

Improved Text Reading System for Digital Open Universities

Mahamadou ISSOUFOU TIADO

Department of Mathematics and
Computer Science,
University of Abdou Moumouni,
UAM
Niamey, Niger

Abdou IDRISSE

University Institute of Technology
University of Tahoua
Tahoua, Niger

Karimou DJIBO

Department of Mathematics and
Computer Science,
University of Abdou Moumouni,
UAM
Niamey, Niger

Abstract—The New Generation of Digital Open Universities (DOUNG) is a recently proposed model using m-learning and cloud computing option and based on an integrated architecture built with open networks as GSM and Internet. The goal of achieving the ubiquitous ability of the m-learning is having the large number of languages as a serious issue. It needs to use many teachers in order to repeat the same course in various languages. In this paper, an extended system is proposed under the consideration of the low capacities of the cell-phone device in terms of computing and visualization. The model uses the possibility to build a voice warehouse which can be used to generate the audio format of every course provide in a text format and in a particular language. The Advanced Text Reading System (ATRS) is proposed to use that voice warehouse and to produce the audio format of a course, giving facility to teachers to easily overcome the constraints of language barrier. The new proposed model is described and its contributions are discussed.

Keywords—m-learning; distance learning; digital open universities; cloud-computing; audio warehouse

I. INTRODUCTION

The advent of the New Generation of Digital Open Universities was previously proposed to achieve the goal of realizing the conversion of the traditional distance learning to digital solutions with multiple options. Some of the DOUNG options allow to record the sessions given by a teacher in a multimedia format or to convey the multimedia stream instantly through the network towards the learners. In this operating model, the saved multimedia files contain a particular course in a specific language. An incoming limitation is related to the need to deliver the same course in another language. The nowadays solution uses another session given by a teacher chosen according to the targeted language, with also another work of recording to do. That operating model becomes heavier and generates more issues when addressing large public of learners around the world according to the ubiquitous nature of the m-learning. Thus, the learner's language becomes a barrier for the teacher and another parameter for recording new sessions. The goal of providing courses to a wide public of learners around the world in as many languages as possible brings to the proposition of the audio warehouse and its exploitation script. They are used to facilitate the conversion of the courses in audio format within a selected language. In addition, this new operating model that

generates audio format of courses meets the constraints of the extension of the DOUNG services over GSM network.

After presenting the basic concepts of the distance education in point II, the hybrid model of the DOUNG is presented in point III. The point IV presents the voice warehouse with its building process. The point V highlights the ATRS algorithm used on the cloud computing architecture to improve the DOUNG offered services. The point VI presents the cloud computing architecture of the DOUNG improved by the audio warehouse and the ATRS.

II. BASIC CONCEPTS OF THE DISTANCE EDUCATION

The distance education is implemented with various techniques such as the use of the e-learning by the means of computers network. Following the evolution process, the e-learning techniques and the mobile learning [1] bring to the advent of the m-learning technology. The new concept integrates the expansion of the education including mobile learners that use a wireless network without architecture (ad hoc technology) interconnected to the initial Internet backbone. The m-learning uses a wide range of mobile devices including wireless technologies, different protocols and applications. The designed architecture operates under the service-oriented technology philosophy. The increasing number of mobile devices becomes a favorable factor for the implementation and the extension of the m-learning, particularly because of the interconnection between wired and wireless networks, and between telecommunications and computer networks.

In a second side, the technological evolution process brings to the advent of the cloud computing technique that allows to optimize the use of the computers in a network. The cloud computing provides the possibility to put in common the overall storage capacity and processing power of the computers in the network. With this technique, the network becomes a channel that carries program and data from one point to another to ensure remote collaboration between nodes seamlessly to the user. The cloud computing allows to design a configurable architecture of computing resources in the network such as storage servers, applications, and services. Following the success of the cloud computing, the technology evolution process brings to the advent of the cloud learning. The cloud learning is a service offering distance learning based on cloud computing. It uses the availability of network servers offering data that contains sources and training documents in

different formats such as text, multimedia, images. The environment used to design the system is called Cloud Learning Environment (CLE). It is free of an organization and put together learners and trainers.

