TELS: A Voice-Response Internet-based Learning System

V. Kolias [1], C. Kolias [2], I. Anagnostopoulos [2], G. Kambourakis [2], E. Kayafas [1]

[1] **National Technical University of Athens**
vkolias@medialab.ntua.gr, kayafas@cs.ntua.gr

[2] **University of the Aegean**
{kkolias, janag, gkamb}@aegean.gr

Abstract

During the last decade the academic world is continuously capitalizing on the use of Internet and web-based learning solutions, because of the simplicity and immediacy in creating, organizing and managing educational material and student data. However, the delivery of educational content to the end-user is characterized by visual presentation and the requirement of some sort of access either wired or wireless to the Internet, which blocks visually impaired individuals or people who don't have access to the Internet in one way or another from accessing educational content. In this paper we describe the design and implementation of the Internet Telephony Learning System (TELS). Besides all other, TELS exploits mature Internet/ web standards and the most popular communication mean in the world, the telephone, to provide audio interactivity between an otherwise traditional web application and the end-user. Unlike other similar applications, TELS does not need any special software or hardware to be accessed and since it is an open source traditional web application it can be custom-tailored to the individual needs of each institution. Since it is accessible to almost every communication device, we argue that it is useful for visually impaired, technologically uneducated, and underprivileged people for accessing information originally intended to be accessed visually via a Personal Computer.

Keywords: e-Learning; m-Learning; Internet / Web-based Education; Prototype implementation; Evaluation.

1. Introduction

The web had such an impact in modern societies that it has become a vital part of them. From commerce to entertainment, the web revolutionized a plethora of everyday processes. This momentum has not left education untouched. Educational institutions adopted web-based solutions not only to promote distance education but also to improve the quality of their services to both students and educators. The advent of e-learning as a teaching method during the last decade elevated the research and development of novel applications that leverage both important technological advances and prototype teaching practices.

From a technical perspective, modern web-based educational applications take advantage of a variety of features introduced mainly with web 2.0. Knowledge exchange mechanisms like wikis and forums, and standard communication mechanisms such as email and chat rooms are widely used to assist students and instructors to exchange educational material and ideas. The combination of hypertext with multimedia allowed the investigation of alternative navigation paths through educational content in high interactivity, according to the individual needs of each learner. Information exchange and distribution of educational activities was also significantly facilitated by making structured and unstructured information resources like academic library catalogues and databases widely accessible.

At the same time, the evolution of mobile communications and the continual proliferation of mobile devices turned the need for *anywhere* and *anytime* access to educational services into a necessity. However, in their attempt to enter the world of pervasive computing today's web-based educational applications seem to lack serving the third and perhaps most important dimension of it: *anyone*. Most of the existing web applications are primarily dependent on web standards such as the Hypertext Markup Language (HTML), which present information in a visual manner. In this way, visually impaired individuals are unable to use and take advantage of a variety of services. Another fundamental constraint of web applications is that they can only be accessed via some kind of Internet/web-enabled device. According to [Miniwatts Marketing Group, 2009], the penetration of the Internet among the population in Greece is estimated at 46%, while in Europe the same percentage is 48.9%. These numbers are significantly lower, in emerging economies in Africa or Asia, where Internet usage is about 5.6% and 17.2% on the total population respectively. This fact highlights the undesirable blockade to important web-based services that is mainly caused by the inability of underprivileged people to afford a computer or a web-enabled handheld device and the disinterest of some people to acquire basic IT skills.

On the other hand, traditional (fixed) Public Switched Telephone Network (PSTN) and mobile phones are clearly much more accepted and they have dominated the population in many countries due to the fact that their cost is significantly lower than a personal computer and their use does not require any special IT skills. In Greece, for example, the penetration of mobile phones is estimated at 110.3% of the total population, while in Europe it is estimated around 110.9% [International Telecommunications Union, 2007]. Additionally, a relative research showed that the 44% of Greek mobile phone users use their cell phones only for making and receiving calls and not for other services (e.g. SMS), while declaring that they are not interested to learn about them [Focus Bari, 2008]. Mobile phone use is much higher than Internet use in Africa and Asia as well where it is estimated at 28.44% and 37.66% respectively. When one combines the aforementioned indicative percentages with the wide use of fixed telephone lines, which usually correspond to more than one person, clearly conceptualizes that access to the telephony network is nowadays considered a common commodity practically everywhere.

In this context, harnessing the most natural and intuitive communication modality available to man, namely speech, via the telephony network would serve the delivery of web content dually: not only would users enjoy a quicker and smoother interaction with traditional web applications, but also they would have a truly anytime/anywhere access affordably. Speech processing technologies render such an attempt feasible with the development of user friendly Voice User Interfaces (VUIs) that understand natural, conversational language. Speech enabled VUIs do not interfere with other (visual) manual tasks, such as driving, while being able to synthesize dynamically generated content via text-to-speech technologies, this way outclassing other human-computer interaction paradigms.

Furthermore, the World Wide Web Consortium (W3C) web standards that facilitate the development of interactive voice dialogues between a human and a computer have matured through the last decade. Voice eXtensible Markup Language (VoiceXML), which is considered the audio equivalent of the HTML, enables the development of high level, platform independent Interactive Voice Response (IVR) dialogs [W3C-VoiceXML]. The set of words that a speech recognition engine may recognize during one interaction is specified with the Speech Recognition Grammar Specification (SRGS) [W3C-SRGS]. Finally, the overall aesthetic result of synthesized speech is further improved with the Speech Synthesis Markup Language (SSML), which is a standard mechanism for controlling various aspects of speech, such as pronunciation, volume, pitch and rate [W3C-SSML].

The Multimedia Technology Laboratory (Medialab) of the National Technical University of Athens (NTUA) responds to the challenge of offering pervasive educational services with TELS: the TElephony Learning System. Developed for the needs of Medialab's taught courses, namely Multimedia Technology and Computer Graphics with OpenGL, TELS aims to deliver teaching material visually, through a web browser via a Personal Computer (PC) or mobile device and acoustically through a fixed, mobile or VoIP phone. TELS is currently addressed to an audience familiarized with information technologies, that is, the students of the Electrical and Computer Engineering Department of NTUA, the vast majority of which owns a PC and mobile phone. This audience can benefit from the system's availability and directness. For example, a student is able to call TELS on his way to the campus to hear about an unexpected cancelation of the day's course or a quick synopsis of the day's lecture. On top of that, the integration of TELS with Medialab's existing e-learning services, which deliver a larger set of courses to a wider audience of all age and educational backgrounds, would reinforce the system's pervasive nature. Therefore, the multiple ways of content representation contributes to the purpose of pervasive learning and may also offer great help to disabled individuals and people in developing countries.

The rest of this paper is organized as follows. The next section presents related work on the topic. Section 3 presents the system's services in detail while section 4 analyzes its architecture. Section 5 is twofold; first it presents the results of a quantitative pilot survey conducted on the first people who tested the application and secondly, some quality of service results regarding the effectiveness of a typical call to the system. The last section offers some concluding thoughts and future directions for this work.

2. Related Work

The significant advances in speech recognition and synthetic voice production only recently allowed for efforts regarding the presentation of dynamically generated content via acoustic means. In the e/m-learning field, this work can be classified mainly in two categories that differ in the way they manipulate data and the audience they refer to. The first involves the conversion of dynamically generated web content that was meant to be presented visually in an audio format that can be accessed either acoustically by phone or by a web browser. The second category involves the generation and management of dynamically created audio content without any textual counterpart. The target group of applications included in both categories is the IT illiterate people in developing economies or visually impaired individuals. In the following we present three distinctive works from the first category and three from the second.

The MobilED initiative [MobileEd Initiative, 2009], motivated by the fact that mobile phones have penetrated more than PCs into young people, due to their low cost and ease of use, involved an exploratory research study considering the use of mobile phones in lower education in South Africa [Ford & Botha, 2007]. By using the popular open source feature-reach wiki engine MediaWiki to develop a source of educational material for the pupils, the main idea behind this project was to enable the acoustic access of wiki content via phone. The proposed implementation, built solely upon open source components and requires students to send a short message (SMS) from their mobile phone with the title of the article they wish to hear. After a while the service responds with a narration of the article content with a synthetic voice directly to the pupils' mobile phone. During the call the user can navigate to different sections of the article using Dial Tone Multi Frequency (DTMF) and contribute to an article by entering voice annotations. The results of a two pilot studies showed that the pupils were focused on learning the content itself and not the technology. The whole process not only loosened the role of the instructor making him a participant but also the anytime-anywhere concept, motivated students to engage more often. However, since this service targets students in developing countries, deployment costs (mainly sending SMS and providing content via phone) could become prohibitive if the service availability is increased. Also, the fact that the application requires the typing of an SMS makes it inappropriate not only to visually impaired individuals, but also to people who do not own a mobile phone or cannot/don't know how to send an SMS (like students in rural areas or even elder people). Although mobile phones are very popular nowadays, ordinary PSTN phones are still used by the majority of people.

In [Werner et al., 2009] the authors propose a client/server architecture to integrate speech technology into web pages to be used for e-learning purposes. They utilize a central speech server with speech synthesis, speech recognition and speaker verification capabilities, which is responsible for the transformation of speech to text and vice versa. The content produced is presented to the user through his web-browser with a Java applet that implements audio input and output capabilities. However, the requirements on the client side for this approach, i.e., a JavaScript enabled browser and Sun's Java plugin, may not be supported by current mobile devices. Additionally, access to the Internet is not guaranteed when the user is on the move or simply when he does not have an Internet connection or a PC.

