A Study on the Use of a Voice Interactive System for Teaching English to Italian Children

Diego Giuliani, Ornella Mich and Marianna Nardon
ITC -irst Centro per la Ricerca Scientifica e Tecnologica – Trento, ITALY
{giuliani,mich,nardon}@itc.it

Abstract

In this paper, we report on a study whose objectives were first to investigate the interaction of children with a multimedia system designed for learning a foreign language and featuring automatic speech recognition functionality. Second, we investigated if, using such a system, users meet some particular difficulties in dealing with the microphone or in solving tasks which implies voice interaction. The study was based on the Italian version of “Tell me More. Kids: The City”, a multimedia system dedicated to children aged 8 to 10. The results of the study indicate that, when learners were involved in tasks where they could freely decide if and how many times to use the vocal interaction modality, they did not hesitate to use this modality. The study also revealed the importance of the form of feedback chosen to report the system’s judgment on the learner’s pronunciation to her.

1. Introduction

In recent years, due to the rapid advances of speech recognition technology, a number of systems for computer assisted language learning have been proposed that support interactive speaking practice [1,2,3]. Some of them are explicitly conceived for children [4,5,6].

In this work we investigate the interaction of children with a multimedia system designed for learning a foreign language that features automatic speech recognition functionality, in particular pronunciation assessment capability. Moreover, we investigated if children meet some particular difficulties dealing with the microphone or in solving tasks which entails voice interaction with computer.

The study was based on the use of Italian version of Tell Me More. Kids: the City, by Auralog [5].

2. Experimental environment

Two systems were initially considered: Happy English by Editori Riuniti [4], based on the IBM VIA Voice speech recognizer, and Tell me More Kids: the City by Auralog [5], based on proprietary speech recognition technology. Three teachers tried both the systems. Considering their observations and our experience with both the systems, we decided to use the Auralog’s one.

In order to have homogeneous experimental results, the children had to consider the same four tasks among those available, and to carry them out in an order defined randomly for each child. The four considered tasks were: task 1 - Library, the user can choose among 12 words, which ones to listen and/or to utter, having feedback about the goodness of her word pronunciation. Three types of feedback are proposed: the waveform of the uttered word is displayed, for comparison purposes, together with the waveform of the reference utterance of the system; the expression (sad or happy) and the colour (green or red) of an animated character, a parrot, change according to the degree of goodness of the learner’s pronunciation; the expression (more or less smiling) and the position (craning his neck) of a clown change according to the system’s judgment; task 2 - Hidden Words whose objective is to find 8 words hidden in a square full of characters (10 x 10) and repeat their pronunciation. For each word, the user had text, a picture and audio; task 3 - Story, whose objective is to re-order four pictures representing a story and then repeat the sentence associated with each picture; task 4 - Color whose objective is to color a picture, uttering the appropriate color for each one of the 12 parts.

25 subjects, 15 females and 10 males, aged from 8 to 12 (average 10.3), participated in the experiment. Children were classified according to their English expertise: 15 beginners, who have studied English for 3, 4, or 5 years, and 10 starters, with 0, 1, or 2 years English study.

The experiment followed a 4 (task) x 2 (English expertise) x 4 (task order) mixed design (both between-subject and within-subject factors).

3. The study results

Because the Library task is completely different in objectives and modality of execution from the other three, we decided to analyze it separately. Regarding the relation between the number of listening per word and the English
level, a Univariate Analysis of Variance reveals a significant difference between the two groups of starter and beginner students of English, \( (F = 5.203, df = 1, p = 0.000) \) and also between the words and the English level \( (F = 1.512, df = 11, p = 0.000) \). On average, starter and beginner learners listened to each word 1.7 and 1.0 times, respectively. Regarding the number of repetitions per word, a Univariate Analysis of Variance reveals a significant difference among the 12 considered words \( (F = 5.153, df = 11, p = 0.000) \), but not a significant difference between starter and beginner learners, see Figure 1. On average, learners used the voice interaction modality 14 times. Compound or complicated words were selected fewer times. We observed that both starter and beginner users listen at least once to a word before uttering it.

![Figure 1. Mean number of word repetitions in Task 1 – Library, considering English expertise](image)

During the experiment we did not highlight relevant problems with the use of headphones and microphones. In general, all children were totally concentrated during the experiment. They were anxious to start and did not appreciate lengthy descriptions of the tasks. A lot of children uttered the word they should say several times in a low voice before activating the microphone. Someone was sceptical about the feedback provided by the system in relation to the right pronunciation, because they did not understand what they changed in the word pronunciation when receiving a positive feedback the second time they repeated the word.

### 4. Conclusions

In this work we investigated some aspects that emerge when children use multimedia software for foreign language learning based on speech recognition technology.

While the three teachers and the twenty-five children that tested the system considered their experience with the software positive, some criticisms emerged. Teachers pointed out that the functionalities of the system should be designed in order to meet the learner's needs. Furthermore, 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. This was confirmed by our experiments. In general children enjoyed speaking into the microphone. However, quite often they were not able to understand why the system judged their utterances as mispronounced.

Speech recognition technology has a great potential for application in the field of computer assisted language learning. However, in designing system functionality, limitations of this technology should be considered.

### References


