Voice Interactive Learning: A Framework and Evaluation

MPuerto Paule-Ruiz
Department of Computer Science, University of Oviedo, Spain
+34 98510 4338
paule@uniovi.es

Victor Álvarez-García
Department of Computer Science, University of Oviedo, Spain
+34 98510 4338
victoralvarez@uniovi.es

J.R. Pérez-Pérez
Department of Computer Science, University of Oviedo, Spain
+34 98510 5094
jpp@uniovi.es

M. Riestra-González
Department of Computer Science, University of Oviedo, Spain
moi.riestra@gmail.com

ABSTRACT
Multi-channel access to information has gained interest in the past years. Yet, the usage of alternative modes of interaction has not been extended into mainstream e-learning systems. This paper illustrates how the elements of a multi-channel learning framework can be identified and used in practice to enable complementary aural access to visual-only web-based environments. To complement these findings, this research proposes an evaluation method that considers usability and didactic effectiveness parameters to support the assessment of voice interactive learning solutions, and allows for the exploration of the meaning and measure of enabling voice interaction in traditional Internet-based learning systems. The results obtained from developing and evaluating audio features with postgraduate students from the University of Oviedo, allow us to present an analysis of the benefits and implications of following the proposed approach, and better understand the influence of voice interaction in e-learning.

Categories and Subject Descriptors
K.3.1 [Computer Uses in Education]: Distance Learning. H.3.4 [Systems and Software]: Performance evaluation (efficiency and effectiveness)

General Terms
Measurement, Design, Experimentation, Human Factors

Keywords
Voice interaction, evaluation, e-learning, adaptation

1. INTRODUCTION
Speech is the primary means of communication for human beings. While not all human languages have a written form, every language has a spoken form. The challenge of using machines to mimic human behaviour, and particularly the capability of speaking naturally, has intrigued engineers and scientists for centuries. Speech technology gets more attention than usual these days because it is a good fit for pervasive computing solutions, and it is commonly used in communications in multi-channel computing environments.

Multi-channel systems offer a communication interface mainly based in visual means, but with the potential of providing additional and complementary communication channels and interaction styles according to the different contextual settings. More particularly, we consider mainstream Internet-based learning systems can take advantage of aural communications to complement their visual-based interfaces, in such a way that, multi-channel features become more prevalent and give rise to new learning scenarios users can benefit from. This proposal finds support in our previous research findings [1], which suggest a series of learning scenarios that make it either necessary or highly recommended to complement visual learning with audio interactions (see Figure. 1).

Our interest to extend traditional e-learning platforms to accommodate access by aural means, led us to take an empirical approach for a voice interactive learning framework that takes into account the principles of the Universal Design for Learning (UDL), as well as the theoretical fundamentals of Educational Hypermedia Systems, to enable creating voice driven learning management systems. This framework finds application in a software process that allows for the design, development and integration of voice-based learning features onto e-learning platforms [2]. We employed this process to produce a series of voice-based educational services that have been tested with postgraduate Computer Science students from our University. In order to support the assessment of voice interactive learning solution, and allows for the exploration of the meaning and measure of enabling voice interaction in traditional Internet-based learning systems, we propose an evaluation method based on different usability aspects along with didactic effectiveness measurements for voice-enabled learning features, and which aims at better understanding the influence of voice interactive learning in education.
2. RELATED WORK
In the study of software for education, and in the particular case of web-based e-learning, recent research works [1] propose to factorise e-learning systems into a collection of dedicated services to address the issues of multi-channel access and meet the needs of mobile learning communities. These studies suggest creating a mobile learning framework which makes use of a variety of communication channels such as Web, email, SMS or phone call, in such a way that, channels are combined in a same unit of interaction depending of students’ context-of-use, and learning services are developed to support and adapt the various access modes.

One of the challenges in the design of software systems based on learning frameworks consists in determining measurement parameters to evaluate learning benefits in an objective and accurate manner. Literature on e-learning systems evaluation has mainly focused on usability, suggesting different approaches for measurement. In [3] the authors define SUE (Systematic Usability Evaluation), a methodology for evaluating e-learning applications. SUE combines inspection technique and user testing for the evaluation of software applications. The main assumption made by other study [4] is that e-learning continuance intention is determined by students' satisfaction, which is in turn jointly determined by perceived usability, among other parameters.

