SPEECH ENABLED SERVICES IN A WEB-BASED e-LEARNING ENVIRONMENT

Matthias Eichner*, Marc Göcks**, Rüdiger Hoffmann*, Matthias Wolff*

*Technische Universität Dresden, Mommensenstraße 13 01062 Dresden, Germany
**Brandenburgische Technische Universität Cottbus Universitätsplatz 3-4 03044 Cottbus, Germany

Abstract

In this paper we introduce a project conducted by a consortium of German Universities which aims to support presence teaching by a web-based e-Learning environment and the respective content from the field of information, communication and media technology. In a first part we will define requirements on e-Learning and describe a design of e-Learning modules fulfilling these requirements. We will also give a short overview of the implemented e-Learning modules and their multimedia enabled content. In the second major part we describe the technology and deployment of speech enabled services – speech recognition, speech synthesis and speaker verification – in the e-Learning environment and discuss first evaluation results of these technologies.

Key Words
E-Learning, WBE, Speech Enabled Services, Multimedia, Web-Technology, eL-IT

1. Introduction

In the course of an intensified establishment of new media and efficient information and communication technologies a considerable number of new applications of e-Learning within the university have opened up [1]. The chances and needs resulting from this must be well considered. New technical and media-didactical concepts for e-Learning applications can be realized utilizing new technical resources more effectively than ever before. A systematic didactical research is necessary for the acceptance of the new media in the university as well as for a supplement and a partly substitution of the traditional learning paradigm of presence teaching by time and local independent teaching and learning forms. However, a complete replacement of presence teaching in favor of a pure virtual university remains a rather unrealistic vision of the future [2] and is, in the last consequence, also not desirable. But using new media in teaching offers a considerable new flexibility and added value for both, teachers and students.

The development, deployment and support of e-Learning applications causes high expenditures. Without an appropriate personnel and technical equipment a majority of the new chances of e-Learning cannot be realized. Conclusively, a financial support is compellingly necessary for the realization of the described concepts and visions - today and in the future. In that respect support or funding programs like those of the German Federal Ministry of Education and Research BMBF ("new media in the education") are a valuable initiative [2]. Accordingly, the question for a long term sustainability becomes increasingly important.

2. Presentation of the „eL-IT“ Project

With regard to the initiative "new media in the education", the research and development project eL-IT (e-Learning modules for curricula in information, communication and media technology) is promoted by the BMBF during its operational period running from 01.06.2001 to 31.12.2003. Within the eL-IT project we develop and deploy interactive, multimedia e-Learning modules for university courses in information, communication and media technology. The emphasis of the project is thereby in the web-based content development and multimedia content preparation. The conversion of spreading infrastructure concepts is coordinated closely with the central objectives of the project.

The research and development project is operated by an association of universities consisting of the Technical University of Dresden, the Technical University of Berlin, the University of Kassel and the Technical University of Brandenburg in Cottbus. Together, the seven chairs at the four partner universities develop fourteen e-Learning modules.

First of all, the developed e-Learning modules serve the current presence teachings and, moreover, represent the basis for a planned, independent master or postgraduate study course. This interstate course of studies shall be
offered in future by the university group as combined presence and distance course of studies. The entrance to this course of studies should be equally possible by all partner universities.

All e-learning modules represent adequate and accepted test credits, with appropriate Credit POINT evaluation assigned to them. The extent of a module includes usually 4 SWS (hours per week) and is equal to at least 4 Credit POINTS. Thus the mutual exchangeability and acknowledgment of the study achievements are to be guaranteed within the project and in the international comparison. A project-accompanying framework concept and basic conditions for the module development with obligatory quality assurance guidelines are to form the conditions for a high success in learning and a high acceptance in the target group. The attractiveness and quality of the teachings are to be clearly improved by the broad introduction of innovative multimedia learning and didactical forms with the universities.

The Technical University of Brandenburg operates as an Application Service Provider for the learning management system, which is the central platform for the developed modules and in future forms the basis for the virtual learning environment of the planned master course of studies.

2.1 Requirements on e-Learning

e-Learning is often used as a generic term for web-based teaching and learning programs [3]. In the eL-IT project we shall attempt to fulfill the following minimum list of requirements and guidelines for the e-Learning environment and its learning modules (c.f. [4],[1],[8]):

- The educational content taught is enhanced by the integration of multimedia elements like animations, video, graphics and audio elements.
- Communication channels between the students are implemented by, e.g., creating chat rooms, discussion rooms, or boards.
- Consulting of a teletutor for a personal support via e-mail or video conferencing is possible.
- The e-learning modules are parts of a network which integrates a huge variety of information resources.
- The student himself and his ability to receive individually tailored information is the centre of interest.