In [Borodin et al., 2007] the authors present a non-visual web browser that enables visually impaired people to navigate the contents of web pages acoustically. A synthetic speech feature transforms the contents of web pages into sound and the open source Sphinx voice recognition engine [CMU Sphinx Group] transforms the user's voice into signals which are recognized by the system. In this way one can navigate through pages with his voice only and can avoid the sequential narration of their content. This application provides significant advantages especially to partially blind users that are not willing to invest time and effort to learn new communication means or acquire IT skills. On the other hand, it runs exclusively on PCs, making it inaccessible to users who do not own or do not know how to use a computer. In addition, there is no lightweight version targeting to mobile devices making it inappropriate for roaming users.

In [Wang et al., 2008] the authors claim that the collaborative nature of wikis is not well served because it limits its users to computer environments or, when deployed in mobile environments, it restricts them by means of input (keyboard stroke or stylus) and output (small screen). They identify that synchronous communication technologies like teleconferences or simple calls are gaining attention as channels to carry out collaborative tasks. Therefore, the combination of the wiki collaboration paradigm with audio communication means can improve the overall usability of wikis. Under this context they propose a wiki implementation to facilitate asynchronous audio-mediated collaboration when on the move. It is based solely on the manipulation of audio files. Despite the fact that the proposed system enhances collaboration with a more personalized feeling – each user contributes with his own voice – it does not have any web counterpart and therefore it cannot be accessed by any mean other than acoustic.

The World Wide Telecom Web (WWTW) [Kumar et al., 2007] is a promising concept that envisions the delivery of services similar to that of WWW to the underprivileged communities, by enabling the access to information and services through voice driven channels that leverage the existing Internet infrastructure. The WWW WebSite counterpart in the WWTW is the VoiceSite and its unique identifier is the VoiNumber, a virtual phone number that maps onto a physical phone number or a Uniform Resource Identifier (URI) such as a SIP URI. VoiceSites can be created easily even by IT illiterate people through a voice driven interface described in [Kumar et al., 2007b] that in essence composes custom voice applications with reusable existing components such as dialogs and other web services. VoiceSites are linked together via VoiLinks over the HyperSpeech Transfer Protocol (HSTP) described in [Agarwal et al., 2007] that enables cross organizational transactions and navigation in speech. An important aspect of WWTW is searching information in VoiceSites or the T-Web in general. Since the primary content is voice, searching audio data is essential, yet difficult compared to that of restricting search in related meta-information. Finally, another aspect of great importance here is the integration of VoiceSites with information and services available in the WWW. This can be achieved either with the integration with local applications (like databases) or with remote applications exposed through standards such as web services. The WWTW is a whole new world of information exchanging, however limited in the scope of application development since its applications rely in a set of redistributed components.

A DAISY Digital Talking Book (DTB) is a file with a specific format developed by DAISY Consortium [DAISY Consortium]. A DTB presents content in multimode form including text, audio and graphics. Readers can easily navigate in the document's logical structure by chapters, headings, pages, paragraphs and sentences. A DTB can be accessed either from DAISY Format capable special devices or PCs. The purpose of this technology is to enable people with reading disabilities to access information provided by mainstream publishers, governments and libraries. Authors in [Spinczyk & Brzoza, 2008] prove in practice the feasibility of developing an educational system based in DTB format. However, the DTB is not a widely adopted standard like VoiceXML and therefore it is not suitable to be adopted for the integration to large scale dynamic systems and web applications. Additionally, from a user's point of view the DTB requires special software and/or hardware and the related skills to use it that may lead to extra costs.

From the related applications that were presented above, one can clearly distinguish some major or minor difficulties and limitations, that although are insignificant to many people, to others are barriers to accessing their content. In more detail, sending a request for content via SMS, is a difficult task for the visually impaired. Also in at least three from the applications above, special software is needed to be installed in the client's PC, in order to access the desirable content. At the same time, some of the aforementioned applications can only be accessed in one way, thus dispossessing the ability to interact with them when on the move. Finally, almost all applications require the direct or indirect use of a PC. This is a mojor restricting factor which blocks out those who do not own or do not have access to PCs, or even those who simply cannot use it. On the other hand TELS combines the convenience and ease of development of a web application with the immediacy and the low cost of the telephone. These characteristics enable – and in some cases encourage - access to educational material to the widest range of people possible.

3. TELS Features

TELS is a fully functional e/m-learning platform that can be seen as an aggregate of typical e-learning tools. This result aims to promote student collaboration in traditional courses whilst providing a comprehensive solution to the problem of learning content management. While being tailored to the specific needs of Medialab's curriculum, TELS maintains the generic characteristics that allow for use in a wider spectrum. What differentiates TELS from other similar approaches as already mentioned is the acoustic representation of its content. Though serving the pervasive nature of the platform, this characteristic dictates the adjustment of content that was meant to be available visually into content for acoustic presentation. For example, the only way to present pictures acoustically, is by presenting their metadata.

In Medialab's deployment of TELS four roles of users exist: (a) Administrators who set the platform's general settings, integrate the desirable curriculum into the platform and assign the tutors to each lesson, (b) Tutors who specify and organize the learning material into the course's calendar they were been assigned, (c) Students or generally learners, who receive the learning services, and (d) Assistants whose main mission is to provide help and guidance to the students in the learning process. Further down we analytically present the characteristics and functionalities for each TELS features.

3.1. Administrative Console

Among other administrative tasks, the administrative console provides access to the platform’s basic settings, user management, and lesson and curriculum specification. It is accessible only to administrators.

3.2. E-Lecture

A central part regarding the TELS platform is the E-Lecture. It provides a graphical interface for the management and access of learning material specified in each lesson’s calendar. More specifically, a Tutor uses E-Lecture for one of the following tasks: (a) to add a lecture’s presentation according to the lesson’s calendar, (b) to add the summary of each lecture of that lesson, add notes or make extra comments, provide links to other parts of the platform or any other relative multimedia with educational content etc., (c) to assign tasks or exercises to students. It is clear that E-Lecture is in essence the part that initiates study and collaboration. Therefore, a student will access E-Lecture to catch up with the content of lectures he might have missed or access any existing additional learning resources or educational material for further study. Although all features mentioned above are meant to be accessed via a web browser, when accessed acoustically they are presented partially. More specifically, the system provides just an indication for the existence of multimedia content, such as pdf documents, powerpoint presentations, videos, pictures etc. rather than attempting to transform the content of such files into sound. Additionally, all the functionality for creating and managing the content of lectures has been intentionally left out. In this case the goal was to offer a quick way for students to keep pace with their lectures and assist educators on their task and not to provide a full content management system over the phone.

3.3. Wiki

Wikis - perhaps the most important feature introduced with web 2.0 - are generally considered as collaboration platforms where users, access and/or contribute knowledge on specific topics. Although they were introduced more than ten years ago, their use in higher education was only explored in the last five years. The success of Wikipedia, a web encyclopedia and perhaps the most famous wiki application nowadays, attracted the attention of educators who anticipate that wikis will significantly assist essential educational processes such as communication and collaborative finding, shaping and sharing of knowledge [Reinhold, 2006].

The wiki of TELS can be viewed as a central repository of educational material mainly produced by the asynchronous contribution of students on specific topics organized in lessons. Users are able to create new wiki articles, contribute by adding their own knowledge to existing ones, reviewing the articles or rollback unwanted changes. In fact students are obligated to contribute an article as part of the lesson's requirements. Along with a mandatory peer assessment of the article quality, contribution on a specific article is guaranteed at least upon its creation. As far as the audio representation of wiki content, though feasible, the creation of an article is allowed only through the web interface. However, updating wiki content can also achieved via phone by adding voice annotations that can be accessed both through the visual and the acoustic interface.

3.4. Discussion Forum

In the context of education, online discussion forums can be seen as virtual learning environments where discussion on educational topics can prove at least as beneficial as course materials or lectures. During the last decade it has been shown that online discussion increases participation and collaborative thinking while promoting critical analysis and the social construction of knowledge through the provision of asynchronous, nonhierarchical and reciprocal communication environments [Ruberg et al., 1996; Warschauer, 1997; Dehler & Parras-Hernandez, 1998].

TELS Discussion Forum allows users to initiate discussions and receive replies on a specific topic. Typically, the forum is comprised of topics organized in lessons. A user is able to start a new topic only through a web browser but can access the full discussion either by the visual web interface or the voice interface pretty much as in the case of wiki. On the top of that, it is possible to be part of the topics discussion by giving an answer vocally via telephone. In this case the voice of the user will be recorded and stored on the server as an audio file. Then, when accessing the answer from a web browser the user will listen to it from the media player components embedded to the page, or the answer will be given in the user's own voice from the phone.

3.5. Home

As the name implies it is the initial page of both the visual and acoustic interface that provides a summary (map) to the lessons, wiki articles, and discussions status in which a user participates. Its primary goal is to facilitate the quick access to trivial information via telephone, such as last minute announcements.

4. TELS General Architecture

TELS is a set of software and hardware components that divide functionality among multiple applications while cooperating in order to deliver the creation, organization and management of educational material. It facilitates the organization of traditional courses while enhancing them with mechanisms that promote collaboration among learners. TELS can only be hosted in Windows operating system. However its twofold content delivery, meaning the use of HTTP to transmit data for visual representation and Session Initiation Protocol (SIP) or the traditional PSTN to transmit data for acoustic representation, makes it accessible to users on a variety of devices and client platforms, such as wired and mobile phones or Personal Digital Assistants (PDA) and PCs running Linux, Macintosh or Windows. As already mentioned TELS is made of multiple cross-interacting components, each being highly configurable in order to be extended or modified as needed. In the following we provide a technical description of the TELS architecture by examining its major components and the way they interact.

4.1. TELS Components

Further down we describe the TELS components and their purpose, while Figure 1 depicts the overall architecture of the system:

The database server component houses the TELS database. It is responsible for supporting all the aforementioned features of the system. The database stores all information regarding administration, users, lessons, wiki articles, discussions and other multimedia files. All wiki content is stored in the database along with presentation information. This information has been defined by the user during an article creation. For convenience, wiki information (content and presentation information) is stored as an XML document complying to the TELS ML, a custom markup language designed to fill the needs of TELS.

