# Universidad Carlos III de Madrid
## e-Archivo

Institutional Repository

# Avoiding communication barriers in the classroom: the APEINTA project

Ana Iglesias[a]*, Javier Jiménez[b], Pablo Revuelta[c] and Lourdes Moreno[a]

[a]Computer Science Department, Carlos III of Madrid University, Avda. Universidad, 30, 28911 Leganés (Madrid), Spain; [b]APTENT be accessible!, Avenida de Gregorio Peces-Barba, 1, 28919 Leganés – Madrid(España), Spain; [c]Electronic Technology Department, Carlos III of Madrid University, Calle Butarque, 15, 28911 Leganés (Madrid), Spain

Education is a fundamental human right, however unfortunately not everybody has the same learning opportunities. For instance, if a student has hearing impairments, s/he could face communications barriers in the classroom, which could affect his/her learning process. APEINTA is a Spanish educational project that aims for inclusive education for all. This project proposes two main accessible initiatives: (1) real-time captioning and text-to-speech (TTS) services in the classroom and (2) accessible Web-learning platform out of the classroom with accessible digital resources. This paper presents the inclusive initiatives of APEINTA. Also an evaluation of the into-the-classroom initiative (real-time captioning and TTS services) is presented. This evaluation has been conducted during a regular undergraduate course at a university and during a seminar at an integration school for deaf children. Forty-five hearing students, 1 foreign student, 3 experts in captioning, usability and accessibility, and 20 students with hearing impairments evaluated these services in the classroom. Evaluation results show that these initiatives are adequate to be used in the classroom and that students are satisfied with them.

Keywords: inclusive education; education for all; real-time captioning; text-to-speech services; students with special needs

## 1. Introduction

Nowadays, universities and instructors are required to ensure that every student, regardless of their personal features, can enjoy the same educational opportunities. This is why inclusive education has been a prominent area of research since the 1990s (Newell & Gregor, 2000), even more when this requirement is included in the European Higher Educational Area (Adam, 2013).

Inclusion or inclusive education is defined as the practice of including all students of all abilities in all learning opportunities (Clough & Corbett, 2000).

The use of new technology can be useful for students and teachers in order to avoid communication and information barriers that unfortunately we can find yet in education (Bingimlas, 2009). For example, historically students with disabilities have experienced

---

*Corresponding author. Email: aiglesia@inf.uc3m.es

1

inadequate access to lecture material in the classroom and presence of communication barriers among other students and the teachers. A typical situation of accessibility barrier for deaf or hard of hearing students is when the teacher is writing on the board while still talking. Students who need to read the lips miss all the information that the teacher is transmitting in this situation.

Out of the classroom, students with disabilities find information barriers accessing learning resources due to their inability to satisfy the necessary accessibility requirements for all the students (Iglesias, Moreno, Martínez, & Calvo-Martin, 2011).

The proper use of the technology would permit inclusive experiences in education, for instance, adapting the pedagogical resources and strategies to each student individually. Moreover, the proper use of technology would also permit to support the teachers by adapting the courses and pedagogical strategies for all, aiming for inclusive education.

The APEINTA project[1] introduced in this paper presents several inclusive initiatives based on the use of new technologies in and out of the classroom. This project received in 2009 the FIAPAS[2] award for research and innovation in education.

Specifically, the paper is centred in evaluating if the inclusive proposals into the classroom are adequate (according to quantitative results) and if the students are satisfied with them (according to qualitative results).

The paper is structured as follows: Section 2 outlines related work. Section 3 describes the architecture of the APEINTA project. Section 4 presents the evaluation design of the real-time captioning and text-to-speech (TTS) services of APEINTA. Section 5 summarizes the main results obtained. Finally, Section 6 provides some conclusions, the study limitations and suggestions for further work.


## 2. Related work

Students with disabilities have historically experienced communication and information barriers in and out of the classroom, being one of the main handicaps in their learning process. Traditionally, Sign Language interpreters sometimes traduced the speech of the teacher, questions and comments of other students during the class to deaf students. However, nowadays not every student with hearing disability uses Sign Language and it is not a cheap method of accessibility in the classroom. Additionally, some researchers are working in lectures transcription based on automatic speech recognition (ASR). The liberated learning project (Bain, Basson, & Wald, 2005) is one of the most active researches in this area, collaborating with IBM to develop the ViaScribe software based on the ASR ViaVoice (VV) software. The Synote system (Wald, 2013) is centred on using the ViaScribe software for annotating recorded videos of the lectures. There are other initiatives based on other ASR systems as the VUST initiative from the Villanova University, which uses the Microsoft Speech Recognition Engine (Kheir & Way, 2007). 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, Bilinski, Adda, Gauvain, & Schwenk, 2006) or the LECTRA Project (Trancoso, Martins, Moniz, Mata, & Viana, 2008) for European Portuguese.

The APEINTA Project is the first initiative in real-time transcription and captioning of spoken Spanish lectures. On the one hand, it provides students with different methods to see transcription within individual devices (Personal Digital Agendas – PDA, laptops, etc.) or in a projection screen for all the students. Furthermore, APEINTA can be considered as one of the few projects that pay attention to speech problems of the students. TTS tools are used

in the classroom, so foreign language students or students with speaking difficulties are able to participate in the class with their comments or questions thanks to this service.

