Evaluating the Use of Speech Technologies In the Classroom: The APEINTA Project

Javier Jiménez Dorado
Spanish Center for Captioning and Audio Description (CESyA),
Carlos III University of Madrid, Spain
jjdorado@cesya.es

Pablo Revuelta Sanz
Spanish Center for Captioning and Audio Description (CESyA),
Carlos III University of Madrid, Spain
prevuelt@cesya.es

Lourdes Moreno Lopez
Computer Science Department
Carlos III University of Madrid, Spain
aiglesia@inf.uc3m.es

Ana Iglesias Maqueda
Computer Science Department
Carlos III University of Madrid, Spain
lmoreno@inf.uc3m.es

Abstract: APEINTA (Spanish acronym for *Proposal Aiming for an Inclusive Education based on Assistive Technology*) is a Spanish educational project that aims for inclusive education for every student of all abilities in and outside the classroom. The APEINTA project is focused in two main inclusive proposals: One *In the classroom* and the other one *Outside the classroom.* This paper is focused on a subjective evaluation of the *In the classroom* proposal, where Automatic Speech Recognition (ASR) and Text to Speech (TTS) technologies have been used in order to avoid communication barriers among teacher and students. The project has been evaluated during different classes in 3rd course of Computer Science degree at the Carlos III University of Madrid, and during a talk in the ACAPPS (Federation of Families and Deaf People of Catalonia) Congress.

Introduction

Nowadays, the use of new technology is growing in every field of education. Teachers are more and more introducing these digital processes and contents in their lessons and there are many resources related to learning via internet. This technology can be useful not only for disabled students, older adults or foreign students but it also can serve as learning resources for everybody, independently of his/her abilities. The proper use of the technology permits to support the student learning by adjusting the pedagogical resources and strategies to each student individually.

Technology and multimedia contents are very useful tools for education and learning process. However, there are barriers for some students that must be considered if we want to ensure universal access to education and a complete integration of students with disabilities.

Historically, students with disabilities have experienced inadequate access to lecture material in the classroom, and insufficient access to the academic resources necessary to sustain their progress. If an equal access to educational resources is guaranteed, following the Universal Design Principles (Connell et al. 1997), then we can avoid students feeling that their learning capacity is limited due to a possible inaccessibility to the educational resources. Inclusive methodologies have to be followed to reach these objectives. These methodologies need to be carried out following standards such as Spanish regulation for television captions (AENOR 2003), eXtensible Markup Language (XML) standard, widely used software, etc.

One of the most useful tools used in inclusive education is the Automatic Speech Recognition (ASR). This technology is especially useful in the classroom for the hard of hearing or deaf students. It is also useful as a tool for taking notes for visually impaired students or those with cognitive or physical disabilities (Bain, Basson & Wald 2005).

Sign Language interpreters have been traditionally used during the class for deaf students' inclusion. However, nowadays not every student with hearing disability uses Sign Language, thus signing does not provide deaf students full access to classroom information (Marschark et al. 2005). Taking this into account, Sign Language is not a global solution for the inclusive education of the deaf.

The APEINTA project aims for providing accessibility in education in and outside the classroom, bearing the functional diversity of the students. In the classroom, the APEINTA project uses ASR to provide personalized real time transcription and to create automatically digital educative resources. It also uses a Text To Speech (TTS) system to allow students with locative problems to type questions which are played loud. Meanwhile, outside the classroom, a website allows students to download accessible pedagogical resources, some of them built during the class thanks to the ASR and TTS services.

This paper is focused on evaluating the use of ASR and TTS in the classroom by surveying the students. It is structured as follows. First of all, it shows some related works. Secondly, it details the global architecture of the APEINTA project. Thirdly, it describes the evaluation of the In the classroom proposal and the results obtained. Finally, conclusions and future works are shown.

Related Work

There are different groups around the world that have worked or are working in lectures transcription based on automatic speech recognition. All of them demonstrate the benefits of speech technology in the new educational framework. Not only for students with disability but also for every person involved in the educational process.

The Liberated Learning Project (LLP) is one of the most active researches (Saint Mary's University 1999), (Leitch & MacMillan 2001). The LLP has collaborated with IBM to develop the ViaScribe software based on the ASR ViaVoice (VV) software (Bain, Basson & Wald 2005). ViaScribe improves readability by detecting pauses in speech and inserting sentence and paragraph breaks. Moreover, it provides phonetic spellings when the recognition is uncertain. And it has a less-accurate speaker independent mode to accommodate multiple speakers (Bain, Basson, Faisman & Kanevsky 2005).

There are other initiatives based on other ASR systems as the VUST initiative, from the Villanova University, which uses the Microsoft Speech Recognition Engine (MSRE) (Kheir & Way 2008). Some initiatives have developed their own ASR system, instead of using commercial ones. That is the case of iCAMPUS developed in the M.I.T. (MIT 2003), the CHIL Project (Lamel et al. 2006) or the LECTRA Project (Trancoso et al. 2008) for European Portuguese.

APEINTA Project is the first initiative in real-time transcription of spoken Spanish lectures. It also provides students with different methods to see transcription within individual devices. And it also can be considered as one of the few projects that pay attention to speech problems using text to speech tools.

APEINTA Project

The APEINTA Project (CESyA 2010) was developed in Carlos III University of Madrid within collaboration between the Spanish Centre of Captioning and Audio Description (CESyA) and the Advanced Database Laboratory of the Computer Science Department.