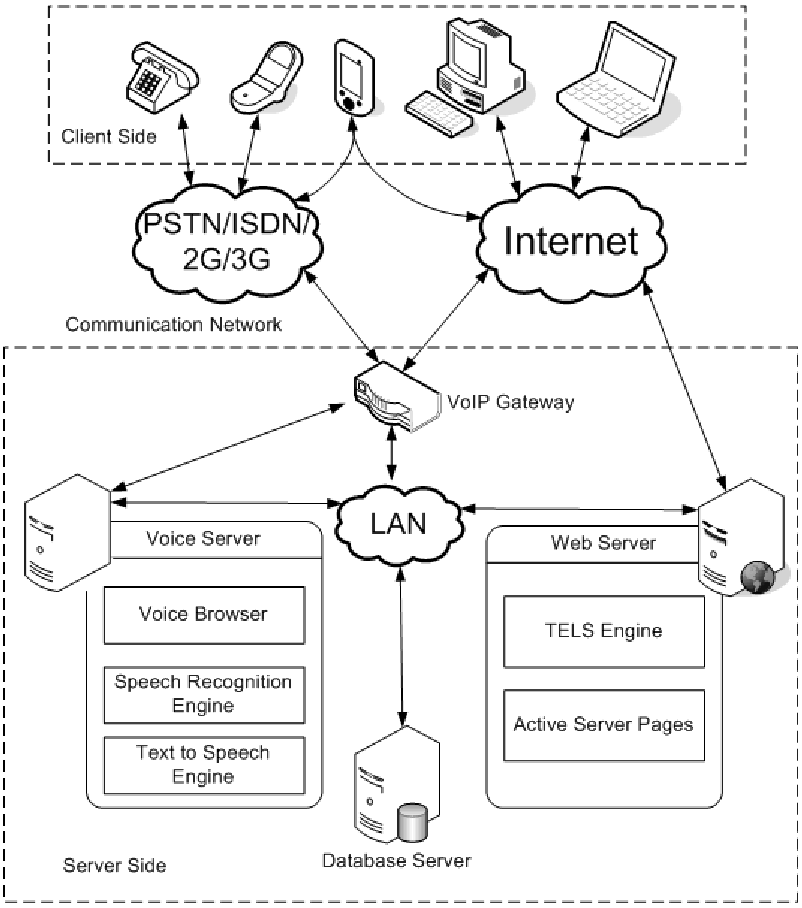


Figure 1. The general architecture of TELS

The Web server component accommodates the TELS web pages and it is responsible to provide the users with the corresponding web services.

The TELS engine is the core component of TELS. It is responsible for the coordination of all other subsystems and for the implementation of the business logic[1] of the system. It is essentially a web application residing on the web server, which depending on the client requests, generates XHTML or VoiceXML pages dynamically and sends them to the user's web browser or the Voice Server component respectively. Upon a wiki content request, it transforms TELS ML to XHTML or VoiceXML according to specific XSLT rules.

[1] Business logic is a non-technical term generally used to describe the functional algorithms that handle information exchange between a database and a user interface. It models real life business objects (such as accounts, loans, itineraries, and inventories), prescribes how business objects interact with one another and enforces the routes and the methods by which business objects are accessed and updated.

The Voice Server is responsible for the acceptance of audio requests from and the delivery of audio content to the PSTN or VoIP network through the VoIP Gateway component. The Voice Server is further comprised of a Voice Browser, a Text to Speech engine and an Automatic Speech Recognition engine each in charge with the implying tasks.

The VoIP Gateway receives calls from the PSTN, converts PSTN signals to VoIP signals and forwards them to the Voice Server component.

The Communication Network component binds together the Database server with the web Server and therefore the TELS engine. This network may be implemented using a Local Area Network (LAN) or remotely via a corresponding Internet connection.

The Client is any registered user to TELS. Clients may access TELS using various devices: wired or mobile phones, PDAs and laptop or desktop computers.

The PSTN is utilized to transfer incoming client calls to the system.

The Internet infrastructure is used to retrieve the corresponding data depending on the request. Therefore users may request and receive web pages via their web browser or audio signals via a softphone like X-Lite.

The technical requirements of the Telephony Learning System are presented in Table 1.

4.2. A closer look at the Voice Server

The Voice Server is responsible for the transformation of text documents to sound and the validation of the user’s voice input against predefined sets of recognizable words. The voice server further consists of a set of subcomponents: (a) the Voice Browser, (b) the Text-To-Speech (TTS) engine and (c) an Automatic Speech Recognition (ASR) engine. The Voice Browser:

1. Receives a request from the user.
2. Receives VoiceXML and grammar files from the Web Server.
3. Specifies the execution flow according to the instructions in the VoiceXML file. It isolates the text to be spoken, forwards it to the TTS engine and it forwards the grammar to the ASR engine.

4. Generates the request made by the user and forwards it to the Web Server.

The Text To Speech engine receives the text from the VoiceXML file meant to be spoken, transforms it into streaming sound and sends it to the VoIP Gateway to forward it to the end-user. On the other hand, the ASR engine receives a grammar, that is a set of terms that is able to recognize along with the client prompt and identifies if the prompt corresponds to any word in the grammar. If true it returns the term textually. The VoIP gateway receives calls from the PSTN, converts PSTN signals into VoIP signals and forwards them to the Voice Server. It is to be noted that the voice server will accept only VoIP signals. Signals originating from the Internet (from VoIP clients) might be SIP signals or signals of some proprietary VoIP protocol, like Skype. The VoIP gateway is also responsible for the transformation of VoIP signals from the Internet to the protocol that the Voice Server recognizes.

4.3. Components Interaction

All users are essentially served by the TELS engine that resides at the web server; however they may interact with the system visually or acoustically, depending on the device they use. When a client calls the system via a mobile or fixed telephone or through a VoIP client installed in a PC or PDA it interacts with Voice User Interfaces When a client accesses the system through a web browser via a computer it interacts with web Interfaces. A typical interaction of the system's components during a telephony session is the following:

1. A client request (speech as a telephone call signal) arrives at the VoIP Gateway through the PSTN or the Internet. The VoIP Gateway component transforms the request signal accordingly and forwards it to the Voice Server. The Voice Browser component of the Server analyzes the request and tries to recognize the given command according to a specific grammar file with the ASR engine. Then it makes a request for the requested VoiceXML document with the appropriate content (that resides in the Database Server) to the Web server. Note that all content resides in the Database Server.

2. The TELS engine receives the corresponding data from the database to construct the appropriate response. At this phase the content retrieved is in raw TELS ML format which means that it contains the requested data along with some markup presentation information. During this step, the transformation process of the TELS engine converts the TELS ML document into the equivalent VoiceXML document according to specific XSLT rules.
3. The TELS engine forwards the dynamically generated VoiceXML document to the Voice Server for further process. The Voice Browser component of the Voice Server analyses the instructions contained in it in order to plan the execution flow of the call. At the same time it sends to the TTS component, any text to be converted to sound and the grammar file with all words to be expected as further input to the ASR component.
4. The response in the form of sound produced by the TTS engine is sent back to the VoIP gateway that converts it into normal PSTN signals or VoIP signals. The latter are sent back to the client.

On the other hand, during a typical web session the above procedure differs in the first step, where the client requests reach the Web server directly through the Internet, in the second step where the TELS ML is converted to XHTML, and the fourth step where the dynamically produced XHTML document is returned to the client directly through the Internet. The third step is not included since the TTS or speech recognition procedure is not taking place.

4.4. TELS User Interface Description

In this section we describe a typical user interaction with perhaps the most representative feature of TELS, regarding voice interaction: the wiki. The scenario presented includes all the steps required in order for a user to connect to the system, retrieve and update an article and finally exit the application. This will provide a thorough understanding of how the application functions, since interaction with the rest of the system components is quite similar. Each TELS feature can be accessed by either using a telephony client or a web browser.

4.4.1. The Voice User Interface

Initially the user makes a call to a publicly known number bound to the application, either from his landline phone, his mobile phone, or his VoIP client installed on his PC or PDA. TELS gives a generic welcome message and prompts the user to provide his credentials, and more specifically to pronounce the digits of his unique ID one-by-one, in order to sign in. After a successful sign in, TELS greets the user with a personal message and redirects him to the Home section. From there he is informed about the status of his active projects, any new announcements or deadlines of his essays. He also gets informed about activities in the rest of the components of the system, for example that he has received new posts on one of his topics on the Forum etc. Finally, he has the option to navigate to any of the other system components. In this case, we assume that the user chooses to interact with the Wiki component in order to listen and update a specific article with a short comment.

First off, he is prompted to provide an article title or a term contained in an article. The difference between the two options lies in the grammar files created and sent to the user. In the first case the grammar file contains all the possible article titles and it is usually chosen when a user knows exactly what he is searching for. In the second case however, the grammar file contains the most popular terms of each article not including noise words i.e., extremely common words contained in each article such as "and", "or", "the", etc. Also, in the second case, the grammar files produced can be extremely large in size therefore, for the sake of efficiency, the search terms are split into more than one grammar files. If the search term is found then the articles containing it are presented to the user ordered by a relevance percentage. It is obvious that this method might be time consuming, especially if many lengthy articles exist, therefore it is recommended only when a user has a small clue on what he is searching for. The chosen article is acoustically presented to the user. During this presentation he is able to navigate through each section, by giving the appropriate commands like "next", "previous" or just by speaking the paragraph titles. Throughout the process the user decides to add a voice annotation (by pronouncing the appropriate voice command) at a specific point in the article flow, thus he provides a short voice message. The message is recorded and stored at the database server. The next time someone would want to access the same article the comment with the user's voice will be heard at the point where this annotation was added, interrupting the synthetic speech produced by the TTS engine. Figure 2 depicts the sequence diagram for the use case described above, while Table 2 provides the dialog for this scenario.

