment for learning*. TESOL Journal 1(1);13-15.
- D. Giuliani and M. Gerosa. 2003. *Investigating Recognition of Children's Speech*. Proceedings of ICASSP, Vol. 2, pp. 137-140.
- D. Giuliani, O. Mich and M. Nardon. 2003. *A study on the use of Voice Interactive System for teaching English to Italian Children*. Proceedings of ICAALT 2003, 3rd IEEE Int. Conf. on Advanced Learning technologies.
- S. Lee, A. Potamianos and S. Narayanan. 1999. *Acoustic of children's speech: Developmental changes of temporal and spectral parameters*,

Journal of Acoust. Soc. Amer., 105(3):1455-1468.

- S. Narayanan and A. Potamianos. 2002. *Creating Conversational Interfaces for Children*, IEEE Trans. on Speech and Audio Processing, 10(2):65-78.
- A. Neri, C. Cucchiarini and H. Strik. 2002. *Feedback in Computer Assisted Pronunciation Training: When Technology meets pedagogy*. Proceedings of CALL Conference "CALL professionals and the future of CALL research", Antwerp, Belgium, 179-188.
- J. Piaget. 1983. *Piaget's theory* in P. Mussen (ed) "Handbook of child psychology", Vol. 1, New York: Wiley.
- B. Smallwood. 2002. *Thematic literature and curriculum for English language Learners in early childhood education*. ERIC Digest.
- J. Trelease. 1995. *The Read-Aloud Handbook*. Penguin USA.
- A. Uberman. 1998. *The use of Games for vocabulary presentation and revision*. Forum 36(1).