The above-mentioned evaluation methods have found application in traditional e-learning systems, available only when using web browsers and visual interaction. However, in the case of access by aural means, evaluation must be particularised for the specific perspective of spoken dialogues. This type of evaluation has been traditionally carried out in terms of instrumentally or expert-derived measures (usually called “objective” evaluation) and quality judgements of users who have previously interacted with the system (also called “subjective” evaluation). One of the most thorough analysis published [6] illustrates the progress in the assessment of the usability in spoken language dialogue systems (SLDS) during fifteen years. The result is a review of a series of European and US-based projects, which have produced major results on product evaluation and usability assessment. A special highlight is given to PARADISE (PARAdigm for DIalog Evaluation System), the most widely proposed methodology to perform a global evaluation of dialogue systems.

3. A VOICE INTERACTIVE LEARNING FRAMEWORK
Voice interactive learning is a framework, which combines voice communication technologies with pedagogical considerations. In addition, this research finds practical application in the design of a software architecture that provides the technical basis for the development of voice interactive learning features.

3.1 Pedagogical Considerations
The success of LMS deployments within organisations is determined by the integration of pedagogical and technological aspects in e-learning implementation exercises. Adaptive learning is based on the idea of adapting learning methodologies to students’ learning styles, preferences, goals, knowledge level, background, interest, preferences, stereotypes or cognitive preferences [9]. The development and use of adaptive learning environments allow for the integration of the concepts and theories of education and information technologies. Educational hypermedia was one of the first application areas for adaptive learning in Internet-based learning systems. Adaptive Educational Hypermedia caters to the needs of each individual student, adapting learning content and navigation accordingly to individual needs or abilities. The development of Adaptive Educational Hypermedia systems is intimately related with the methods and techniques of Adaptive Hypermedia to adapt the information to students' knowledge and preferences.

Figure 2 Design principles for voice-enabled learning systems
Our voice interactive learning framework is also founded on the principles of “The Universal Design for Learning” (UDL). UDL is used as a reference to allow accommodating design principles into the characteristics of a generic voice-enabled learning system (see Figure 2). At the same time, our proposed framework adopts learning as an active process where it is vital that the information is presented and organised according to the needs of each individual student.

3.2 Voice Interactive Classroom
Voice Interactive Classroom (VIC) [1] is a software architecture that provides the mechanisms for adapting the user interface of Learning Management Systems (LMSs) in order to create audio experiences for learners. In addition to following guidelines for interaction design, Voice Interactive Classroom introduces a user-centred iterative design process that helps to make audio interactions more user-friendly and better suit learners’ requirements.

We propose to address the problems of accommodating design principles into the characteristics of voice-enabled learning systems, and adapting the interfaces from visual-only learning into naturalistic voice dialogues, by means of applying methods and techniques of adaptive hypermedia. At the same time, Voice Interactive Classroom uses of a service-oriented architecture, supported by the e-learning specifications IMS Abstract Framework (IAF; http://www.imsglobal.org/af/) and the Open Knowledge Initiative (OKI; http://www.okiproject.org/), and compliant with W3C Voice Browser recommendations (http://www.w3.org/Voice/), to provide a complementary audio communication channel and enable the intra-domain/inter-domain interoperability of audio features among the various e-learning platforms and components. As result, VIC provides multi-channel access to existing Internet-based e-learning platforms. More particularly, we have successfully enabled multi-channel access to Segue (http://segue.middlebury.edu/) and Sakai (http://sakaiproject.org/), providing users with the option of consulting the information using the traditional access through a web browser or the alternative access through a VoIP audio device.
3.3 Vocal LMS Tools and Services
Learning Management Systems (LMSs) provide an integrated environment for designing and delivering learning materials and activities. In our view, aural-based LMSs share the goals with their visual counterpart, which makes it possible to represent them as a set of tools that support services such as learning content management, education administrative tasks, students enrolment and learning performance reporting.