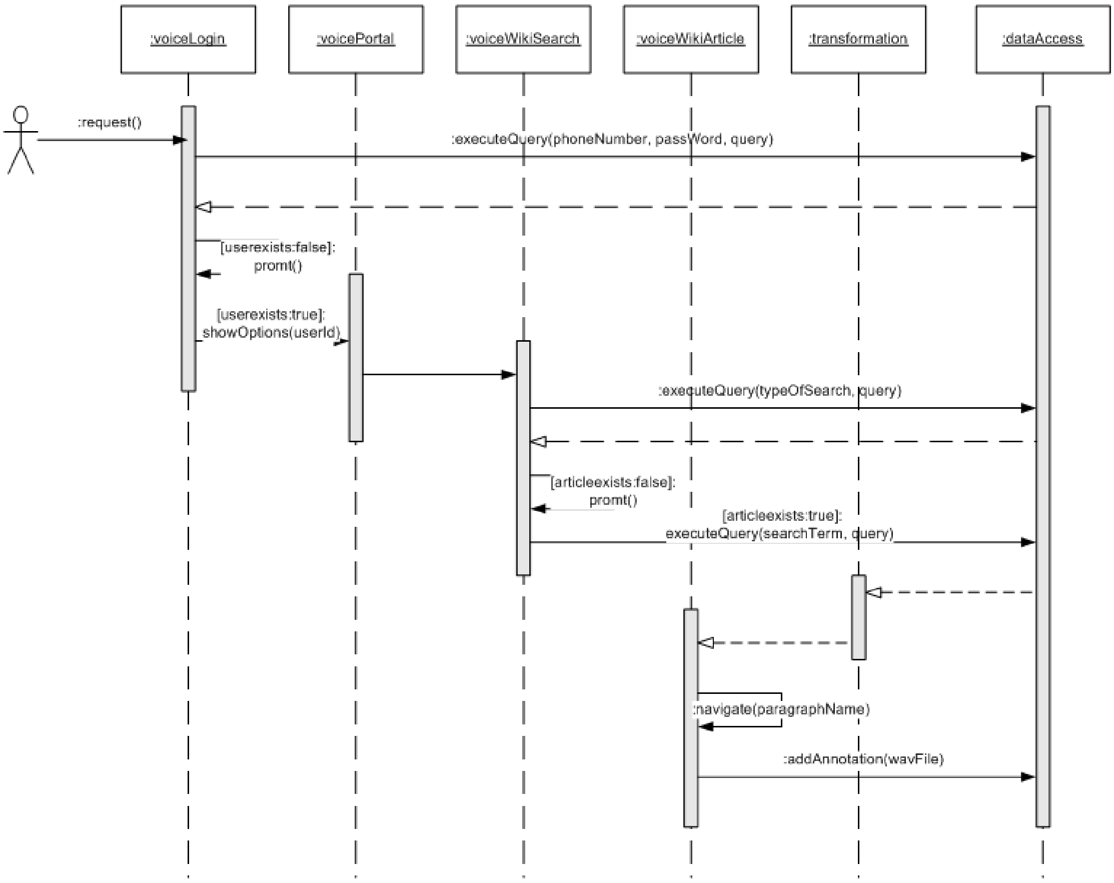
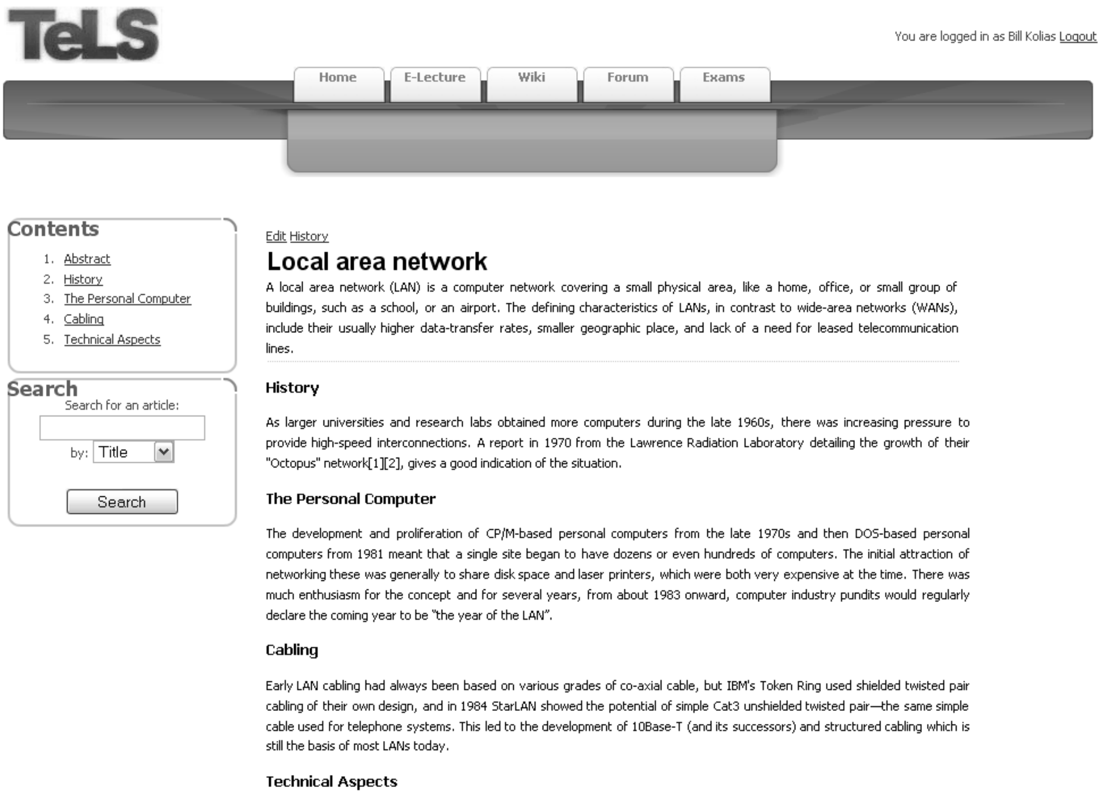


Figure 2. Sequence Diagram for accessing and updating an article of the wiki component

4.4.2. The Web Interface

The Web Interface is a set of common web pages that provide access to the TELS services through a web browser. A typical user interaction scenario includes the following actions taken by the user. First the user connects to the initial page where he provides his login credentials. After a successful login, he connects to the Home page, where he is informed about the status of his active projects, new announcements, homework deadlines and other activities in the rest of the system components, much as he does in the voice user interface. He is also able to navigate to the rest of the TELS features. The Web user interface of the wiki component in particular is presented in Figure 3.



Local area network

A local area network (LAN) is a computer network covering a small physical area, like a home, office, or small group of buildings, such as a school, or an airport. The defining characteristics of LANs, in contrast to wide-area networks (WANs), include their usually higher data-transfer rates, smaller geographic place, and lack of a need for leased telecommunication lines.

History

As larger universities and research labs obtained more computers during the late 1960s, there was increasing pressure to provide high-speed interconnections. A report in 1970 from the Lawrence Radiation Laboratory detailing the growth of their "Octopus" network[1][2], gives a good indication of the situation.

The Personal Computer

The development and proliferation of CP/M-based personal computers from the late 1970s and then DOS-based personal computers from 1981 meant that a single site began to have dozens or even hundreds of computers. The initial attraction of networking these was generally to share disk space and laser printers, which were both very expensive at the time. There was much enthusiasm for the concept and for several years, from about 1983 onward, computer industry pundits would regularly declare the coming year to be "the year of the LAN".

Cabling

Early LAN cabling had always been based on various grades of co-axial cable, but IBM's Token Ring used shielded twisted pair cabling of their own design, and in 1984 StarLAN showed the potential of simple Cat3 unshielded twisted pair—the same simple cable used for telephone systems. This led to the development of 10Base-T (and its successors) and structured cabling which is still the basis of most LANs today.

Technical Aspects

Figure 3. TELS web Interface: the wiki component

5. Evaluation of TELS

The ability of TELS to be accessed via the Internet or a telephone network, wired or wireless is the main characteristic that joins together TELS with pervasive applications. Additionally, the extensibility of TELS allows it to be incorporated to existing applications of diverse institutions such us, museums, city guides etc. Users may significantly benefit from the use of TELS because:

- They can access most of the functionality vocally/acoustically when on the road or in places where Internet access is not available. Especially students, who are the main focus of TELS, can benefit from that since they are able to access almost all the features of the platform.

- TELS has also low operational costs. Users are charged with rates for a common call when they access TELS over a PSTN network, with the rates of a mobile phone call when they access it over a GSM or UMTS network and of course they access the application free of charge, from a SIP softphone or from a web browser.
- The acoustic nature of TELS makes it also ideal for persons with special needs such as the ones with sight disabilities.
- Technologically illiterate people would surely appreciate the use of the voice interface.

On the other hand, TELS has some limitations that originate mostly from its dual content presentation nature:

- It is already clear that some of the functionality is not possible to be ported to a voice equivalent. For instance, multimedia resources can only be described when accessed acoustically, while the functionality relevant with the administrative tasks are left out since the idea is to facilitate the instant access to educational content only.
- The quality of services the users enjoy and therefore the success of the application heavily depends on the performance of the TTS and Voice-Recognition components. In turn, the efficiency of the Voice-Recognition component depends on the quality of the user’s input. Therefore, it is expected that the experience of the user will not be the same when accessing the application by different means. More specifically, the user might be forced to provide inputs to the system repeatedly when accessing the application from a mobile phone under noisy environments.
- The sequential nature of the speech interface might act as a suppressive factor for the application’s success to users familiarized with PC use. Indeed, the amount of information that a user can receive when accessing the application acoustically on the same unit of time is significantly smaller. On top of that, locating specific parts of the information in this way might demand more interaction with the system which also gives room for error.

