propor 2008

International Conference on Computational Processing of Portuguese Language

Applications of Portuguese Speech and Language Technologies

Applications of Portuguese Speech and Language Technologies - Propor 2008 Special Session

Hosted by:

Universidade de Aveiro

Promoted by:

Microsoft Language Development Center

Propor 2008 Special Session Commitee

Special Session Chair

- **António Teixeira -** DETI/IEETA, Universidade de Aveiro, Portugal

Organising Committee

- **Daniela Braga,** Microsoft Language Development Center, Portugal
- **Miguel Sales Dias,** Microsoft Language Development Center, Portugal
- **António Teixeira -** DETI/IEETA, Universidade de Aveiro, Portugal

Programme Committee

- **António Teixeira -** DETI/IEETA, Universidade de Aveiro, Portugal
- **Daniela Braga,** Microsoft Language Development Center, Portugal
- **Vera Strube de Lima,** Pontifícia Universidade Católica do Rio Grande do Sul, Brasil
- **Luís Caldas de Oliveira,** INESC-ID/IST, Portugal

Editorial Board

- **Daniela Braga,** Microsoft Language Development Center, Portugal
- **Miguel Sales Dias,** Microsoft Language Development Center, Portugal
- **Luanda Braga Batista,** Microsoft Language Development Center, Portugal

A Computer-Assisted Learning Software to Help Teaching English to Brazilians

Nelson Neto, Ana Siravenha, Valquíria Macedo and Aldebaro Klautau

Universidade Federal do Pará, Signal Processing Laboratory,
Rua Augusto Correa. 1, 660750110 Belém, PA, Brazil
{nelsneto,siravenha,vgmacedo,aldebaro}@ufpa.br
http://www.laps.ufpa.br

CAI (Computer Assisted Instruction) is a learning process with the aid of computers. In the last 40 years, there was an exponential growth in the use of computers as providers of instructions. During this fast technological revolution, the CAI has become more refined and computers turned into a sophisticated, and why not say essential, instrument of aid, to the extent that they invented a new way of teaching and learning, involving more easily the student in the universe intellectual. Over the years more and more systems are incorporating CAI multimodal interfaces such as sound, image, video and interactive programs that support a wide range of applications. More recently, started up the incorporation of speech interfaces in environments called CALL (Computer-assisted Language Learning).

Despite creating new possibilities for the development of learning tools, the practical impact of CALL applications in education of foreign languages has been quite modest, consequence of evaluations and interfaces without standardization, and the absence of concrete evidences proving the educational benefits of these systems [CHA 97]. But experts, as in [WAR 96], say that with the drastic and rapid invasion of computers in schools and homes, the languages teachers need, as soon as possible, revise their concepts and work in the search of mechanisms that help to overcome the obstacles set by the limitations of these speech technologies.

In recent years, the performance of "personal computers" has evolved with the production of ever faster processors, a fact that encourages the implementation of speech processing algorithms in real time and economically viable. The integration of high-quality audio, high-resolution images and videos with the output of text and graphics of conventional computers has generated recent multimedia systems that provide powerful "training tools" explored on CALL. The results of these efforts are prototypes with speech interfaces for training the human pronunciation in a wide variety of languages, texts reading and foreign languages teaching in pre-defined contexts [EHS 98].

An important area within the speech processing is the training of pronunciation. Intelligent tutors interact and guide students in the repetition of words and phrases, or reading sentences, in order to practice both the fluency (quality of the phonetic pronunciation), as the pitch (manipulation of the parameter [pitch]) in a certain language. These prosodics characteristics are particularly important in the language learning processing, since a incorrect prosody can block the communication between the speakers. Errors in intonation can give to the listener a misconception impression of the attitude taken by the person who is speaking, as well as a sentence without fluency makes it virtually impossible for the listener understands the dialogue context.

There is not a consensus among the linguists about the presence of teachers or tutors in the process. To [FRA 00], pronunciation training softwares must be employed to enhance the interaction between teacher and student, where the tutor must concentrate their efforts on speech characteristics that may affect their comprehensibility, while [ANA 03], believes in replacement of "human tutors", and suggests that the students can follow their own learning rhythm. However, linguists may attend the technicians, suggesting which speech aspects of the practitioner need to be focused and stipulating limits for the shunting lines identified in the pronunciation.