At the third side, the Virtual Private Network [2][3][4][5] (VPN) has emerged and becomes another technology of distance education used by many centers. The VPN technique uses several remote computers, in the image of a real private Local Area Network (LAN), but communicating by using a public infrastructure. The computers operate by creating a tunnel and imposing the constraints of (1) user authentication, (2) encryption of the data, (3) access keys management and (4) multi-protocol availability. The following protocols are used to ensure the VPN operation. The Point to Point Tunneling Protocol (PPTP) creates a virtual private network by using the Point to Point Protocol (PPP) and the Internet Protocol (IP). The Internet Protocol Security (IPsec) [6] or the Multi-Protocol Label Switching (MPLS) [7] are used at the level 3. The Secure Sockets Layer (SSL) [8] is used at level 4 and ensures the security and the confidentiality of the communication through the network. The SSL protocol uses a socket connection to associate clients to server stations. It can be used by any application for securing the traffic such as Hyper Text Transport Protocol (HTTP). Other applications are used to offer multiple choices to the learners such as File Transfer Protocol (FTP), Telnet [9], Internet Relay Chat (IRC) ...

III. THE HYBRID MODEL OF THE DOUNG

A. The DOUNG architecture

As defined in [10], it is designed to achieve the ubiquitous faculty of the m-learning associated to the capacities offered by the cloud computing. The architecture is based on the hybrid model with set of solutions based on computers and telecommunication networks. The model interconnects teachers and learners within a course produce in a support made available for learner's access. The package 10 in 1 that summarizes the set of solutions gives different format of the link between the teacher and the learner. It includes also the format of the content of the course, the course medium and the methods of the course production. The DOUNG is designed with various method destined to make the course available (broadcast and/or storage). It is designed with the definition of many kinds of learner access (synchronous and/or asynchronous transfer files). The definition of the nature of the channel that can be used completes the description of the DOUNG services. The transmission channel consists of a backbone network connected to the Internet, accessible via Internet and GSM. The course can be subjected to immediate transmission in multimedia stream, or stored as multimedia files, treated text files, untreated text files or pdf files. The operation of the DOUNG is based on the cloud computing with the goal to achieve the balancing of the servers load in the network combines to the need to incorporate the limited capacity of mobile devices.

At the learner side, the required materials consist of a computer with Internet access or a cellular phone. Among many choices offered by the DOUNG, the learner can use multimedia applications (real time and deferred time), or file

transfer mode with file displaying applications according to their format [1].

B. The DOUNG applications

The implementation tools of the DOUNG are previously defined in [10] with hardware, protocols and applications. The wide range of applications is given by the m-learning and is extended by the use of the cloud computing, making helpful their grouping into categories. The applications used to produce the courses are grouped in category (1). The second (2) category contains applications used to make those courses available. The applications allowing the learner to access the DOUNG services are grouped in the third (3) category. The fourth (4) category regroups applications used for the visualization of the courses by the learner and the fifth (5) category concerns the applications of exchange between teacher and learner.

The list of applications used by the DOUNG is extended because of the need to integrate the new proposed voice warehouse. As previously stated, the first list of applications of the category (1) regroups multimedia stream capture and recording applications, text processing applications, spreadsheets and pdf file generator. The category (2) still remains in multimedia recording with instant transmission server and an ftp server. Immediate multimedia viewer and ftp clients are used for the category (3). Deferred visualization multimedia client, word processing application, spreadsheets and pdf viewers are used for the category (4).

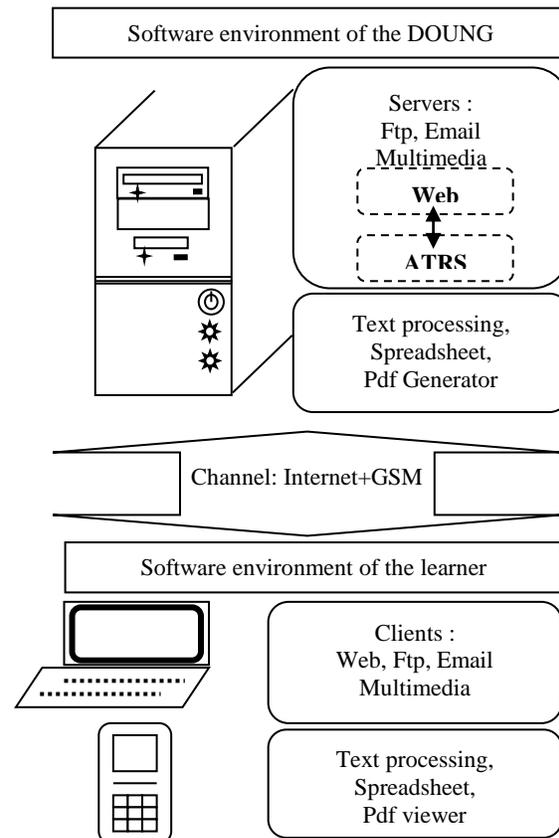


Fig. 1. The improved DOUNG software environment

The category (5) concerning applications used for the exchange between learner and teacher includes multimedia clients, multimedia servers with instantaneous transmission, Internet telephony applications, email clients and servers applications. A web server is still being suitable to be used for its multifunctional operating mode that incorporates web pages containing courses, multimedia content, ftp site, search of archives, ... The following figure 1 shows the improvement of the DOUNG software environment when integrating the ATRS.