- Despite the fact that its use comes to a low cost for the end user, TELS has high deployment costs. Although TELS is an open source application its implementation components are not. The Speech server, SQLServer and Windows operating system have commercial licenses for their use. Depending on the choice of hardware components and the editions of the necessary software (see Table 1) a rough estimation of the implementation cost would range from 3000 € to 10.000 €. In more details, the minimum initial implementation cost would require a low cost server system which would host all TELS software components and a low cost voice gateway. In that case TELS would be set using the limited (and free) editions of the database, server operating system and speech server software, without loosing any of its original functionality. Even if this cost is relatively low, for some institutions it is still prohibitive. Cooperation with other institutions or academic versions of the necessary software is a possible solution to this problem..

5.1. Call and Voice Quality Evaluation

5.1.1. Methodology and related work

A typical call to the system as stated in the previous paragraphs may originate either from the PSTN or a VoIP client. In each case the call arrives at the voice gateway which routes the call over the IP network. Conversely, it routes calls from the IP network back to the Switch network. Call routing is set in its simplest form: only the Voice Server is configured to receive calls through the gateway. Thus, all inbound all inbound Switch-to-IP calls are sent to a single endpoint. The underlying signaling protocol that enables the Voice Server to support VoIP is SIP (RFC 3261), while voice is delivered using RTP/RTCP. When a client wants to place a call to the system, an INVITE message is sent to the Voice Server, which contains data such as its location and IP address. Once the server accepts the INVITE, the two parts can communicate directly. A detailed description of the exchanging messages can be found in the RFC 3261.

Systems used in speech communication, need to transmit voice signals (analog/digital) or data in an acceptable quality threshold, in order to have a comprehensive vocal information exchange from at least two communication parts. However, the definition of that threshold varies according to the medium used, the application and the information needs. That is why voice quality is a multi-dimensional and a non-trivial problem. For measuring voice quality two distinct evaluation methods exist, namely the subjective and objective testing methods.

In the subjective-based methods, speech samples are presented to an evaluation group of listeners who rate the quality of the vocal information using an integer opinion score. All scores are then averaged to produce a Mean Opinion Score (MOS) value. However, to minimize the variability of the rating, a significant amount of evaluators are required [Ding, 2009]. The most common rating method is Absolute Category Rating (ACR), where the evaluators use a five-point scale to judge the quality without comparison to a reference vocal information. Subjective testing is highly reliable, but it is time-consuming, expensive, and the results are subject to human perspective and the test settings. As a result, such tests are usually performed once and cannot be automated. Thus, in a need to substitute these methods, the objective-based testing methods have been developed as a solution for effectively measuring voice quality. These methods consist of algorithms carried out by devices, which do not require any intervention from evaluators. Objective tests are further categorized in the intrusive and non-intrusive methods, based on whether a reference voice signal is used or not respectively.

Through the intrusive-based speech quality estimation techniques, researchers compare a test speech signal, as reconstructed in respect to the reference signal [Raja et al., 2007]. A well-known intrusive estimation model is the ITU-T P.862 (PESQ model) [ITU-T, Rec. P.862, 2001]. However, its results depend on the time or frequency domain analysis of the speech signal under test. It also requires the test call to be recorded for a considerable duration before it can be analysed. That is the main reason for not being suitable for real-time and continuous monitoring of speech quality.

On the other hand, the non-intrusive-based methods estimate the quality of the transmitted signal without having a reference signal for comparison purposes. Sun an approach is highly effective in environments where the reference speech signal is not accessible. A relatively new ITU-T Recommendation of this kind is P.563 [ITU-T, Rec. P.563, 2005]. However, the main application of this recommendation is to narrowband telephony applications [Raja et al., 2007].

So, in this paper we used a non-intrusive objective evaluation method, calculating parameters of the ITU-T Recommendation G.107 Protocol (also known as E-model) [ITU-T, Rec. G.107,, 2005]. E-model was developed by the European Telecommunications Standards Institute (ETSI) and adopted by International Telecommunication Union (ITU). This model presents different versions for various network conditions such as codec type and bursty or non-bursty network conditions. Moreover, it is restricted to a limited number of codecs and network conditions due to its reliance on subjective tests [ITU-T, Rec. P.833, 2001].In this objective-based method, several QoS parameters and metrics are involved for measuring voice quality such as the Signal-to-Noise Ratio (SNR), packet delays and inter-arrival variations, jitter, etc. There are several works in the respective literature identifying areas that could potentially influence the performance of voice when transported over IP-based networks and simultaneously competing with those QoS parameters and metrics [Zeadally et al., 2004; Zeadally & Siddiqui, 2004]. Through these international standardized metrics, the protocol provides a transmission Rating factor (R-Factor), which varies between 0 and 100, and can be then interpreted in several subjective evaluation values, such as the commonly used MOS value. Equation 1 provides the estimated MOS value when using the objective E-model, through the calculated R-Factor defined in Equation 2.

$$\begin{aligned} &MOS = 1 + 0.035R + 7\cdot 10^{-6}R(R-60)(100-R), \quad when\ 0 < R < 100 \\ &MOS = \{1\ \ or\ 4.5\}, \quad when\ R \le 0\ or\ R \ge 100 \end{aligned} \qquad (1)$$

$$R = R_o - I_s - I_d - I_e + A \qquad (2)$$

In Equation 2 R_o is a basic SNR value, which includes several noise sources such as circuit and room noises and is defined in the following Equation 3.

$$Ro = 15 - 1.5(SLR + No) \qquad (3)$$

Moreover, N_0 is the power addition of several separate noise sources as defined in Equation 4.

$$No = 10\log\left[10^{\frac{Nc}{10}} + 10^{\frac{Nos}{10}} + 10^{\frac{Nor}{10}} + 10^{\frac{Nfo}{10}}\right] \qquad (4)$$

In Equation 4 N_c (in dBmp) stands as the sum of all circuit noises referred to the zero dB reference point (0dBr), while N_{os} is the equivalent circuit noise at 0dBr, caused by the noise measured in the room at the sender side (P_s) and is defined according to Equation 5, where D_s ranges between -3 and 3 (default value is set equal to 3). In this work D_s was set equal to 3.

$$Nos = Ps - SLR - Ds - 100 + 0.004(Ps - OLR - Ds - 14)^2 \qquad (5)$$

Similarly, N_{or} (in dBmp) is the equivalent circuit noise at the levels of 0dBr, caused by the noise measured in the room at the receiver side (P_r), and it is defined according to Equation 6, where $Pre = Pr + 10\log\left[1 + 10^{\frac{(10-LSTR)}{10}}\right]$ is the "effective room noise" caused by the enhancement of P_r by the receiver's sidetone path (in dB). *LSTR* (in dB) stands for the Listener Sidetone Rating and it ranges between 13 and 23 (default value is equal to 18). In this work *LSTR* was set equal to 18.

$$Nor = RLR - 121 + Pre + 0.008(Pre - 35)^2 \qquad (6)$$

Finally, N_{fo} (in dBmp) represents the so-called "noise floor" at the receiver side referred to the zero dB reference point, and it is equal to $Nfo = Nfor + RLR$, where $\mathrm{N_{for}}$ equals to -64dBmp. In all equations above $OLR = SLR + RLR$, where OLR, SLR, and RLR is the Overall, the Sender, and the Receiver Loudness Rating respectively (in dB).

The rest metrics in Equation 2 (I_s, I_d, and I_e) are dedicated to the calculation of distortions occurred in the voice signal (combination impairment factor), distortions caused by end-to-end delays and echoing (delay impairment factor), as well as distortions caused by coding-decoding and packet loss (equipment impairment factor) respectively. Finnally, the Advantage Factor (A) is a threshold value used for fine-tuning the acceptable level of a voice signal. This factor ranges between o and 20 but usually its value is set equal to zero (in this work *A*=0). According to the ITU recommendations described in [ITU-T, Rec. G.107,, 2005], there is no relation of the Advantage Factor to all other transmission parameters, while its assigned value depends on the telecommunication technology, the context and the devices used. Table 3 depicts the absolute upper values for *A*, in respect to some communication systems applied.

For simplicity reasons, we do not analyze in detail the formulae used when calculating the combination impairment factor (I_s), the delay impairment factor (I_d), and the equipment impairment factor (I_e). However, having read the detailed analysis of the R_o metric above, the reader can easily find similar explanation for the rest metrics in [ITU-T, Rec. G.107,, 2005].

5.1.2 Results

In our non-intrusive objective method used for evaluation purposes, we performed in parallel, true real-time jitter measurements. The main advantages of a true real-time jitter measurement are two-fold. On one hand, there is no dependence on packets needing to be sent at a known interval, while on the other hand the method can measure jitter on a bursty traffic like the one we expected to have in our application. Also, this method does not restrict test duration. This is because the calculation occurs in real time as packets are received, with no need of packet capture. Finally, this method compensates for lost and out-of-sequence packets, producing results in real-time for instant feedback even when traffic load or device parameters vary [Spirent Communications, 2007].

For real time acquisition of the parameters and metrics needed, we use CommView, which is a tool capable for real-time monitoring in Internet and Local Area Networks (LANs) as well as in analyzing activity of captured data network packets (http://www.tamos.com/products/commview/).

Table 4 depicts the averaged results regarding a set of evaluation experiments conducted for measuring the call quality of TELS. The evaluation procedure was performed as follows. We used 15 individuals (students), separated in three groups, namely A, B and C, each consisted of 5 TELS users. Group A and C consisted of VoIP users, while Group B consisted of telephone (either PSTN or GSM) users. During the first sets of experiments, we recorded the averaged values of the R-Factor, the available bandwidth of the voice server, as well as the jitter between the two communication parts, when a single VoIP session was performed from each user. We then performed the same voice sessions five additional times, by generating traffic in the voice server through protocols UDP, TCP and ICMP. This was made in order to assess the performance of TELS under different network stressing conditions. The size of the generated packets was 42, 54 and 106 Bytes for UDP, TCP and ICMP protocols respectively, while their generation rate was 20 packets/sec.