Although visual and aural-based LMSs are suitable for providing equivalent tools, the way the information is structured and organised differs from visual to verbal interactions. In a visual-based LMS, information is built from a set of (X)HTML pages (dynamic or static), where the user selects which route to take according to options available. In an aural-based LMS, the information is determined by voice dialogues, and the routes are selected from the options offered in audio prompts.

Another aspect that must be kept in mind is the cognitive overload caused by the use of aural systems [10] and aural LMSs, which can be classified among the former. Thus, both the amount of information presented and the number of navigation options have to be restricted. In order to create a naturalistic dialogue flow and avoid information overloads, the vocal LMSs can use the methods and techniques of adaptive hypermedia. More specifically, conditional inclusion of fragments can be used to reduce the amount of information included in the spoken interface, and link removal allows to restrict the number of navigation options available in voice dialogues and give them a more naturalistic and simplistic flow.

4. EVALUATION
One essential achievement of this research consists in devising an evaluation model for voice-oriented e-learning systems. Although the scientific literature describes methods and processes for usability evaluation in audio and e-learning settings, they have not been related to one another. Our proposed empirical evaluation model benefits from the common characteristics of audio and e-learning systems by including the assessment of usability parameters. At the same time, we take into account that e-learning system developments need to accommodate domain-specific characteristics, and they should also be evaluated considering their didactic effectiveness [3].

4.1 Instruments and Measurements
Usability is among the determining factors of pedagogical success in an e-learning system, since the lack of usability can result in an obstacle for students’ learning [12]. Usability is a measure of satisfaction, effectiveness and efficiency with which users achieve their goals in a specific context-of-use. Our process for evaluating the usability of the prototypes suggests taking into account a combination of usability aspects in e-learning and speech-enabled applications. It applies both quantitative and qualitative measures to evaluate the effectiveness and efficiency of the prototypes as well as the satisfaction of e-learners.

ISO’s effectiveness is the accuracy and completeness with which users achieve certain goals. There are several indicators of effectiveness, including quality of solution and error rates. We use error rates as indicator for effectiveness. In our approach, error rates are calculated as the ratio between TurnsPerfect, a constant number predefined by the dialog interface design, and TurnsPractice, obtained in practice from logs files and audio recorded files where e-learner's voice is recorded.

3.3.4 Vocal LMS Tools and Services
Learning Management Systems (LMSs) provide an integrated environment for designing and delivering learning materials and activities. In our view, aural-based LMSs share the goals with their visual counterpart, which makes it possible to represent them as a set of tools that support services such as learning content management, education administrative tasks, students enrolment and learning performance reporting.

Although visual and aural-based LMSs are suitable for providing equivalent tools, the way the information is structured and organised differs from visual to verbal interactions. In a visual-based LMS, information is built from a set of (X)HTML pages (dynamic or static), where the user selects which route to take according to options available. In an aural-based LMS, the information is determined by voice dialogues, and the routes are selected from the options offered in audio prompts.

Another aspect that must be kept in mind is the cognitive overload caused by the use of aural systems [10] and aural LMSs, which can be classified among the former. Thus, both the amount of information presented and the number of navigation options have to be restricted. In order to create a naturalistic dialogue flow and avoid information overloads, the vocal LMSs can use the methods and techniques of adaptive hypermedia. More specifically, conditional inclusion of fragments can be used to reduce the amount of information included in the spoken interface, and link removal allows to restrict the number of navigation options available in voice dialogues and give them a more naturalistic and simplistic flow.

4. EVALUATION
One essential achievement of this research consists in devising an evaluation model for voice-oriented e-learning systems. Although the scientific literature describes methods and processes for usability evaluation in audio and e-learning settings, they have not been related to one another. Our proposed empirical evaluation model benefits from the common characteristics of audio and e-learning systems by including the assessment of usability parameters. At the same time, we take into account that e-learning system developments need to accommodate domain-specific characteristics, and they should also be evaluated considering their didactic effectiveness [3].