On the other hand, regarding the inclusive proposal out of the classroom, several research groups and organizations developed technologies and standards related to acquiring accessible educational resources and developing Web-platforms. For instance, the guidelines for developing accessible learning applications of the Global Learning Consortium (IMS) with the initiative of AccessForAll Meta-data (IMS Global Learning Consortium, 2004) the World Wide Web Consortium (W3C), and the Web Accessibility Initiative (W3C, 2009) of W3C with the Web Content Accessibility Guidelines 2.0 (W3C, 2008), which plays a leading role in promoting the importance of accessibility and developing guidelines. These guidelines are useful for developing accessible Web resources.

In additional, in order to elaborate and edit accessible digital resources, some normative must be followed such as the NCAM study in the educational area of Accessible Digital Media (NCAM, 2006) and the NIMAS Report v1.0 (2004) from the National Instructional Materials Accessibility Standard.

Nowadays, several contributions on accessibility applied to digital educational resources and learning environments like e-learning platforms are found in the literature. Most of them use automated accessibility testing tools in order to profile the community to gain an understanding of common accessibility problem areas (Abascal, Arrue, Fajardo, Garay, & Tomás, 2004). However, some studies found that a great percentage of them failed in the automated checking, so a manual analysis of the accessibility may find worse results (Curran, Walters, & Robinson, 2007). Only some systems are evaluated by combining semi-automated evaluation tools and manual evaluation by an accessibility expert, as the ALPE project (2009) or the ATutor system (ATutor, 2008). Some of these studies include disabled users who have access with the help of assistive technology (Moreno, Iglesias, & Martínez, 2007). In other cases, e-learning platforms provide direct access and they are oriented only to a predefined group of users and functionality, for instance, an e-platform providing resource adaptation to deaf or blind students (Nevile, Cooper, & Heath, 2005). Furthermore, some systems classify the students according to their abilities or disabilities in order to provide accessibility, as the Portland Plus project (Harrison, Stockton, & Pearson, 2008).

The APEINTA project includes an accessible educational Web-platform that provides universal access to multimedia learning resources, regardless of the learning or personal characteristics of the students. This Web-based learning platform was designed according to the current standards of accessibility, guidelines and Universal Design Principles in order to guarantee all students of all abilities access the pedagogical content of the system, without the necessity of classifying the students or distinguish them according to their abilities or access characteristics.

## 3. APEINTA architecture

The APEINTA architecture shown in Figure 1 presents different inclusive services in two well-differentiated environments: IN the classroom, where real-time captioning and TTS services are applied (Revuelta Sanz, Jiménez Dorado, Sánchez Pena, & Ruiz Mezcua, 2008; Moreno, Iglesias, Castro, & Martínez, 2008); and OUT of the classroom, where a Web-learning platform with accessible pedagogical contents can be used in every moment for students of all abilities including benefits in their learning (Revuelta Sanz et al., 2008; Moreno et al., 2008).

Real-time captioning is used mainly in the classroom, but it can be used also for captioning pre-recorded video or audio materials. In the classroom, this service avoids
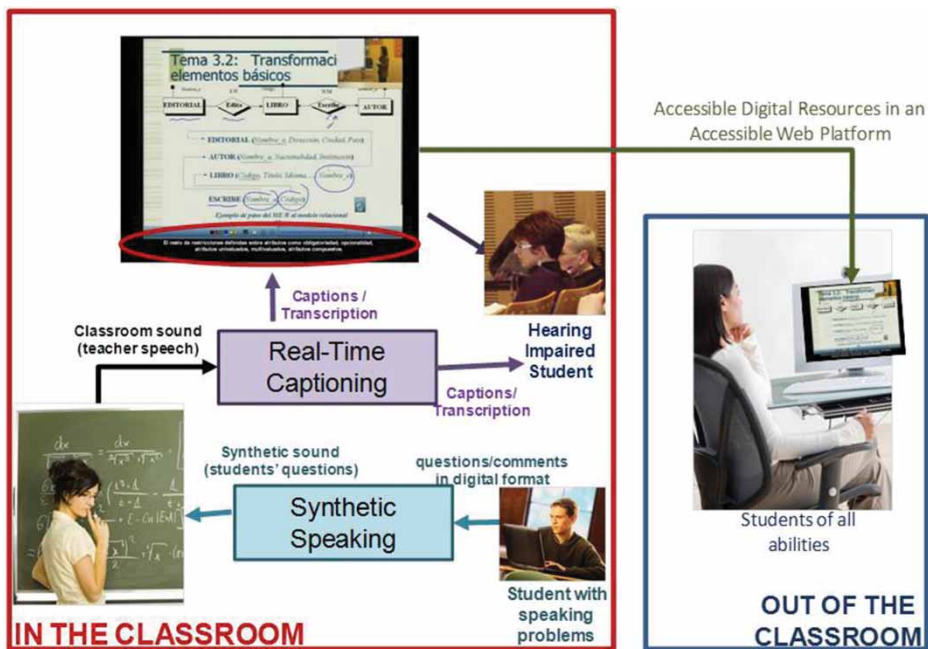
Figure 1.    APEINTA architecture.

communication barriers for hearing impaired students or foreign students, for instance, whose oral level of Spanish usually is lower than their reading level of the language. On the other hand, TTS service permits students with speaking problems participate during the classes with their comments or questions. The APEINTA real-time captioning service uses commercial software for ASR. The ASR used in the evaluation of this paper was Dragon Naturally Speaking v.10.[3] (DNS). Then, APEINTA divides the verbatim transcription provided by DNS into captions according to the UNE153010:2012 Spanish regulation (AENOR, 2012). Finally, the captions are projected into the students' personal devices.