A widespread style of CALL, especially in schools of languages, makes use of an high-level interface and grammar set within a context, regular situations in real life, to practice a foreign language [CCA 98]. Basically, the student must choose a response within a limited number of alternatives shown on the screen, or may be challenged to answer a question without any help from the software. These methods are called closed response and open response, respectively.

The proposed application is able to merge the use of these two methods by foreign language learning (English). The student is introduced to respond orally, or even manually, the questioning by written, spoken and visual stimulations. Besides the proposed objective and subjective exercises, there is the possibility of a prior training, where students can enhancement their vocabulary by listening Text-To-Speech individualized words or phrases, always with the respective visual return, illustrating the meaning of the word, or the action associated with the sentence.

There are a large number of engines available for English, which is one of the reasons for the development of this system, however the existence of engines for others languages, including Portuguese, makes possible the implementation of a software CALL for foreigners. Therefore, changes on the source code and in the structure of the original software components are essential.

The personal agent, "Merlin", created from the MSagent [MSA 05] component, provides feedback and, when necessary, assistance. The agent returns verbally and textually the responses, "excellent" or "wrong", depending, clearly, of the reply sent by the student via keyboard or mouse. The Merlin also participates in the "help option" created to assist the user during the exercises. To call this function, the students must pressure the corresponding button, then tips will be presented to guide them in the attempt of resolution the question.

Using engines provided by Microsoft for ASR (Automatic Speech Recognition) and TTS (Text-To-Speech), it is possible to interact with the user through the English language. However, the familiarity with the interface by the user, was very dependent on TTS in the native language, in our case the Portuguese language, for instructions and feedback. This situation is shared with other projects under development in Brazil (e.g., [VEL 04]). Thus, in the beginning of the project development, a multilingual structure was developed with the incorporation of the CSLU (Center for Spoken Language) toolkit.

The CSLU toolkit [CSL 05] is a platform for dialog application development. The software provides four interface levels for applications development. For example, developers can use code written on the C# language to access. However, there is no support for a generic API (Application Programming Interface) such as SAPI (Microsoft Speech API) [SAP 05] or JSAPI (Java Speech API) [JSA 05]. The approach adopted to synthesize Brazilian Portuguese texts using the CSLU toolkit did not required changes in the CSLU code. Batch files that called the CSLU TTS were used but the solution is inefficient (obviously, reading and writing from/to disk was rather slow). This strategy could be used in other situations but was abandoned in favor of a completely SAPI-based solution employing Microsoft Agents.

Since the characters Genie, Merlin, Peedy and Robby are compiled to use the English TTS engine L&H (Lernout & Hauspie) TruVoice as default, it was necessary the installation of L&H TTS3000 for Portuguese language, also licensed by Microsoft. Besides, the L&H engine has proved essential for the installation of Portuguese language components. These components are libraries (DLL files) that add support for synthesis on a given language. With them it is possible to change the characters TTSModeID property.

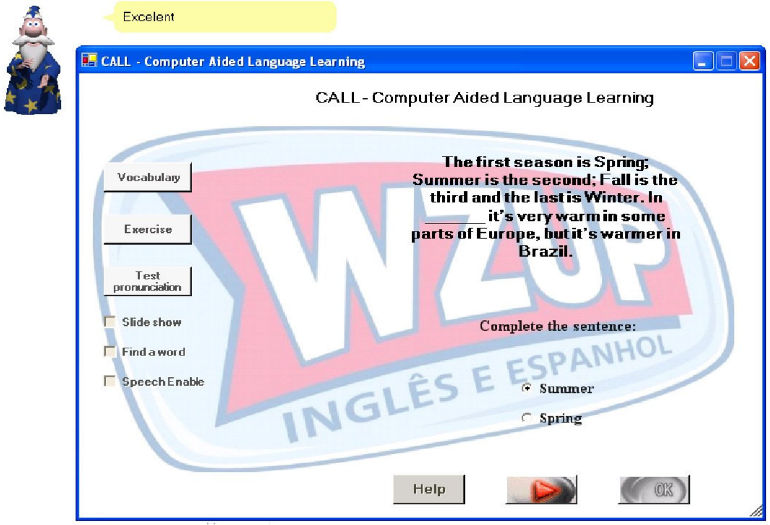
This work presents a computer-assisted software for English language learning. As mentioned, all the speech interfaces (recognition and synthesis) were implemented using the Microsoft Speech Application Development (SAPI 5.1) toolkit. Also, adopting the Microsoft provided engines: Microsoft English Recognizer 5.1 and L&H TTS, the system can interact with the user through English and Portuguese languages. Another feature is the custom agent, Merlin, a Microsoft Speech Agent component that provides feedback and assistance.