4.1 Instruments and Measurements
Usability is among the determining factors of pedagogical success in an e-learning system, since the lack of usability can result in an obstacle for students’ learning [12]. Usability is a measure of satisfaction, effectiveness and efficiency with which users achieve their goals in a specific context-of-use. Our process for evaluating the usability of the prototypes suggests taking into account a combination of usability aspects in e-learning and speech-enabled applications. It applies both quantitative and qualitative measures to evaluate the effectiveness and efficiency of the prototypes as well as the satisfaction of e-learners.

ISO’s effectiveness is the accuracy and completeness with which users achieve certain goals. There are several indicators of effectiveness, including quality of solution and error rates. We use error rates as indicator for effectiveness. In our approach, error rates are calculated as the ratio between TurnsPerfect, a constant number predefined by the dialog interface design, and TurnsPractice, obtained in practice from logs files and audio recorded files where e-learner's voice is recorded.

E-learners’ satisfaction is the interface ability to interact with a user in a pleasant way. Surveys are a common method to gather students’ perceptions about their experiences with e-learning environments and assert qualitative parameters. This study customises an empirically validated survey instrument developed by [11] for measuring learners’ satisfaction in asynchronous e-learning systems. The survey questionnaire is available at http://www.pulso.uniovi.es/learninganalytics/questionnaire.pdf, and it includes a set of nine questions that have to be answered using a Likert scale from 1 (Strongly Disagree) to 4 (Strongly Agree). The middle point was omitted in order to eliminate the tendency of students to avoid committing in one direction or another.

4.2 Background Didactic Effectiveness
Literature reports two different approaches to evaluate didactic effectiveness, instructional design, and student-centered pedagogical practices. Instructional design considers aspects related to educational planning of content elements or correctness and accuracy of contents [3], but without providing an objective way to measure those aspects. From a student-centered perspective, didactic effectiveness is associated with changes in students’ performance over their instruction time [8]. There are different opinions about how to determine its measurement. In [7], the authors suggest to measure didactic effectiveness as the total amount of completed, visited, or studied learning activities during a learning phase. However, it can be also measured by using a value added approach [5], this is, didactic effectiveness is the ratio between the change in performance of the student, and his/her knowledge needs before attending the course.

After considering the previous factors, we decided to follow Cowan’s index. The focus on Cowan’s index allows measuring educational and academic results by offering a specific and precise index of the efficiency in students’ knowledge acquisition over a period of time.

4.3 Usability Testing
We have applied the aforementioned described instruments and measurements to the tools for checking grades, calendar notes/events and RSS news. These modules, which are responsible for common e-learning administrative tasks, allow us to assess the usability of the system from a user perspective.

4.3.1 Method, Participants and Materials
The assessment of the three case studies was conducted within a mixed class of 20 students enrolled in the second course of the Web Engineering Master's degree at the University of Oviedo. All the students were regular users of the virtual classroom and they had also prior knowledge in the use of audio-based software. Students were asked to complete the three tasks using both the visual and audio interfaces of the learning services. In the case of visual access, students made use of a PC’s web browser to access the virtual campus, whereas vocal access was performed from a softphone and wearing headphones to avoid environmental noise and mutual disturbance. Computers and headphones used for the experiment were of the same make and model. It is also important
to note that before experiments started, e-learners were instructed on the steps and completion of tasks, with the goal of improving the quality and efficiency of interactions of novice users with the system.

4.3.2 Procedure
Samples for the studies were collected using three different procedures, console log and log files, speak-aloud, and a survey on users’ satisfaction.

Console and log files monitor and register information about e-learner’s interaction with the system; they show and collect information about the inputs and outputs as well as starting and finishing times for each task. The participants were asked to speak aloud while doing the activities. Their thoughts and comments were recorded in audio files, which also registered starting and finishing times, as well as the time instant when the e-learner made any mistake. Finally, once the activities were completed, e-learners were asked to fill a survey to assess their satisfaction with the system.