The Web-learning platform provides universal access to pedagogical contents such as podcast videos of the lectures, slides with the presentation of the teachers and other academic documents. Moreover, the educational resources generated automatically during the classes by the real-time captioning service (audio, synchronized subtitles, notes in different formats, etc.), in addition to the podcast videos of the lectures, can be easily published in the Web-platform as an accessible resource available for all the students.

## 4.    Evaluation design

In this paper, the evaluation of the real-time captioning and TTS services of APEINTA are presented in two well-differentiated case studies in order to evaluate if these services are adequate to be used in the classroom and the students are satisfied with them:

(1) *Evaluation of real-time captioning and TTS services at university.* They were evaluated in undergraduate studies at university, during 2 regular lectures of 50 minutes each in a 6-month course. Hearing students, a foreign student, experts in

4

captioning, usability and accessibility, and students simulating different hearing disabilities participated in the evaluation.

(2) *Evaluation of the real-time captioning service at an integration school.* The system was evaluated in a 50-minute seminar. Ten- to sixteen-year-old students with different level of hearing disabilities participated in the evaluation.

## 4.1. *Instruments and measurements*

A quantitative evaluation of the real-time captioning was carried out through two typical measures in order to evaluate if the service is adequate to be used into the classroom:

(1) The word error rate (WER) in accordance with the Levenshtein Edit Distance (known as an edit metric as well) (Levenshtein, 1966).
(2) The delay in displaying the captions/transcriptions to the users.

The delay is derived from the ASR software due to the computational complexity of the task. The ASR process takes a certain time to transcribe teacher's speech. Besides, the ASR software waits until there is a short silence to give the whole transcription. Therefore, shown transcription to students is delayed with respect to the moment in which the teacher spoke. This delay mainly depends on when ASR software detects a silence. If the teacher speaks for a long time without making pauses, transcribed phrases by the ASR are longer and the delay increases. If the teacher makes frequent pauses, transcribed phrases are shorter and the delay decreases.

Apart from delay, ASR performance is not perfect. Any ASR software produces errors while transcribing. These errors are divided into three categories. Insertions (I), when there is a transcribed word that was not spoken; deletions (D), when a spoken word is not transcribed and substitutions (S), when a spoken word is transcribed by a different word. The most popular way to measure the ASR performance is the WER. The WER derives from Levenshtein Edit Distance between the transcribed text and the original transcription (reference) at the word level.

$$\text{WER} = (S + D + I)/N,$$

where S is the number of substitutions, D the number of the deletions, I the number of the insertions and $N$ the number of words in the reference.

On the other hand, a qualitative evaluation was done by surveying the users' satisfaction in the classroom when receiving the verbatim transcriptions of the teacher's speech in different devices.

Finally, we interviewed captioning, usability and accessibility experts, a foreign student and two more students (chosen in a random way and completely voluntary) in order to obtain a more precise feedback.

## 4.2. *Participants*

The assessment of the two case studies was conducted during sessions of 50-minute lectures within mixed (men and women) regular students at university and at school, respectively.

On the one hand, real-time captioning and TTS services were evaluated in undergraduate studies at university, during a six-month course of third-year Computer Science studies

at Universidad Carlos III de Madrid. The evaluation sessions were carried out during 2 lectures of 50 minutes each. Forty-five hearing students, 10 of them simulating hearing disabilities (with earphones and high music, earplugs, etc.), 1 foreign volunteer and 3 experts in captioning, usability and accessibility evaluated these services in the classroom.

On the other hand, real-time captioning service was evaluated as well at an integration school. The study was conducted during a 50-minute seminar at Tres Olivos school,[4] an integration school for deaf children. The design of this experiment was similar to the experiment at the university. Twenty volunteers with different levels of hearing disabilities participated in this evaluation.

It is important to underline that the class orator was the same person in both case studies.

### 4.3. *Evaluation procedure and materials*

Before the lectures, the speaker was trained in the ASR software for 40 minutes approximately. The training was conducted by reading predefined texts with non-specific vocabulary (30 minutes). Afterwards, for each case study the speaker was trained in the ASR software adding new specific words to the vocabulary according to the lecture subject at the university and at the school (10 minutes).

Moreover, it is important to underline that the lectures at the university and at school were carefully selected in order to avoid significant differences in the vocabulary and the sentences complexity for the ASR system. We also paid special attention to similitudes in the training phase in order to minimize the differences in the evaluation process.

During the 50-minute lectures at the university or school, students could visualize the verbatim transcription of the speech of the orator in different devices: a big screen in the auditorium, laptops and PDAs. Figure 2 shows the orator transcription in a big screen.

Moreover, laptops and PDAs allowed students to visualize the captions in two different modes:

(1) *Captions*. The transcription was divided into captions following the Spanish standard regulation (AENOR, 2012).
(2) *Transcription*. Plain text, where the sentences were not divided. Moreover, students could navigate through the transcription with a scroll bar.

Figure 3 shows the difference between transcriptions and captioning in a laptop and Figure 4 shows the captions in a PDA.

Moreover, these two personal devices permitted students to assess the TTS service, allowing them to write their comments or questions, which were lately repeated aloud with synthetic voice into the classroom through the teacher computer. At school, eight students could evaluate the TTS service, comparing its functionality by using laptops and PDAs. At university, 6 of the 10 students who simulated hearing disabilities, the foreign student and the 3 usability/accessibility experts could assess the TTS service interacting with both laptop and PDA devices.