The conversion of the specified requirements is to be interpreted in each case of the module-specific context.

3. Project Specialties

Under the supervision of the "video technology" group at the multimedia center of the Technical University of Brandenburg (BTU Cottbus), different practical operations and usage scenarios from the field video technology were tested and realized. In selected modules potential practical operations are analysed by groupware solutions. Moreover ranges of application are tested and evaluated by 3D-learning environments. By the combination of presence and distance forms of the study, particularly female students shall explore new possibilities to flexibly combine their studies with the occupation and/or family life. A particular study has been conducted in order to identify potential, gender specific differences with the preferences of e-Learning-applications [7]. 3D-learning environments and speech technologies will be described in detail in the due course.

Another project focus lies on the integration and testing of speech enabled services – like speech recognition, speech synthesis and speaker verification – in the e-Learning environment and modules. The underlying technology was supplied by the speech processing workgroup at the Institute of Acoustic and Speech Communication of the Technical University of Dresden.

3.1 Testing of 3D Learning Environments

In the context of developing modules various scenarios and possibilities for 3D learning environments are tested. For the module "multimedia law" a three-dimensional, virtual law court has been programmed based upon a real life district court in Berlin. VRML (Virtual Reality Modelling Language) serves as the basis for the three-dimensional description of rooms and physical devices. It enables the student to move and/or navigate within the building in order to enter topic-specific court rooms to receive information on particular cases, background information, issues of law texts and contents as well as working cases and help manuals.

Figure 1: 3D-entrance hall, eLearning-module “multimedia-law”

Within each of the fourteen existing chapters contains a data-base supported paragraph search program called “§Finder”. Whenever the student finds a reference to a paragraph without explanation, he can type the topic into
an integrated search mask to receive the exact legal text. Beside these relevant materials for the exams and tests further more interesting information’s are offered to the user.

With the conception of the 3D learning environment, priority has been given to the creation of contrasts between the historical, traditional legal doctrine and the modern requirements on the multimedia-law. That idea is reflected particularly in appearance of old-venerable space metaphors (figure 1), which are in contrast to current contents of the multimedia-law. The intention of these provoked contrasts is the creation of areas of conflict, in order to energize and arrange knowledge transfer more exciting.

4. Speech Enabled Services in eL-IT

At the partner university in Dresden different scenarios for application of speech technologies in eL-IT have been developed. In the second part of the article the integration of speech services in the e-Learning environment will be introduced. The following scenarios have been investigated:

1. Voice control of the e-learning environment (speech recognition),
2. Voice control of multimedia environment in lecture rooms (speech recognition),
3. Speech output functionality of the e-learning environment (additional content and read-aloud service), and
4. Speech enabled online exercises and tests (speaker verification, speech recognition, speech synthesis and dialogue control).

For seamless integration of speech technology into web pages a client/server based architecture is applied (figure 2). The content of the e-Learning module is presented to the user in a web-browser. In the web page, a Java Applet (Speech Applet) is embedded that implements audio input/output capabilities. The applet communicates with a server (Speech Server) that provides the speech services speech recognition, speaker verification and speech synthesis. Using these services voice control and speech output can be easily integrated into existing web-pages. The only requirement on the client side is a JavaScript enabled browser and Sun’s Java-Plugin version 1.4.

The user interface of the Speech Applet consists of a recording level indicator and a button for deactivating the applet. The level indicator gives the user important feedback about the correct adjustment of the sound interface. Also important is to inform the user whether the Speech Applet is recording, playing or recognizing. This information can be easily integrated into the layout of the web page using standard HTML elements. A JavaScript library supports web page developers to track the status of the applet, to define a vocabulary for the recognizer and to associate events to the recognized commands. The possibility to define separate vocabularies for each web page improves the performance of the system, since only commands needed for the current page are activated in the recognizer. This offers a powerful way to define the dialog model implicitly by the structure of the pages.

In figure 3 a voice enabled multiple choice test is shown. The speech applet is displayed in the lower left corner of the page. The user can navigate within the test by naming the question and the corresponding answer. All speakable commands are listed in the navigation frame on the left.
In case of incorrect user commands, i.e. naming the answer before selecting a question or trying to finish the test without answering all questions, the user gets a reply by a synthetic voice. The integration of speech synthesis allows a read aloud functionality of arbitrary text in the web page. The user simply marks the text and starts the synthesis by clicking a button.