The second set of evaluation experiments was made under the same concept, but subject to two simultaneous calls from users that used different context when calling TELS (either a PSTN and VoIP call or GSM and VoIP call – Group B and A respectively). However, for measuring the voice quality the users received through telephone calls, we asked them to use the 5-points MOS scale for their evaluation. The experiments ended with the evaluation made for two simultaneously VoIP calls, from users of Group A and Group C. Thus, the total amount of separate experiments conducted for voice quality evaluation over different context and networking conditions were 90 (30 for single VoIP calls, 30 for two simultaneous calls of different context and 30 for two simultaneous VoIP calls).

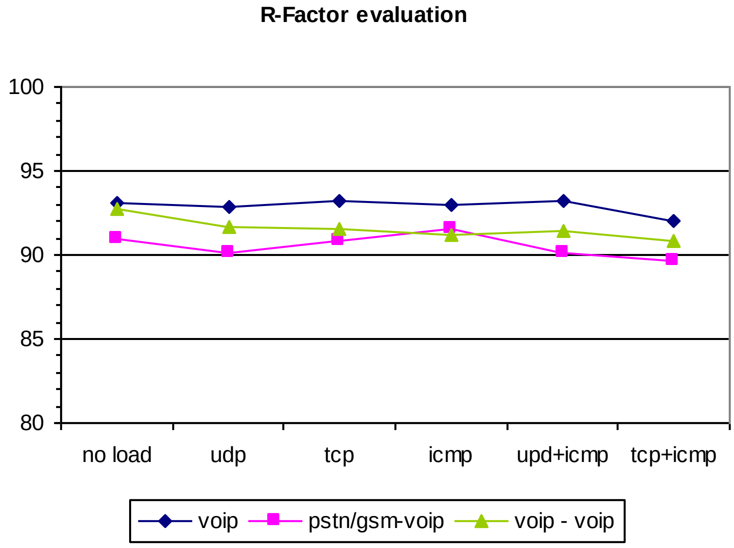


Figure 4. R-Factor evaluation over call context and traffic load

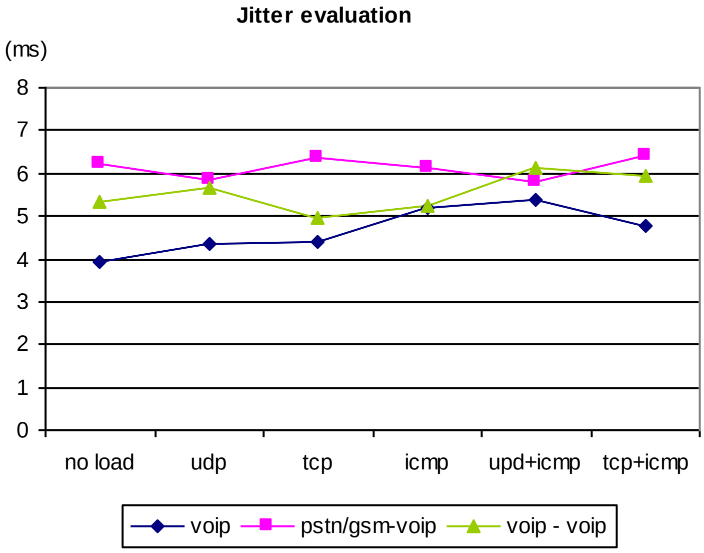


Figure 5. Jitter evaluation over call context and traffic load

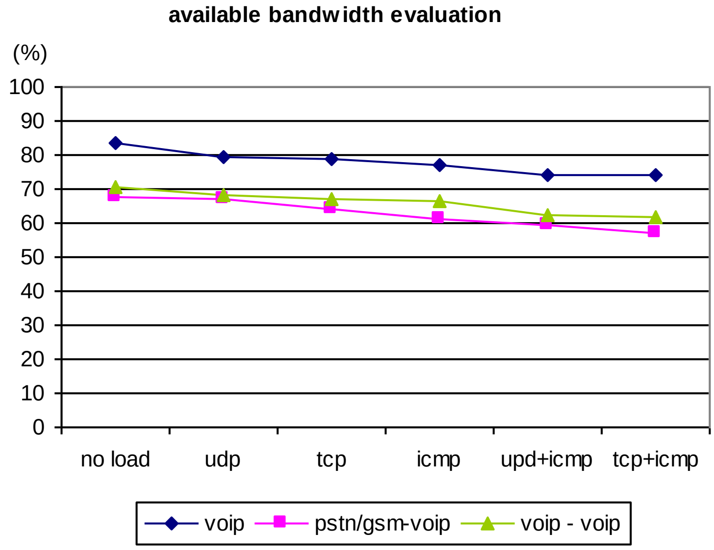


Figure 6. Jitter evaluation over call context and traffic load

Every value depicted in Table 4 is an averaged value from five different experiments. Based on these values, Figures 4, 5 and 6 illustrate the averaged values for all series of experiments explained above. It was evaluated that R-Factor is not significantly influenced by traffic or stressing network conditions, but it is more sensitive when different context calls are performed simultaneously (Figure 4 – green/purple curves). However, even in such a case, the voice quality reduction is nearly negligible (close to 0.1 or 0.2 in the 5-points MOS scale). Now, as far as jitter is concerned, Figure 5 clearly shows that this metric is totally correlated with the amount of simultaneous calls and/or calls of different context, as its averaged values considerably increases in all cases. Finally, Figure 6 shows that the percentage of available bandwidth presents a similar behavior in respect to the R-Factor trends. Nevertheless, the slope in the R-Factor reduction is notably lower in comparison to the one of the available bandwidth. It was assessed that even if the available bandwidth fall down by nearly 10% in all three separate group experiments (83.59 to 73.96, 67.92 to 57.16, and 70.37 to 61.60 respectively), R-Factor averaged values were dropped only by 2 units at the most (Table 4 and Figures 4, 6). This was very encouraging, indicating that TELS is quite robust when it comes to provide high quality vocal services. Finally, Figures 7 and 8 illustrate the influence of the available bandwidth and jitter in the R-Factor in all tested cases depicted in Table 4.

The curves in these figures are the moving averages of the measurements taken in respect to the three separate case studies described above (single VoIP calls, two simultaneously calls of different context and two simultaneously VoIP calls).

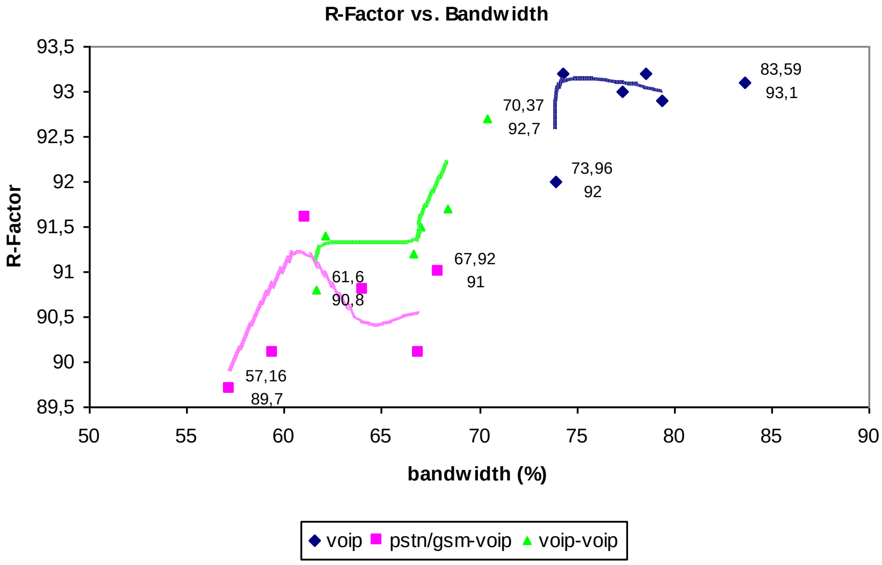


Figure 7. The influence of available bandwidth in the R-Factor

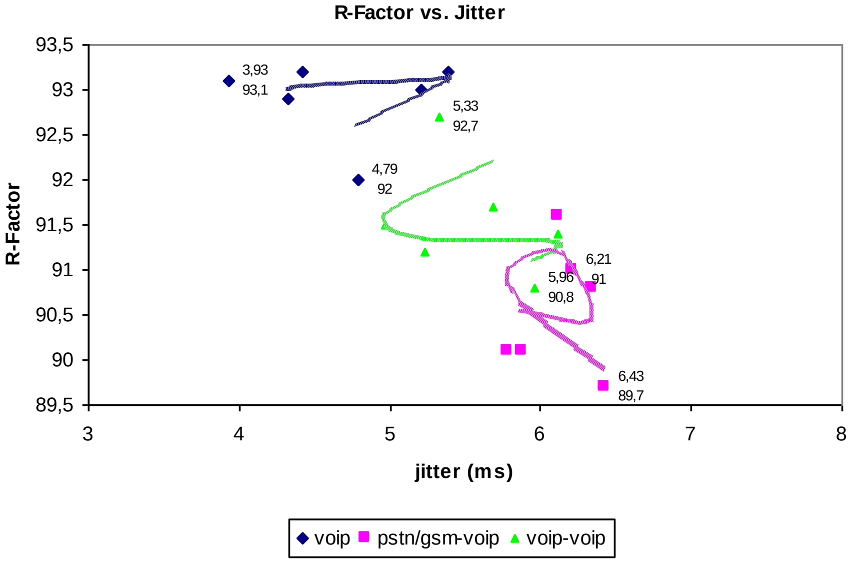


Figure 8. The influence of jitter in the R-Factor

5.2. Evaluating TELS from a user's perspective

5.2.1. Methodology

The data used in this pilot study were gathered through a survey among NTUA postgraduate and undergraduate students and instructors involved in Medialab's courses. A questionnaire consisting of ten questions was designed to collect users' experiences and perceptions about TELS. Specifically, the main objectives of the present study are: (a) to identify if the users are satisfied when using the system and to what degree, (b) to collect users' opinions on improving certain functionalities of TELS. To do so we use positive worded statements in nine and thirteen-item Likert-type corresponding questions. Each statement has five alternatives to choose from: strongly disagree, disagree, neither agree nor disagree, agree, and strongly agree.