During the 50-minute lectures, the audio of the lectures was recorded, and the verbatim transcription was stored. After that, the lectures were captioned by a human expert and then the WER and the delay derived from the ASR could be studied by comparing the captions automatically produced during the class and the captions generated by the human expert after the lecture.
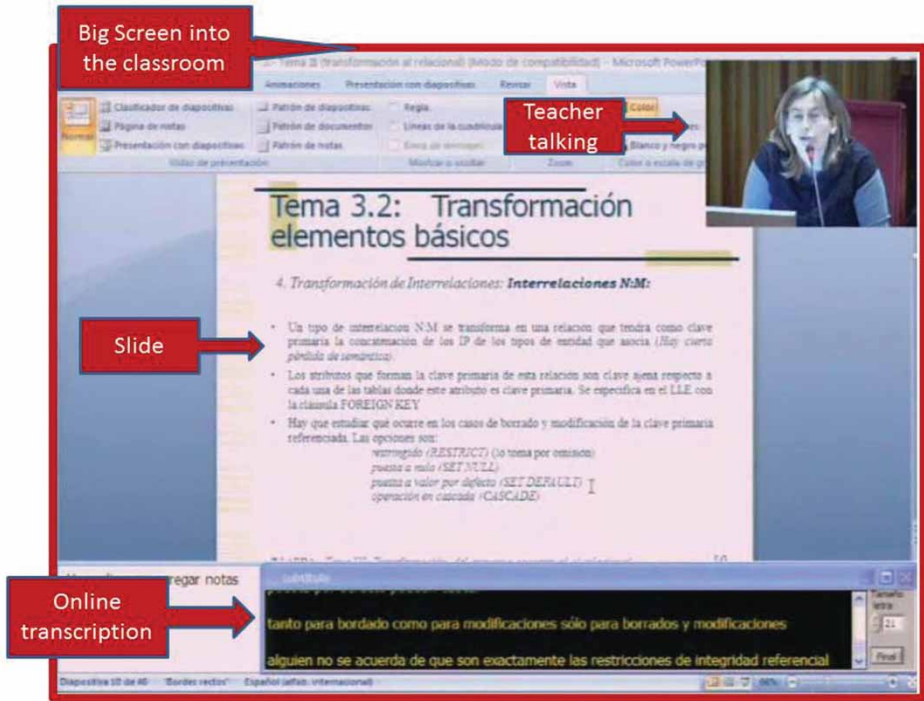
Figure 2.    Transcription in a big screen.



Figure 3.    Transcription (left)/caption (right) in laptops.

In addition, after each 50-minute lecture (at university or at school), students were asked to answer some survey questions measured in a five-point Likert-type Scale (Brooke, 1996). All the students completed the survey voluntarily and some of them provided written comments about improvements and satisfaction to the inclusive initiatives.

Finally, three experts in captioning, usability and accessibility were interviewed and they gave us their opinion of the services in the classroom after attending to the evaluation sessions at the university.

Figure 4.    Captions in the PDA.

## 5.    Evaluation results

### 5.1.    *Results related to WER and delay*

The average WER of the transcription derived from the ASR software was very similar in both case studies (at university and at the school). It was measured as 10.4%. This similarity is not unusual since the orator was the same and the training process and the complexity of the class topic were similar in both case studies.

The average delay displaying the captions in the classroom directly depends on the time passed between speech pause times (SPT). The SPT are defined as the silent interval between phonations during the ASR process. The ASR software used for the experiments presented in this paper is fast, processing the recognition in 0.5 seconds or less. So the delay depends directly on the way of speaking of the orator.

The SPT for both case studies were also very similar. The orator also took special attention on this characteristic. Specifically, for the 50-minute lectures, the ASR transcribed a total of 5484 words, so she spoke approximately 1.82 words per second. Moreover, the ASR recognized 1140 SPT per lectures, so the frequency of SPT (or the text blocks) was of 27 blocks per minute. Therefore, the length of each block of text was approximately of four words and the time spent for each block of text was of 2.19 seconds.

It is important to note at this point that there is a big difference in the delay between the captioning and the transcription modes. In the transcription mode, the time spent for each block of text is around 2.2 seconds (as we have explained before). However, the delay in the captioning mode is a compound delay, because it completely depends on the time the previous caption is exposed. The average delay for this visualization mode was approximately eight seconds for each caption, much worse than the transcription visualization mode.

### 5.2.    *User's satisfaction survey results*

Concerning the qualitative evaluation, four main groups of data were collected in the surveys, related to user's satisfaction with the real-time captioning service, the user's satisfaction with the TTS service, the user's satisfaction with the client devices and the general satisfaction with APEINTA inclusive proposals in the classroom. Table 1 summarizes the mean and standard deviation of the students' answers in a five-point Likert scale (1: strongly disagree to 5: strongly agree) and with the Not Applicable (N/A) option.