IV. THE VOICE WAREHOUSE

A language can be singularly defined as a set of reserved words and a set of rules that explain how to put the language words together. The set of words define a dictionary which is usually available in a text format. Every word used to build sentences that composed a text in a particular language is provided by the dictionary. The voice warehouse is here a kind of dictionary containing the audio format of each word in a given language. A difference that can be stated is that words that have the same pronunciation generate the same audio file, even if their orthography is different.

The idea given as part of the main contribution of this paper is to improve the options offered by the DOUNG to a learner. In the improved scheme, the DOUNG provides in addition to the previous courses format, the same courses in audio format and in different language through the Advanced Text Reading System. The ATRS is implemented as a script that uses the audio warehouse containing the pronunciation of every word given by the dictionary of a selected language. The script is written to follow the words juxtaposition in a text and for each word, it will select the corresponding audio file from the warehouse. After the selection step, the audio file is played by using the loudness parameters defined later to help to take the punctuation into account. The audio flow produced is conveyed through the network (Internet + GSM) to the cell-phone of the learner as in the case of the VPN multimedia traffic between Internet and GSM. The audio flow can be also converted in a multimedia file so that a learner having a cell-phone with appropriate option (multimedia client) can download the file and play it locally. This means that another course in multimedia format can be produced by the ATRS without imposing to the teacher the constraints to record again the same course given in the text format.

Such kind of association between a word and its associated audio file can simply be made by given to the audio file, the same character composition of the word regardless to the upper or lower cases of the characters. This brings to make the script more intelligent so that it can resolve the cases of words having different orthography with the same pronunciation. A faire organization can be made in the storage of the sound in the warehouse. The audio dictionary of a language will be put into a separate directory different to the other language one. This may help to facilitate the language word selection while the language itself becomes an input parameter of the ATRS script. It can be assumed that if the text is well written in a particular language, by following the language grammar and its orthography constraints, the ATRS will also produce the correct reading of that text.

The precedent list of applications of category (1) is extended to integrate the ATRS script. The performance of the production of the sound by the ATRS is subjected to the capacity of the computer to run fast that conversion program. Another solution can be implemented used by multimedia applications. It concerns the use of the anticipation window that allows to early provide new sounds by accessing to the warehouse when the precedent audio words are being played. Thus by using the cloud computing mechanism, the task will remain on the server of the GSM provider. And then, the flow of sound provided by the ATRS will be conveyed to the cell-phone through GSM channel as in the case of the transmission of the normal audio emission. The cell-phone will receive the text, but also the sound, and the learner can choose freely to replay the sound unlimitedly, with more connection benefit in the case of a downloaded multimedia file.

Three processes are used to build the entire system. The first process is related to the building of the voice warehouse, the second process allows to implement the ATRS algorithm, and then the third process includes the running of the ATRS script by the learner.

The process of building the voice warehouse as the first step is conduct by recording an audio file for each word in a selected language. The file represents a word and is put in the warehouse. The warehouse is subdivided into folders with a separate folder for each language. A folder contains all the necessary voice files of a language that are previously recorded. Thus the whole audio warehouse becomes an audio dictionary.

V. THE ADVANCED TEXT READING SYSTEM

The encoding of the ATRS script is realized by following some steps in order to facilitate its evolution through new versions. The first step allows the script to take a language as an input parameter. It sets the language folder in the warehouse as default and current folder to avoid the searching of the file path during every warehouse access. The second step of the encoding of the ATRS system brings to write the script that reads the text by following the words sequence inside the course provides in text format. For the next step, the script realizes the conversion of the extracted word into its audio format by accessing to the appropriate part of the warehouse and selecting the audio file corresponding to the word. The encoding of the ATRS script is realized within the text reading algorithm with taking care of punctuation and performing another algorithm for reading numbers. The running of the ATRS script by a learner is managed by the use of the web server.