The software consists of an executable file containing ten (10) modules. Each module corresponds to a lesson, which contains three (3) sections: vocabulary, exercises and pronunciation testing. The user is invite to respond orally, or even manually, the questioning by written, spoken and visual incentives. Besides the proposed objective and subjective exercises, there is the possibility of a prior training, where the user is asked to enrich his/her vocabulary by listening individual words or phrases synthesis, always helped by a picture, which illustrates the meaning of the word, or the action associated with the sentence. The sections are better described below:

- Vocabulary: stage where the user listening the words or phrases via text-to-speech, is invited to provide the transition between the sentences by hand. With the intention to make the transition more comfortable by the user, it was developed the "slide show" function. Still within the vocabulary section, there is a routine called "find a word" that, intuitively, allows the user to search for a word in the whole lesson and listen to it on its correct pronunciation.

- Exercises: tasks are proposed to test the grammar knowledge developed by the user. The software only moves forward if the user correctly answers the proposed question. If the answer is incorrect, the user is alerted by the agent, with the possibility of changing his answer. In case of doubt, the user can use the help option, where a synthesized sentence will help to choose the correct answer, showed in Figure I.

- Pronunciation testing: based on the vocabulary, up to twenty five (25) words are simultaneously made available so that the user can test his/her pronunciation by using automatic speech recognition. When the word is properly recognized, a corresponding visual illustration is shown. The recognition is possible through activation of a grammar with restricted rules (command-and-control). The confidence level control can be adjusted by the user. The available options are: easy, normal and difficult. Intonation and rhythm plays an important role in language learning but they are not referred in this work.



The first season is Spring; Summer is the second; Fall is the third and the last is Winter. In ________ it's very warm in some parts of Europe, but it's warmer in Brazil.

Complete the sentence:

● Summer

○ Spring

Figure I: One of the screens of computer-assisted software for English language learning.

The ASR grammar used in the implemented CALL software is the simplest possible (in the sense of a minimum *perplexity).* As the recognition is always focused on contexts (limited options for response), the grammar is simple and the system is robust even if not trained for non-native speakers.

The current version of the software does not control or monitor recognition errors. Informal tests showed that the determining factors for the accuracy are: the training stage suggested by the engine during the installation procedure, the use of "head-mounted" microphone with a constant distance between the mouth and microphone, and the amount of environmental noise during the test (execution) stage.

References

1. [CHA 97] CHAPELLE, C. **Call in the year 2000: Still in search of research paradigms?** Language Learning and Technology, 1(1), 19-43, [S.l.], 1997.
2. [WAR 96] WARSCHAUER, M. **Computer-assisted language learning: An introduction. In S. Fotos.** Ed., Multimedia language teaching, pp. 3-20. Tokyo: Logos International, [S.l.], 1996.
3. [EHS 98] EHSANI F. and Knodt E. **Speech Technology in Computer-Aided Language Learning: Strengths and Limitations of a New CALL Paradigm.** Language Learning and Technology Vol. 2, No. 1, July, pp. 45-60, 1998
4. [FRA 00] FRASER, H. **Phonetics, phonology, and the teaching of pronunciation - a new cd-rom for all esl learners and its rationale.** Eighth Australian International Conference on Speech Science and Technology, pp.180-185, [S.l.], 2000.
5. [ANA 03] ANANTHAKRISHNAN, K. S. **Computer Aided Pronunciation System (CAPS).** University of South Australia, 2003.
6. [CCA 98] CCAA. **CALL Computer-assisted Language Learning.** CCLS PUBLISHING HOUSE, C1626AK11 ISBN 85-340-0602-4, 1998.
7. [MSA 05] http://www.microsoft.com/msagent.
8. [VEL 04] VELHO, L. and Rodrigues, P. L. and Feijó, B., **Expressive Talking Heads: uma ferramenta de animação com fala e expressão facial sincronizadas para o desenvolvimento de aplicações interativas.** Proceedings of Webmídia. SBC, 2004.
9. [CSL 05] http://cslu.cse.ogi.edu/toolkit/.
10. [SAP 05] http://www.microsoft.com/speech/.
11. [JSA 05] http://java.sun.com/products/java-media/speech/.