Table 1. Survey of the real-time captioning and TTS services.

| | | University (49 participants) | | | School (20 participants) | | |
|---|---|---|---|---|---|---|---|
| | | Non N/A answers | Mean | Std. dev. | Non N/A answers | Mean | Std. dev. |
| **Captioning quality** | | | | | | | |
| CQ.1 | I could easily detect errors in the transcription/captioning of the speech (omitted words, different words, etc.) | 49 | 3.90 | 0.70 | 20 | 4.55 | 0.69 |
| CQ.2 | I could easily read and understand the transcription/captioning of the speech, even with a total absence of punctuation marks and accents | 49 | 3.70 | 0.90 | 20 | 4.20 | 0.95 |
| CQ.3 | The captioning delay is not a big problem | 49 | 3.50 | 1.10 | 20 | 3.20 | 1.20 |
| **TTS quality** | | | | | | | |
| TTSQ.1 | I could easily understand the questions and comments | 49 | 4.20 | 0.70 | 20 | 4.22 | 0.97 |
| TTSQ.2 | The TTS service does not perturb the class rhythm | 49 | 3.30 | 0.80 | 20 | 3.80 | 0.84 |
| **Satisfaction using Big Screen/Laptop/PDA** | | | | | | | |
| D.1 | I could easily read the *caption* in the device | 49/ 10/ 10 | 5.00/ 4.00/ 3.80 | 0.00/ 0.60/ 0.7 | 20/ 8/ 8 | 5.00/4.75/ 3.88 | 0.00/ 0.46/ 0.35 |
| D.2 | The use of this device for real-time *captioning* is very useful and I recommend it | 49/ 10/ 10 | 4.90/ 4.3/ 3.80 | 0.3/ 0.5/ 0.7 | 20/ 8/ 8 | 4.94/4.63/ 3.75 | 0.25/ 0.52/ 0.46 |
| D.3 | I could easily read the *transcription* in the device | –/ 10/ 10 | –/ 4.7/ 4.0 | –/ 0.5/ 0.6 | –/ 8/ 8 | –/ 4.75/3.88 | –/ 0.46/ 0.35 |
| D.4 | The use of this device for real-time *transcription* is very useful and I recommend it | –/ 10/ 10 | –/ 4.6/ 3.9 | –/ 0.5/ 0.7 | –/ 8/ 8 | –/ 4.63/3.75 | –/ 0.52/ 0.46 |
| D.5 | I could easily *write comments or questions* in the device | –/ 10/ 10 | –/ 4.8/ 3.7 | –/ 0.4/ 0.8 | –/ 8/ 8 | – /4.75/3.50 | –/ 0.46/ 0.53 |
| D.6 | The use of this device for *writing questions or comments* is really useful and I recommend it | –/ 10/ 10 | –/ 4.7/ 3.6 | –/ 0.5/ 0.8 | –/ 8/ 8 | – /4.63/3.38 | –/ 0.52/ 0.52 |
| **General** | | | | | | | |
| G.1 | The real-time captioning is useful for me | 49 | 4.4 | 1.2 | 20 | 5.00 | 0.00 |
| G.3 | The TTS service is useful for me | 49 | 3.6 | 1.0 | 20 | 4.53 | 0.61 |
| G.5 | It is really interesting the use of real-time captioning in the classroom | 49 | 4.7 | 0.5 | 20 | 4.81 | 0.40 |

### 5.2.1. *Captioning quality*

Students were generally satisfied with the quality of the captioning service.

(CQ.1) Although the speech transcription was not completely correct, the students at school (most of them deaf or with hard of hearing impairments) could easily detect errors in the transcription of the speech (mean ($M$): 4.55 and standard deviation (SD): 0.69). The survey results at the university were a bit worse ($M$: 3.9 and SD: 0.7). This could be due to the fact that the level of the lectures at the university is high and the lectures are not easy to be understood, so a word changed on a caption, for instance, makes it difficult to understand the meaning of the sentence. Moreover, it is important to underline that most of students at school usually use captions in their daily life for communicating, so they usually face this problem in other environments and they have acquired this ability for years and this could affect the final result.

(CQ.2) Students, in general, could easily read and understand the transcription/captioning of the speech, even with a total absence of punctuation marks and accents. Again, the results at school ($M$: 4.20 and SD: 0.95) are better than the results at university ($M$: 3.7 and SD: 0.9). The cause could be the same.

(CQ.3) Concerning the captioning delay, they considered that it is a problem, and they were more divergent in the answers. At school, the data are summarized as ($M$: 3.20 and SD: 1.20) and at university as ($M$: 3.5 and SD: 1.1).

### 5.2.2. *TTS quality*

(TTSQ.1) Most of the students could easily understand the questions and comments sonified in the classroom with computer voice. The results at school ($M$: 4.22 and SD: 0.97) are very similar to those at university ($M$: 4.20 and SD: 0.70).

(TTSQ.2) Moreover, many of them thought that the TTS service does not perturb the lecture rhythm. Actually, the results at school were a bit better ($M$: 4.22 and SD: 0.97) than those at university ($M$: 3.30 and SD: 0.80).

### 5.2.3. *User's satisfaction for each device*

All the students at school (20 students) and at the university (49 students) could visualize the captions/transcription at the big screen and all of them answered the survey questions related to this device. However, only 8 students at the school and 10 users at the university could assess the TTS service interacting with both the laptop and the PDA.

(D.1) All the students (in both case studies) could easily read the captions in the big screen ($M$: 5 and SD: 0). They completely agree in this question. Students answered as well that it was easy to read the captions/transcriptions in the laptop (school → $M$: 4.75 and SD: 0.46 and university → $M$: 4.00 and SD: 0.6), although the size of the PDA screen made this task not so easy for students, so the results were worse for this device (school → $M$: 3.88 and SD: 0.35 and university → $M$: 3.8 and SD: 0.7).