Another scenario for integration of speech technology is the voice control of devices used in lecture rooms. The operation of devices like beamers, switches, cameras etc. by the lecturer using IR remote controls is not feasible due to the number of devices and the distance between lecturer and equipment. Controlling the devices by voice commands is desirable and would be the most convenient way of human-machine-interaction. At the Dresden University of Technology a device control has been developed and published under GPL that complies with the following requirements [20]:

- flexible and easily adjustable user interface (HTML)
- devices are controlled via RS232
- local control and remote control via internet possible
- OS independent (client and server)

This software can be combined with a locally installed speech recognizer or the Speech Applet described above to enable voice control of the devices.

We also tested the integration of speaker verification technology in the e-Learning environment. Speaker verification can serve as additional security feature besides conventionally login/password queries. A demonstration of this technology is used as an example for a biometric coding procedure in chapter “security aspects of eCommerce” in the e-Learning module eCommerce.

4.1 Speech Recognition

For application in an e-Learning environment the speech recognizer needs to be speaker independent and no initial training (per user) should be necessary. The vocabularies for controlling web pages consist of a limited set of commands and command phrases only, but vary depending on the design and content of the page. Therefore fast vocabulary switching and no restriction in the choice of command phrases are additional requirements. The recognizer has been developed within a project for integration of speech recognition and speech synthesis into a unified system [19] and is supported by the Deutsche Forschungsgemeinschaft (DFG). The system uses a generalized form of HMM’s for acoustic modelling and achieves a recognition rate of 60% for unconstrained phoneme recognition [13].
In the given scenario the speech input consists of a mixture of arbitrary speech and single commands that must be distinguished by the recognizer. Therefore a sophisticated rejection scheme is necessary to minimize the number of commands recognized by mistake. Two rejection measures were implemented basing on the comparison of the unconstrained phoneme recognition result and the result under the constrained of the given vocabulary. The relative acoustic score match (RASD) is computed from the acoustic scores and compares both results on the acoustic level. The smaller the RASD, the higher is the confidence of the recognizer. The second measure is derived from the number of matching symbols in both results and is called relative phonetic match score (RPMS). The higher the RPMS, the better is the confidence.

The recognizer was evaluated using a vocabulary of 21 command phrases for navigation and solving a multiple choice test. The test set consisted of 301 utterances from 8 speakers where 113 utterances contained valid command phrases. The measures included the word error rate (WER), the false acceptance rate (FAR, a command was recognized by mistake), the false rejection rate (FRR, a command was spoken but has been rejected) and the equal error rate (EER, FAR equals FRR). WER, FAR and FRR depend on the thresholds chosen for RASD and RPMS. In figure 4 and 5 the results of the evaluation are shown. The operating point was specified by setting RASD=0.05 and RPMS=0.25. Using these thresholds a word error rate of 13.3% and an equal error rate of 10% were achieved.

4.2 Speech Synthesis

DRESS - the Dresden Speech Synthesizer is a multiphone-based time-domain synthesis platform, which enables the conversion of any input text into a synthetic speech signal. The system has been developed at the Dresden University of Technology [18]. DRESS may be adapted with regard to the desired speaker, the voice type, the prosodic style and additional languages. It includes modules for German, Chinese, US English, Italian, Czech and Russian. The system design considers language-independent principles and focuses on a high flexibility. The system includes e.g. seven different modules for intonation and duration control and enables the learning of language- and speaker-specific prosody parameters.

4.3 Speaker Verification

The speaker verification system that is used in the speech server has been developed within European Co-operation in the field of Scientific and Technical Research (COST250)[16]. It can be used free of charge for education and research. The system has the following properties:

- Signal analysis: LPC-Cepstrum
- Vector quantization
- One acoustic model per speaker + one speaker independent model
- Text independent
- Language independent

The performance of speaker verification systems is measured by the false rejection rate (FRR) and the false acceptance rate (FAR) [14][15]. In figure 6 the FRR and
FAR are shown depending on threshold. This threshold is used to adjust the operating point of the system.

Conclusions

Despite the large variety of different developments and the substantial broad content information during the content preparation, the target group orientation is centre of attention for all efforts undertaken. However, analysis results of extensive module evaluations will answer the question whether students will understand this offer as an additional incentive or added value. The intention of provocatively oppositions shall provide areas of conflict in order to help stimulate the knowledge transfer and to create a more exciting environment.

References