The second part was targeted to collect some demographics data about the subjects. That is, age, gender the subject's experience with computers and the Internet, etc. The questionnaire can be found in the Appendix.

The content validity of the instrument was reviewed by a panel of experts in e/m-learning projects. The calculated coefficient alpha reliability from the results of this survey instrument was 0.902, which suggests that this instrument is suitable for measuring the users' perceptions about TELS. The participants of this pilot study were assured that their responses would be anonymous and confidential and that no personal data were recorded during the overall questionnaire submission process. We collected 20 cases in total. The outcomes of this pilot study will provide us with the necessary feedback to improve TELS before it can be fully deployed and help us to refine the survey instrument. A large-scale and more detailed survey, considering a large number of participants, is scheduled in the near future. This is possible because several users are expected to use TELS for at least 2 semesters in their e-learning classes and thus form a well-rounded opinion about its features and functionality.

5.2.2. Results

The subjects were 16 males (80% of the total sample) and 4 females (20%). Their ages varied in four categories. The first category included 14 subjects (70%). The second category included 4 subjects (20%), who were 31-40 years old. The third and fourth categories contained 1 subject each (5%) who was 41-50 and >50 years old correspondingly. The students reported four different levels in relation to the application and its associated tools usage degree (Q_1). The first category contained 3 subjects (15%) who reported some use of the tool. The third and fourth categories came up with 5 items (25%) and 4 (20%) correspondingly reporting moderate use. The final category contained 8 subjects reporting frequent use. The Mean and Standard Deviation (SD) values for Q_1 were 2.85 and 1.13 respectively. According to $Q_{2,3}$ each subject had connected to TELS using a fixed line nearly 4 times in average, 1 time using his mobile phone, and 5.5 times in average using a softphone. The average time duration of connections was 2.3 min. The subject's knowledge and experience with PCs and the Internet ($Q_{9\text{-}11}$) were classified in five different categories. The mean values for these questions were 4.45, 4.40, and 4.00 correspondingly. The results show that subjects are well familiar with IT tools and technologies in general.

The core of our study is questions 4 and 5. The first one identifies if the users are satisfied when using the tool and to what degree, while the latter explores users' opinions on improving some of TELS functionalities. The Mean and St. Dev. Values for these questions are given in Table 5.

In a nutshell, the subjects report that they are satisfied or very satisfied with $Q4_{1,2,3,4,7,8,9}$ statements and neutral for the rest. For instance, 60% of the subjects agree that the voice instructions of TELS are very helpful when navigating around its topics. Also, 50% of the subjects agree that navigating TELS is always easy. On the other hand, 60% of the subjects are neutral about the quality of TELS topics, and a 25% strongly agree that the topics of TELS should be enriched. A 70% of the respondents agree that TELS respond always fast to their queries, and a 95% agrees (45%) or absolutely agrees (40%) that no interruptions occurred during service acquisition. A 55% of the subjects are generally satisfied (50%) or absolutely satisfied (5%) when using TELS. The rest of the respondents remain neutral to this last question. Also, a 95% of the subjects would use (45%) or may use (45%) TELS in the future.

Considering Q_5, the users report that the tool could be further improved in several different aspects. For instance, 65% of the subjects agree that the enrichment of the wiki engine with additional and more detailed voice instructions would be helpful. Also, a 70% of the subjects agree (35%) or absolutely agree (35%) that TELS' voice commands could be further amplified to improve navigation around topics. A 50% or the correspondents absolutely agrees that TELS instructions should be localized to be fully understandable by native users. On the other hand, a 35% of the subjects agree (25%) or absolutely agree (5%) that TELS response speed to voice commands should be improved. This is in accordance to $Q4_7$ as well. Generally, users' opinions are cross-verified when comparing Q4 with Q5. For instance, $Q4_2$ with $Q5_1$, $Q4_5$ with $Q5_5$, and $Q4_4$ with $Q5_8$ to mention just a few.

6. Conclusions

In this paper, a novel e/m-learning system that can be accessed by any wired or wireless phone as well as by a common web browser was presented. TELS, as an aggregate of proven useful tools for collaboration in learning, such as wiki and forum, was made accessible from any communication device, practically everywhere. Speech, the most natural and intuitive communication modality combined with telephone, the most penetrated communication mean can bring learning closer to a wider range of people who do not own or are not comfortable with the use of a PC. At the same time TELS can be useful to the IT literate people, due to its directness and availability. The advantages of the proposed implementation over similar existing approaches are: (a) it is based solely on speech, therefore no training is needed, (b) it does not require installation of special software or a PC in order to be used and (c) it is cost efficient for the end user.

We provided a detailed description of TELS's prototype components, discussing their functionality and analyzing their aspects. Several quantitative results clearly depict TELS applicability to educational realms. TELS can adapt and potentially cover on-demand any informational needs of institutions, which means that it can be custom-tailored to support and further improve modern distance learning services.

References

Agarwal S., Chakraborty D., Kumar A., Nanavati A. A., & Rajput N., (2007) HSTP: Hyperspeech Transfer Protocol. In Hypertext 2007 Proceedings of the ACM Conference on Hypertext and Hypermedia. UK.

Borodin Y., Mahmud J., Ramakrishman I.V. & Stent A., (2007) The HearSay Non-Visual Web Browser, ACM International Conference Proceeding Series, Proceedings of the 2007 international cross-disciplinary conference on Web accessibility (W4A), Vol. 225, pp. 128-129, Banff, Canada.

CMU Sphinx Group, Open Source Speech Recognition Engines, Retrieved on December 30, 2008 from http://cmusphinx.sourceforge.net/html/cmusphinx.php

DAISY Consortium. Retrieved on December 30, 2008 from http://www.daisy.org/

Dehler, C. & Parras-Hernandez, L.H. (1998) Using Computer-Mediated Communication (CMC) to Promote Experiential Learning in Graduate Studies. Educational Technology, 38, 3, 52-55.

Ding L., (2009) Learn about VoIP quality measurements, white paper, Retrieved on May 4 from www.embeddeddesignindia.co.in/STATIC/PDF/200903, EE Times-India.

Focus Bari Marketing Research Services, (2008) Greeks and Mobile Telephony, InfoCom World Conference 2008, retrieved on May 4, 2009 from http://www.focus.gr/includes/download2.asp?file=InfoComWorld_2008.pdf&size=433847 (in Greek).

Ford, M & Botha, A. (in press): (2007) MobilED - An Accessible Mobile Learning Platform for Africa?, IST Africa 2007 Conference. Maputo, Mozambique.

International Telecommunications Union, (2007) World Telecommunication/ICT Indicators Database 2007, Mobile cellular subscribers.

ITU-T, Rec. G.107, Series G: Transmission Systems And Media, Digital Systems and Networks: International telephone connections and circuits – General definitions, The E-Model, a computational model for use in transmission planning, March 2005.

ITU-T, Rec. P.563, ITU-T: Single-ended method for objective speech quality assessment in narrowband telephony applications. International Telecommunications Union, Geneva, Switzerland. (2005) ITU-T Recommendation P.563.

ITU-T, Rec. P.833, ITU-T: Methodology for Derivation of Equipment Impairment Factors From Subjective Listening-Only Tests. International Telecommunications Union, Geneva, Switzerland. (2001) ITU-T Recommendation P.833.

ITU-T, Rec. P.862, ITU-T: Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs. International Telecommunications Union, Geneva, Switzerland. (2001) ITU-T Recommendation P.862.

Kumar A., Rajput N., Chakraborty D., Agarwal S. K. & Nanavati A. A., (2007) WWTW : The World Wide Telecom Web, Workshop on Networked Systems for Developing Regions.

Kumar A., Rajput N., Chakraborty D., Agarwal S., & Nanavati A. A., (2007) Voiserv: Creation and delivery of converged services through voice for emerging economies. In WoWMoM'07 Proceedings of the 2007 International Symposium on a World of Wireless, Mobile and Multimedia Networks, Finland.

Miniwatts Marketing Group (2009), World Internet Usage, retrieved on May 4, 2009 from http://www.internetworldstats.com/europa.htm#gr

MobileEd Initiative (2009), retrieved on May 4, 2009 from http://mobiled.uiah.fi/

Raja A., Atif Azad R.M., Flanagan C., and Ryan C., Real-Time, Non-intrusive Evaluation of VoIP, M. Ebner et al. (Eds.): EuroGP 2007, LNCS 4445, pp. 217–228, Springer-Verlag Berlin Heidelberg, 2007.

Reinhold, S. (2006). WikiTrails: augmenting Wiki structure for collaborative, interdisciplinary learning. In *Proceedings of the 2006 international Symposium on Wikis* (Odense, Denmark, August 21 - 23, 2006). WikiSym '06. ACM, New York, NY, 47-58. DOI= http://doi.acm.org/10.1145/1149453.1149467

Ruberg, L.F., Moore, D.M. & Taylor, C.D. (1996) Student Participation. Interaction, and Regulation in a Computer-Mediated Communication Environment: a Qualitative Study. Journal of Educational Computer Research, 14, 3, 243-268.

Session Initiation Protocol RFC 3261, retrieved on September 4 2009 from http://www.ietf.org/rfc/rfc3261.txt

Spinczyk D. & Brzoza P. (2008) Multimedia System for Accessible Distant Education. Information Technologies in Biomedicine: 513-517.