(D.2) Again, the students answered that they prefer and recommend the big screen for real-time captioning (big screen/laptop/PDA at school → $M$: 4.94/4.63/3.75 and SD: 0.25/ 0.52/0.46 and at university → $M$: 4.9/4.3/3.8 and SD: 0.3/0.5/0.7).

(D.3 and D.4) Regarding the real-time transcription that could be only be assessed by the users interacting with laptops and PDAs, they could read them more easily at the laptop than at the PDAs, agreeing in the answers in both case studies. Moreover, the results were very similar when recommending the devices for this service.

(D.5 and D.6) When attending to the TTS service, which could only be assessed by using the laptop and PDA, we found that the users at school and at university were satisfied with the ease of writing comments or questions in the laptop, but they were less satisfied with the PDA device. For instance, the data in Table 1 for the school case study are (laptop/PDA *M*: 4.75/3.50 and SD: 0.46/0.53). Finally, students preferred the laptop to the PDA for writing questions and comments (data at school → laptop/PDA *M*: 4.63/3.38 and SD: 0.52/0.52).

### 5.2.4. *User's satisfaction in general*

All participants at school and at university were answered about the real-time captioning and TTS service in the classroom.

(G.1) Most of them answered that the real-time captioning is useful for them in the classroom (school → *M*: 5 and SD: 0 and university → *M*: 4.4 and SD: 1.2). The students with deaf or hard of hearing impairments at the school case study answered 5; 0 because they need it for communicating in the classroom, but the results at the university case study were also very high. When we interviewed students and experts after the evaluation sessions, they told us that captions helped them to pay more attention and to refocus on the class when they were distracted.

(G.2) The TTS service is useful for many students (school → *M*: 4.53 and SD: 0.61 and university → *M*: 3.6 and SD: 1.0). The results obtained at the university case study were very varied, but we want to underline that the answer for the foreign student in this questions was 5. Certainly this survey should be repeated with more foreign students in order to obtain a significant values, this is a limitation of this study, but it was not possible due to the fact that in our university there are not many foreign students yet in regular lectures, so we could not obtain more results yet, but we are currently extended this experiment.

(G.3) Finally, students were asked about their general satisfaction with the pedagogical initiative: the real-time captioning/transcription and the TTS services in the classroom. The results were very positive (school → *M*: 4.81 and SD: 0.4 and university → *M*: 4.7 and SD: 0.5).

### 5.3. *Users' Interviews*

In this section, we summarize the main results obtained from the interviews with the experts on accessibility, usability and captioning, and the foreign student and two more students were chosen randomly and voluntarily.

On the one hand, the experts on accessibility remarked that the use of the devices affected directly the user's attention. In the classroom, it is necessary to pay more attention in reading the transcription and, as a consequence, teacher's non-verbal information is eventually missed. For this reason, they agreed thinking that this system would have limitations in practical lectures while it would be very useful in more theoretical lectures.

On the other hand, the experts on usability noticed that when a student wanted to type a question in order to ask something to the teacher, during the typing process the student can sometimes miss important information. That is the reason why the device's usability is an important parameter to take into account. It would be necessary to expedite the typing and reduce spent time in this task for each device used.

Finally, the software of APEINTA allows individual control of the captioning speed in each device, so it may be modified according to user's necessities. Dealing with this functionality, the experts on captioning underlined that if a user configures the captioning speed

and sets it faster than the suggested speed in the Spanish standard (AENOR, 2012), to read all the captions could be a very difficult task for the student. On the other hand, it is important to remark that if the speed of the captions decreases, the delay would increase, unless the teacher speaks more slowly. However, this solution should be discarded since the system should affect teacher as less as possible.

### 5.4. *Discussion and limitations*

A quantitative evaluation was made for real-time captioning and a qualitative evaluation permits to evaluate the student's satisfaction for real-time captioning and TTS services.

According to the literature, a good comprehension of the transcription can be achieved if the WER is not bigger than 15% (Wald, 2004) for English language. There are no studies for Spanish language, but we could assume that the WER obtained in the evaluation (10.4%) is acceptable if applying the same threshold. The relationship between WER and comprehension directly depends on the specific vocabulary used in the speech related with the stored in the ASR system, so we asked the users in the survey if they could detect the errors and understand the overall meaning of the speech. We are currently working on reducing the WER.

Regarding the delay, the software used in this evaluation (DNS) waits until the whole phrase is pronounced to give the resulting transcription. This process is supposed to have lower WER because it uses context information to choose the best match transcription. The delay obtained in transcription is normal in this kind of software (Ando, Imai, Kobayashi, Isono, & Nakabayashi, 2000). However, the delay of captions is a cumulative delay, so it is not very desirable because it can confuse students in the classroom.

A general acceptance of the system was found in the users' satisfaction survey. Deaf people assessed positively the outcome of the real-time captioning provided by APEINTA since it is the first step to get a relative independence of stenotype, manual transcriptions and signing interpreters. Non-disabled students considered the system less useful for themselves, but they also found benefits in it. TTS service obtained a good level of users' satisfaction as well. Therefore, these evaluations show that the use of real-time transcription and TTS in the classroom can be a good opportunity for inclusive education because it opens an access to students with disabilities or communication problems. Moreover, the use of these technologies has demonstrated that they do not influence negatively in the learning process of their classmates.