APEINTA is divided into two distinct parts but dependent on each other. The first part takes places in the context within the classroom and the other outside. The architecture of the multidevice system has been designed as a client/server platform (Fig. 1). In the classroom, there are a real-time transcription and a text to speech (TTS) system to avoid communication barriers for deaf and hard of hearing students, older adults, foreign people or students with speaking problems, for example (Revuelta et al. 2008). The real-time transcription is used to generate automatic educational resources such as audio and synchronized subtitles or notes in different formats. 0ut the classroom, a Web platform provides access to the automatically generated resources as well as other digital content such as video, slides and other documents. The Web platform adapts contents to students' needs, obtaining benefits in their learning process (Moreno et al. 2008).

This paper is focused in the *In the classroom* module. Firstly, the server transcribes the teacher's speech with the help of an ASR system, converting the spoken lesson into a digital resource in real time. This resource is available for students in form of teletext captioning (called "teletext" mode) or as plain text, in paragraphs, where the user can navigate the whole transcription ("plain text" mode). This information may be optionally shown in a big screen (connected to the server) or it can be sent wireless to the client which can be a PDA, a laptop or a mobile phone, for example. The wireless connection is done via Bluetooth.

In order to make the system non-ASR dependent, the server has been developed to receive data from any ASR application programmed to send the transcription via Transmission Control Protocol (TCP) connection. APEINTA is currently compatible with two commercial ASR systems: Dragon NaturallySpeaking (DNS) and ViaVoice (VV), mainly due to the availability of a software development kit (SDK) but it could be easily updated to be compatible with other ASR systems.

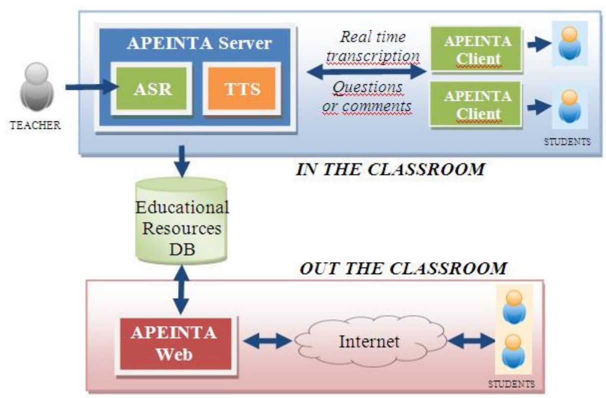

Figure 1: APEINTA architecture.

The system also allows students with speaking problems to type in a keyboard (of a laptop, a PDA, mobile phone, etc.) questions or comments which are sent back to the server where they will be converted to synthetic voice by means of a TTS system.

Finally, the orator's transcription and students' questions and comments are stored into a database (XML and SRT format) together with the audio and the podcast video of the class. These digital resources are accessible through the Web platform of the project. More details can be found in (Moreno et al. 2008).

Evaluation

The evaluation of the *In the classroom* system is presented in this section. The real-time transcription service and synthetic voice service are evaluated by surveying the users. This evaluation has been carried out in two different environments. On the one hand, the evaluation of the project was carried out without deaf or hard of hearing students firstly at Carlos III University of Madrid, during a 3th year subject of Computer Science degree called Database Design – *University evaluation* (Fig. 2). On the other hand, a second evaluation with deaf and hard of hearing users was held during a 30 minutes talk in the ACAPPS (Federation of Families and Deaf People of Catalonia) Educational Congress, in Barcelona on November 2009 – *Conference evaluation*.

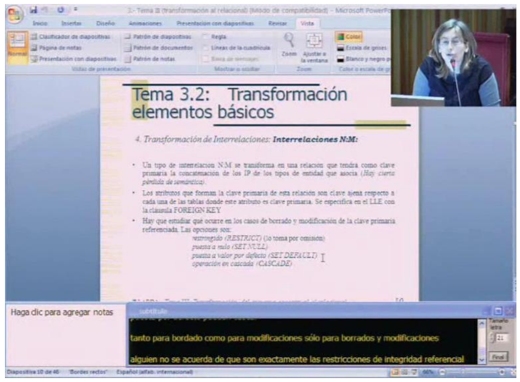

Figure 2: A moment during the *university evaluation* with the real time transcription generated within APEINTA.

The orator trained the ASR software during 40 minutes approximately in Spanish. ASR software was Dragon NaturallySpeaking (DNS) version 10 standard. The training was conducted by reading predefined texts (30 minutes) and adding new words to the vocabulary according to the subject specific vocabulary (10 minutes).

The *university evaluation* was a 50-minute lecture meanwhile *conference evaluation* was a 30–minute congress presentation. The average word error rate measured was 10.4%. The WER was calculated based on the Levenshtein Edit Distance (the sum of the number of substitutions, the number of deletions and the number of insertions) divided by the number of words in the original text which as translated manually.

The client devices evaluated by the users during both evaluations were two: PDA's and Laptop's (Fig. 3-4). Both devices have similar functionality but small differences in the interface due to size and performance differences (such as screen size, usability of the keyboard interface, etc.). Furthermore, the transcription was also showed in a big screen in the auditorium enclosed to the lecture slides connected to the server. Therefore, there were three different ways to read the lecture's transcriptions: in a PDA, in a laptop and/or in a big screen.

Text to Speech system (TTS) is based on Microsoft Speech API v5.1 for Spanish.