4.3.3 Results and Discussion
Our initial evaluation of the module for checking grades already yielded to meaningful results. Efficiency measures showed that the completion time for the aural task was 56% longer than the visual equivalent. This increment is caused by the forced sequentially of aural navigation. Effectiveness indicated that the interaction with the user was correct, with an error rate below 5%

In the measure of learner’s satisfaction, all the students indicated they highly agreed with the content, presentation and navigation of the aural features. In this case, aural satisfaction exceed to its visual counterpart in 3%. Students particularly valued the easiness of browsing grades in the aural system, as opposed to performing the same task on the visual platform.

The previous findings were complemented by the results obtained from the evaluation in the module for checking calendar notes/events. Efficiency measures showed that completion time for the aural task was 23% longer than the equivalent visual interface. This improvement was achieved by reducing the number of dialogue turns. Effectiveness indicated that the interaction with students was correct, scoring an error rate of approximately 10%. In this case, satisfaction scores were slightly higher than in the previous case. As occurred in the previous case, aural satisfaction is 3% higher compared to that in the visual mode, with students considering both aural navigation and content easier and more intuitive than their visual counterpart.

Finally, the module for checking RSS feeds presented additional and significant results with regards to aural interaction. Efficiency assessment shows, in this case, that the aural task takes in average 50% less time than the visual task. In our interpretation of this result, aural interaction benefits more from tasks which are performed sequentially, such as consulting RSS feeds, as opposite to visual interaction, which proves more effective for non-sequential navigation, that is, when users can browse through different interface elements simultaneously. Effectiveness rate scored around 7%, showing similar values with the rest of the modules. Satisfaction shows again similar values for both aural (75%) and visual interaction (79%).

Table 1: Satisfaction results: Visual and aural comparison (with mean and standard deviation)

<table>
<thead>
<tr>
<th>Question</th>
<th>Checking grades</th>
<th>Checking events</th>
<th>Checking RSS news</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Visual</td>
<td>Vocal</td>
<td>Visual</td>
</tr>
<tr>
<td>1</td>
<td>3.80 0.40</td>
<td>3.80 0.40</td>
<td>3.70 0.46</td>
</tr>
<tr>
<td>2</td>
<td>3.70 0.46</td>
<td>3.70 0.46</td>
<td>3.70 0.46</td>
</tr>
<tr>
<td>3</td>
<td>3.70 0.46</td>
<td>3.80 0.40</td>
<td>3.60 0.66</td>
</tr>
<tr>
<td>4</td>
<td>3.70 0.46</td>
<td>3.80 0.40</td>
<td>3.70 0.46</td>
</tr>
<tr>
<td>5</td>
<td>3.40 0.80</td>
<td>3.60 0.66</td>
<td>3.40 0.66</td>
</tr>
<tr>
<td>6</td>
<td>3.10 0.94</td>
<td>3.60 0.49</td>
<td>3.30 0.66</td>
</tr>
<tr>
<td>7</td>
<td>3.60 0.49</td>
<td>3.70 0.46</td>
<td>3.50 0.67</td>
</tr>
<tr>
<td>8</td>
<td>3.10 0.94</td>
<td>3.60 0.66</td>
<td>3.60 0.49</td>
</tr>
<tr>
<td>9</td>
<td>3.70 0.46</td>
<td>3.40 0.66</td>
<td>3.80 0.40</td>
</tr>
</tbody>
</table>

Table 2 Efficiency results: Visual and aural comparison (with mean and standard deviation)

<table>
<thead>
<tr>
<th>Checking grades</th>
<th>Checking events</th>
<th>Checking RSS news</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual</td>
<td>Vocal</td>
<td>Visual</td>
</tr>
<tr>
<td>X σ</td>
<td>X σ</td>
<td>X σ</td>
</tr>
<tr>
<td>37,80 13,53</td>
<td>59,20 8,84</td>
<td>38,50 12,61</td>
</tr>
</tbody>
</table>
Table 3 Effectiveness results: Visual and aural comparison

<table>
<thead>
<tr>
<th>% / Task</th>
<th>Checking grades</th>
<th>Checking events</th>
<th>Checking RSS news</th>
</tr>
</thead>
<tbody>
<tr>
<td>Failures</td>
<td>4.76%</td>
<td>9.29%</td>
<td>7.41%</td>
</tr>
</tbody>
</table>

5. DIDACTIC EFFECTIVENESS TESTING

The instructive character of the learning contents module, made it our subject for the assessment of didactic effectiveness. This module proposes a voice-enabled service to allow students listening to educational resources used in courses they are enrolled in. Technically speaking, contents are audio files, obtained from applying text-to-speech conversions to visual contents. The only requirement for the conversion is that visual contents, provided in any format, have to be capable of being expressed as text documents.