Currently, we are working on looking for new users' devices and on improving the class conditions in order to help the students to integrate their attention with the teacher movements, the captions/transcription reading and the use of TTS services.

### 6. Conclusions and further research

The APEINTA project proposes to use two different inclusive proposals. In the classroom, the use of ASR and TTS services allows to avoid communication barriers among the speaker and the attendants. Out of the classroom, the use of accessible Websites and multimedia resources provide benefits to all the users.

This paper presents the evaluation of the real-time captioning/transcription services in two different educational environments: during two 50-minute lectures at the university and during a 50-minute seminar at an integration school for students deaf or with hard of hearing impairments. Nowadays, we are working in evaluating the APEINTA project (with all its services) in a whole course where students with different levels of disabilities and foreign students are enrolled.

The APEINTA project was originally designed, implemented and evaluated in the educational environment, but after this first evaluation process we noticed the benefits that this initiative provides to people with temporal or permanent disabilities, elder and foreigners or students with speaking problems, for instance. Subsequently, this project is currently being introduced and evaluated in other environments, such as research conferences, talks, etc.

## Notes

1. Spanish educational project where the Spanish Centre of Captioning and Audio-description (CESyA) and the Computer Science Department and the Electronic Technology Department of the Universidad Carlos III de Madrid collaborate in order to provide inclusive education for all, independently of the students' abilities.www.APEINTA.es (last access on April 2014).
2. FIAPAS is the Spanish Confederation of Parents and Friends of Deaf People (www.fiapas.es – last access on April 2014)
3. http://www.nuance.com/naturallyspeaking/ (last access on April 2014)..
4. Tres Olivos school. http://www.colegiotresolivos.org/ (last access on April 2014).

## Notes on contributors

Ana Iglesias Maqueda has been a member of the faculty of the Computer Science Department of Carlos III University of Madrid, since February 2006. She obtained a degree in Computer Science from Carlos III University of Madrid in 1999, and her Ph.D. in Computer Science from Carlos III University of Madrid (UC3M) in 2004. From January 2005 until August she carried out a postdoctoral stay at the Human Computer Interaction Institute of Carnegie Mellon University. Since 2005, she is collaborating in the Spanish Center of Captioning and Audiodescription (CESyA). She has been working in several National research projects on Inclusive Education, e-Learning platforms, Natural Language Processing and Information Retrieval, Advanced Database Technologies and CASE Environments. She is the main research in the APEINTA project. Her research interests include Inclusive Education, Assistive Technologies, Adaptive Intelligent Educational Systems, Natural Language Processing and Information Retrieval, Database Design and Advanced Database Technologies and Software Engineering.

Javier Jiménez Dorado is currently working in APTENT be accessible! This company explores the use of automatic speech recognition and ICTs for accessibility, in particular automatic live closed captioning in education, theatre and television. Previously he was working for four years as Technician-Researcher at the Spanish Center for Subtitling and Audio Description (CESyA) and at the Displays and Optical Applications Laboratory at the Carlos III University for 2 years. He is currently a Ph.D. student at the Carlos III University. He has several publications as co-author in both national and international conferences and book chapters. His current work and his interests are focused on two fields, both related to accessibility and technical aids: Automatic Subtitling based on Speech Recognition and non-invasive EEG-based Brain–Computer Interfaces.

Pablo Revuelta Sanz obtained his degree in Telecommunication Engineering (2006) at the Carlos III University of Madrid (Spain), obtained his Master's degree in 2008 and Ph.D. in 2013 from the same university. He has published several conference papers and three book chapters focused on assistive products, image processing, sonification and political philosophy.

Lourdes Moreno López. She has a Ph.D. in computer science from Universidad Carlos III de Madrid and Mathematician in the specialty of Computation Sciences for the Universidad Complutense de Madrid. She has worked with several IT companies, in R+D Departments, working in matters about Infometrics (information measurement) especially in web-channel. Since 2002 she works as an Associate Professor and since 2004 as an Assistant Professor at the Advanced Database group in the Computer Sciences Department at the Universidad Carlos III de Madrid. She has developed the doctoral thesis in the Design and development of accessible Web applications. She works in several international research projects and is co-author of several publications.

# References

Abascal, J., Arrue, M., Fajardo, I., Garay, N., & Tomás, J. (2004). The use of guidelines to automatically verify web accessibility. *Universal Access in the Information Society*, *3*(1), 71–79.

Adam, S. (2013). The central role of learning outcomes in the completion of the European Higher Education Area 2013–2020. *Journal of the European Higher Education Area*, (2). Retrieved from http://www.ehea-journal.eu/index.php?option=com_docman&task=doc_details&gid=349

AENOR. (2012). Subtitulado para personas sordas y personas con discapacidad auditiva [Captioning for deaf and hard of hearing personas]. Retrieved July 2013, from http://www.en.aenor.es/aenor/normas/buscadornormas/resultadobuscnormas.asp#.UfE9fG1p3sc

ALPE project. (2009). *The accessible e-learning platform for Europe Project, aDeNu (Adaptive Dynamic online Educational systems based oN User modelling)*. Spanish Distance Learning University (UNED). Retrieved from http://adenu.ia.uned.es/alpe/

Ando, A., Imai, T., Kobayashi, A., Isono, H., & Nakabayashi, K. (2000). Real-time transcription system for simultaneous subtitling of Japanese Broadcast News Programs. *IEEE Transactions on Broadcasting*, *46*(3), 189–196.