Spirent Communications, (2007) Measuring Jitter Accurately, white paper retrieved on May 4 2009 from www.spirent.com/documents/4814.pdf

Wang, L., Roe, P., Pham, B. & Tjondronegoro, D. (2008). An audio wiki supporting mobile collaboration. In Proceedings of the 2008 ACM Symposium on Applied Computing (Fortaleza, Ceara, Brazil, March 16 - 20, 2008). SAC '08. ACM, New York, NY, 1889-1896.

Warschauer, M. (1997) Computer-Mediated Collaborative Learning: Theory and Practice. Modern Language Journal, 81, iv, 470-481.

Werner, S.; Wolff, M.; Eichner, M. & Hoffmann, R., (2004) "Integrating speech enabled services in a Web-based e-learning environment," Information Technology: Coding and Computing, 2004. Proceedings. ITCC 2004. International Conference on , vol.2, no., pp. 303-307 Vol.2, 5-7.

World Wide Web Consortium (W3C), Speech Recognition Grammar Specification, retrieved on May 4 2009 from http://www.w3.org/TR/speech-grammar/

World Wide Web Consortium (W3C), Speech Synthesis Markup Language, retrieved on May 4 2009 from http://www.w3.org/TR/speech-synthesis/

World Wide Web Consortium (W3C), VoiceXML, retrieved on May 4 2009 from http://www.w3.org/TR/voicexml20/

Zeadally, S., Siddiqui, F., Kubher, P., (2004) Voice over IP in intranet and internet environments. IEE Proceedings: Communications 151 (3), pp. 263-269.

Zeadally, S., Siddiqui, F., (2004) Design and implementation of a SIP-based VoIP architecture, International Conference on Advanced Information Networking and Application (AINA) 2, pp. 187-190.

Tables

Table 1. Technical Requirements for a TELS implementation

No	Requirement	Description
1	C# 2.0	Used to implement the TELS engine
2	SQL Server 2008	Used to implement database functionality
3	ASP.NET 2.0	Used to generate XHTML and VoiceXML pages
4	IIS 5.1	The server needed for TELS website to work
5	Microsoft Office Communications Server 2007 Speech Server	Used to implement the speech server functionality
6	Dialogic DMG 1004 Media Gateway	Used as a VoIP gateway

Table 2. A typical Voice Session with a wiki article voice annotation addition

Entity	Dialog
TELS	*Welcome to the TELS audio learning system. In order to login, please provide one by one the digits of your unique id, and then say finish.*
User	*1*
TELS	*One*
User	*..*
TELS	*.*
User	*Finish*
TELS	*Now, please provide the digits of your password.*
User	*9*
TELS	*Nine*
User	*Finish*
TELS	*Welcome, George.* *You have 2 new posts on your forum topics.* *1 of your essays is approaching its deadline.* *You can navigate to one of the following sections.* *E-lecture, Wiki, Forum.*
User	*Wiki*
TELS	*Would you like to provide an article title or a search term?*
User	*Article title*
TELS	*Please provide the article title*
User	*Local area network*
TELS	*You have selected to hear the article Local Area Networks* *Table of Contents: …* *Please say how to proceed.*
User	*Begin*
TELS	*Definition. A Local Area Network is a set of connected computers…*
User	*Next*
TELS	*History. The first local area network was developed…*
User	*Comment.*
TELS	*Add your comment after the sound.*
User	*The first network interface card was created…*
TELS	*Annotation added.*
User	*Exit*

Table 3. Provisional examples for the Advantage Factor *A* (as appear in [ITU-TRec, 2005])

Communication system example	***Maximum value of A***
Conventional (wirebound)	0
Mobility by cellular networks in a building	5
Mobility in a geographical area or moving in a vehicle	10
Access to hard-to-reach locations, e.g., via multi-hop satellite connections	20

Table 4. Averaged call quality evaluation results for TELS

Call context	**single VoIP call**			**two simultaneous calls** **(pstn/gsm – VoIP)**			**two simultaneously VoIP calls**		
Evaluation Parameters	**R-Factor**	**Bandwidth (%)**	**Jitter (ms)**	**(MOS) / R-Factor**	**Bandwidth (%)**	**Jitter (ms)**	**R-Factor**	**Bandwidth (%)**	**Jitter (ms)**
No load	93.1	83.59	3.93	(4.3) / 91.0	67.92	6.21	92.7	70.37	5.33
udp	92.9	79.36	4.33	(4.1) / 90.1	66.87	5.87	91.7	68.42	5.68
tcp	93.2	78.57	4.42	(4.0) / 90.8	64.03	6.34	91.5	67.01	4.97
icmp	93.0	77.32	5.21	(4.1) / 91.6	61.08	6.11	91.2	66.59	5.23
Upd+icmp	93.2	74.31	5.39	(4.0) / 90.1	59.44	5.78	91.4	62.14	6.12
Tcp+icmp	92.0	73.96	4.79	(3.9) / 89.7	57.16	6.43	90.8	61.60	5.96

Table 5. Descriptive statistics for survey questions Q4 and Q5

Statement	**Mean**	**SD**
(Q_{4_1}) Connection to TELS is always easy	4.25	.639
(Q_{4_2}) TELS voice instructions are clear and helpful	4.00	.649
(Q_{4_3}) Navigation to contents is straightforward	3.80	.696
(Q_{4_4}) TELS is usually able to understand user voice commands	3.65	.813
(Q_{4_5}) Variety of TELS subjects/topics is satisfactory	2.80	.951
(Q_{4_6}) Quality of TELS subjects/topics is satisfactory	2.80	.834
(Q_{4_7}) TELS responses are immediate (speed)	3.95	.686
(Q_{4_8}) Connections to TELS are never interrupted	4.25	.716
(Q_{4_9}) Calls to TELS are always sharp (sharpness)	4.25	.639
(Q_{4_10}) Browsing TELS topics is easy	2.85	1.137
(Q_{4_11}) The variety of TELS voice commands is satisfactory	3.35	.671
(Q_{4_12}) Overall user satisfaction	3.60	.598
(Q_{4_13}) Future use	3.50	.889
(Q_{5_1}) Enrichment of TELS voice commands	4.15	.587
(Q_{5_2}) Specialization and fine-tuning of TELS commands	4.00	.918
(Q_{5_3}) Localization[*] of TELS voice instructions	4.35	.813
(Q_{5_4}) Localization of TELS contents	4.25	.851
(Q_{5_5}) Enrichment[**] of TELS contents to include more topics	4.00	.795
(Q_{5_6}) Improve the quality of TELS contents	3.80	.834
(Q_{5_7}) Improve TELS response speed	3.05	.887
(Q_{5_8}) Improve TELS voice recognition system	3.40	1.046
(Q_{5_9}) General improvement of TELS	3.85	.875

[*] Localization: the translation of TELS voice commands and contents to the local (Greek) language.

[**] Enrichment: to enrich (improve, populate) TELS contents to include new and more detailed topics.

Appendix

SURVEY
«TELS LEARNING SYSTEM»

The following survey seeks your opinion concerning the quality of service offered by the TELS tool. Your participation is entirely voluntary and your responses will remain confidential.

1. How many times did you make use of the services of TELS?

□ 1 – 2 times □ 3 – 4 times □ 5 – 6 times □ more than 6

2. How many times did you connect to TELS using the following (fill in):

A landline phone: ___
A mobile phone: ___
A VoIP softphone (Internet): ___

3. For how many minutes in average did you stay connected (approximately)?

□ up to 5' □ up to 10' □ up to 20' □ more than 20'

4. Mark the corresponding block that represents the level of your satisfaction from the services of TELS.

(1 = strongly disagree, 2 = disagree, 3 = neither agree, nor agree, 4 = Agree, 5 = strongly agree)

		1	2	3	4	5
1.	Connecting to TELS is always easy	□	□	□	□	□
2.	The voice instructions provided are clear and helpful to the navigation	□	□	□	□	□
3.	The navigation using voice commands is always easy	□	□	□	□	□
4.	Your voice commands are usually understood	□	□	□	□	□
5.	The variety of the content is satisfactory	□	□	□	□	□
6.	The quality of the content is satisfactory	□	□	□	□	□
7.	The response of TELS to your voice commands is always immediate, without notable delay	□	□	□	□	□
8.	The connection was never interrupted during a call	□	□	□	□	□
9.	The communication is always "sharp" (without noise and interrupts)	□	□	□	□	□
10.	You always locate easily the topic you are looking for	□	□	□	□	□
11.	The variety of voice commands that TELS provides is satisfactory	□	□	□	□	□
12.	In general, you are satisfied from TELS services	□	□	□	□	□
13.	In the future you will continue using TELS	□	□	□	□	□

5. Mark the corresponding block that represents your opinion about the ways that the services of TELS could improve

(1 = strongly disagree, 2 = disagree, 3 = neither agree, nor agree, 4 = Agree, 5 = strongly agree)

		1	2	3	4	5
1.	Provide a larger amount and more specialized voice instructions to the user	□	□	□	□	□
2.	Provide a larger amount and more specialized voice commands to the user	□	□	□	□	□
3.	Localization of the instructions and voice commands	□	□	□	□	□
4.	Localization of TELS contents	□	□	□	□	□
5.	Enrichment of TELS contents	□	□	□	□	□
6.	Improvement of the quality of TELS content	□	□	□	□	□
7.	Improvement of TELS response speed to user’s commands	□	□	□	□	□
8.	Improvement of voice recognition regarding the navigation commands	□	□	□	□	□
9.	General improvement of TELS environment so that it is easier to use and more functional	□	□	□	□	□

Other comments (write):

____

____

____

____

Demographics

6. Your gender

□ Male □ Female

7. Your age

□ 25-30 □ 31-40 □ 41-50 □ 51-60 □ 60+

8. Do you have access to computers with Internet connection?

□ No □ Little □ Some □ A lot □ Very much

9. Do you use computers with Internet access?

□ No □ Little □ Some □ A lot □ Very much

10. Your knowledge and experience with the use of computers could be characterized as:

□ Inadequate □ Little □ Moderate □ Good □ Excellent