Our set up for the evaluation of the use of learning contents is proposed for obtaining a measure of didactic effectiveness when accessing aural materials and relate it with students’ prevalent learning styles, in such a way that allows for analysing the relationship between students’ prevalent learning styles (text/aural) and the didactic effectiveness obtained for a specific content type (text/aural). The following describes the method and results in 20 students.

5.1.1 Method, Participants and Materials

The 20 students who participated in this study were divided equally into the control and experimental group. While the ten students belonging to the experimental group gain access to the traditional visual-based learning resources of the learning site from a web browser, the ten students belonging to the control group access equivalent aural learning resources by using a VoIP client and headphones. All the students were novices or beginners in the proposed field of knowledge.

In a blended learning context, changes in students’ performance are affected by both the classroom setting and the e-learning environment, and therefore didactic effectiveness could not normally be measured separately. To allow didactic effectiveness of on-line resources to be measured independently in this study, we created a subject content to be offered exclusively through the e-learning system. The course introduces the same e-learning concepts through visual and aural modes of presentation. The audio format has a fixed duration of three minutes approximately. In the case of the visual mode, the equivalent content complements text with images, within two sides of one page, and a average reading time of two and half minutes.

5.1.2 Procedure

The following describes the steps for carrying out the evaluation:

Firstly, we ask students to perform a test on learning styles, more specifically, the VARK questionnaire (A Guide to Learning Styles; http://www.vark-learn.com/english/page.asp?p=questionnaire). VARK provides four scores for visual, aural, read-write and kinesthetic learning styles. Our goal consists in understanding the relation between a predominant learning style and didactic effectiveness.

Secondly, we ask students to answer 5 simple-answer questions designed to test the knowledge about the subject.

Then, we provide students with a didactic unit specifically designed for this study. The minimum requirement consists in accessing and listening to, or reading, the learning contents included in the unit.

Finally, we ask students to answer again the same 5 simple-answer questions test. By comparing the results in the pre and post-test it is possible to ascertain whether it has been a change in knowledge acquisition and measure the increment.

5.1.3 Results and Discussion

Our previous studies confirmed the didactic effectiveness of aural learning contents and allowed to state the auditory contents have didactic value for e-learning courses [1]. In the present study, we focus on another factor that may influence the convenience of using aural materials. Specifically, we examined the relation between students’ prevalent learning styles (text/aural) and didactic effectiveness of visual and auditory contents. Results show a positive tendency for didactic effectiveness in students with predominant read/write learning style when using text contents, as well as in students with predominant aural learning style when using auditory contents (see Figure 4 and Figure 5). The prevalence of a learning style has been calculated from the VARK index and using the ratio between a particular learning style and the sum of all the scores. Overall, we consider students scoring above 25% for a particular learning style has a Cowan index higher than average. The results achieved in this study are consistent with research works on the adaptation of e-learning systems to students’ learning styles, which are based on the principle that the performance of the students increases when their learning styles are matched with the adequate educational strategies. However, results also show how some students achieve lower didactic effectiveness than expected according to their Cowan index scores. This behaviour makes us conclude that didactic effectiveness is also conditioned by other aspect intrinsic to learning, such as students’ motivation, interests and goals.

Figure 3 Didactic effectiveness in students read/write learning style
7. ACKNOWLEDGMENTS

This work has been funded by the Department of Science and Innovation (Spain) under the National Program for Research, Development and Innovation: project TIN2011-25978 entitled “Obtaining Adaptable, Robust and Efficient Software by including Structural Reflection to Statically Typed Programming Languages”.

8. REFERENCES