ATutor. (2008). *The Adaptive Technology Resource Centre (ATRC) at the University of Toronto Atutor*. Open source web-based Learning Content Management System (LCMS), 2008. Retrieved from http://www.atutor.ca/

Bain, K., Basson, S., & Wald, M. (2005). *Speech recognition in university classrooms: Liberated learning project*. Proceedings of 5th annual international ACM conference on Assistive Technologies, 192–196.

Bingimlas, K. A. (2009). Barriers to the successful integration of ICT in teaching and learning environments: A review of the literature. *Eurasia Journal of Mathematics, Science & Technology Education*, *5*(3), 235–245.

Brooke, J. (1996). Sus: a 'quick and dirty' usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & I. L. McClelland (Eds.), *Usability evaluation in industry* (pp. 189–194). London: Taylor & Francis.

Clough, P., & Corbett, J. (2000). *Theories of inclusive education: A student's guide*. London: SAGE Publications Ltd. doi:http://dx.doi.org/10.4135/9781446220436

Curran, K., Walters, N., & Robinson, D. (2007). Investigating the problems faced by older adults and people with disabilities in online environments. *Behaviour & Information Technology*, *26*(6), 447–453.

Harrison, M., Stockton, C., & Pearson, E. (2008). *Inclusive, adaptive design for students with learning disabilities*. Eighth IEEE International Conference Advanced Learning Technologies, 2008. ICALT '08. Cantabria, Spain: Santander.

Iglesias, A., Moreno, L., Martínez, P., & Calvo-Martin, R. (2011). Evaluating the accessibility of three open-source learning content management systems: a comparative study. *Computer Applications in Engineering Education*, *19*(2), 320–328.

IMS Global Learning Consortium. (2004). *IMS access for all meta-data*. Retrieved July 2013, from http://www.imsglobal.org/accessibility/accmdv1p0/imsaccmd_bestv1p0.html

Kheir, R., & Way, T. (2007). *Inclusion of deaf students in computer science classes using real-time speech transcription*. 12th annual SIGCSE conference on innovation and Technology in Computer Science Education, ITiCSE '07. 261–265.

Lamel, L., Bilinski, E., Adda, G., Gauvain, J., & Schwenk, H. (2006). The LIMSI RT06s lecture transcription system. *Machine Learning for Multimodal Interaction*, *4299*, 457–468.

Levenshtein, V. L. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, *10*, 707–710.

MIT. (2003). *ICAMPUS projects*. Recuperado el December de 2009, de Spoken Lecture Processing. Retrieved from http://icampus.mit.edu/projects/SpokenLecture.shtml

Moreno, L., Iglesias, A., Castro, E., & Martínez, P. (2008). *Using accessible digital resources for teaching database design: Towards an inclusive distance learning proposal*. In Proceedings of the 13th Annual Conference on innovation and Technology in Computer Science Education (ITiCSE'08), 32–36.

Moreno, L., Iglesias, A., & Martínez, P. (2007). *Accessibility of multimedia resources in web*. In Proceedings of WEBIST – International conference on web information systems and technologies, Barcelona, España, 470–473, March 2007.

NCAM. (2006). *CPB/WGBH*. Retrieved July 2013, from http://ncam.wgbh.org/cdrom/guideline/

Nevile, L., Cooper, M., & Heath, A. (2005). *Learner-centered accessibility for interoperable web-based educational systems*. Proceedings of the 1st international workshop on interoperability of web-based education systems, 14th International World Wide Web Conference, Chiba, Japan.

Newell, A. F., & Gregor, P. (2000). User sensitive inclusive design – in search of a new paradigm. *CUU '00*. Retrieved from http://doi.acm.org/10.1145/355460.355470

NIMAS Report v1.0. (2004). *National file format technical panel, national instructional materials accessibility standard report – Version 1.0, Resource Document*. National Center on Accessing the General Curriculum. Retrieved July 2013, from http://www.cast.org/ncac/nimas/executive_summary.htm

Revuelta Sanz, P., Jiménez Dorado, J., Sánchez-Pena, J. M., & Ruiz-Mezcua, B. (2008). *Online captioning system for educational resources accessibility of hard-of-hearing people*. In Proceedings of the 11th IASTED Conference on Computers Helping People with Special Needs (ICCHP'08), 22–32.

Trancoso, I., Martins, R., Moniz, H., Mata, A. I., & Viana, M. C. (2008). *The LECTRA Corpus – Classroom lecture transcriptions in European Portuguese*. Proc. of the Sixth International Language Resources and Evaluation (LREC'08).

W3C. (2008). Web Content Accessibility Guidelines (WCAG) 2.0. Overview, Resource document. *World Wide Web Consortium*. Retrieved July 2013, from http://www.w3.org/WAI/intro/wcag.php

W3C. (2009). Web Accessibility Initiative (WAI). Resource document. *World Wide Web Consortium*. Retrieved July 2013, from http://www.w3.org/WAI/

Wald, M. (2004). *Using automatic speech recognition to enhance education for all students: Turning a vision into reality.* 34th ASEE/IEEE Frontiers in Education Conference, 22–25.

Wald, M. (2013). *Synote discussion. Extending Synote to support threaded discussions synchronised with recorded videos*. Proceedings of the eighth international conference on internet and web applications and services. The eighth international conference on internet and web applications and Services (ICIW 2013), International Academy, Research and Industry Association.