A. *The text reading algorithm*

Let assume that the course is produced in the web page format and the ATRS system has to skim the text with the word recognition philosophy by using separators. It is useful to integrate the difference between the word of the course and the reserved words of the Hyper Text Markup Language (HTML) so that, the HTML reserved words must not be pronounced by the ATRS script. The algorithm will use the beacon format in the HTML language to distinguish the reserved words of the language and the course words. The text will be read character

by character. Between HTML beacon characters starting with "<" and ending with ">", the words are ignored. Outside the range of the HTML beacon containing eventually attributes and delimited by the precedent characters, each character not includes in the list of separator is put in a string. The string is built by considering the characters between two separators. The extracted string triggered the loading and the running of the corresponding audio file from the voice warehouse. This is referenced here by the term "resolving the word sound". The text reading algorithm can be written in a low level language as the C language used in the development of the Linux version and the other open sources.

B. Integrating the punctuation in the ATRS

When reading the text, taken care of the punctuation offers the guarantee to make the reading understandable. We will consider two kinds of influence of the punctuation. The first influence is related to the break time before continuing the reading; the second one is related to the sound volume at the end of sentences. When reading the text, the ATRS will use the Normal Inter Words (NIW) interval time that allows the listener to distinguish the "space" separator between words. The speed level of reading can be set according to a value evolving inside an interval given by [LS, HS], where LS is the allowed Low Speed value and HS the allowed High Speed value. These parameters are paramount important for the regulation of the speed of the words resolution in the warehouse. They must be chosen according to the minimum time that can take the system to resolve the word in the warehouse, even if an anticipation window can be used at the convenience of the implementation. When encountering a coma (","), the ATRS will use the Short Inter Words (SIW) interval time. When encountering a full stop character ("."), the algorithm will use the Long Inter Words (LIW) interval time. When playing the pronunciation of the words extracted from a text, the algorithm will use a Normal Loudness Level (NLL). But the pronunciation of the word that comes before the full stop will use a Decrease Loudness Level (DLL) from the NLL level to the Limit of Hearing Level (LHL). The association of the NLL and the LIW at the end of a sentence can give the pronunciation of the question mark "?". Thus, LS, HS, NIW, SIW, LIW, NLL, DLL, LHL are classified in the range of the algorithm parameters that can be adjusted to bring an improved listening quality.

C. The left-to-right reading numbers algorithm

A particular algorithm must be used to read numbers differently to the reading of strings. A string is delimited by separators and its audio format is in a single file. A number is also a string delimited by separators but need to be read digit by digit and by mixing the ten powers showing the position of the digit inside the number. When reading a number, also sub numbers that are in the language base must be detected. Every language has its own base containing digits (single numbers) and sub numbers composed with digits and having their own name. The ten powers are particular sub numbers that are in the base and are usually used to determine the position of the precedent number in the sequence forming one string number. For example, 2345 can be read as follow: two thousand three hundred forty five. Thousand and hundred are the ten powers and two, three and five are digits, then forty is a sub number.

For recognizing a sub number when reading a string number, the script must take into account the current digit and the number of the remaining digits that follows it. If the system crosses a digit that begins a basic sub number, and if the number of symbols (digits) that follows the current digit allows to recognize the sub number, the system will get the other next digits before resolving the word sound. The number of symbols that follows the current digit is also used to determine the time at which the correspondent ten power must be used. For example, for "23", the system recognizes first the digit "2" and then detects that the number of the remaining digits allows to recognize a sub number. The script waits to the get the next digit before resolving the word sound. The algorithm will resolve "20" before "3". It will pronounce after getting the digit "3": twenty three. But in the case of the number "230", after recognizing the digit "2", the number of the remaining digits can't allow to recognize a sub number. Then the system will resolve "2" with using the ten power "100" and adding "30". It will pronounce: two hundred thirty.

A string that consists of characters and numbers is assumed to be a string and can be read in two ways: (1) by spelling the characters that compose the string or (2) by pronouncing the substring formed by the characters and spelling the integrated number digit by digit. For example, the string "MAN123" can be read "M-A-N-one-two-three" or "MAN-one-two-three".

The need to integrate the reading of numbers brings to the completion of the warehouse by including digit sound files, ten powers sound files, and by extension, all the sub numbers that are in the numeric base and that have their own name. For example for the English language, the base consists of the digits from zero to nine, the sub number from eleven to twenty, and thirty, forty, fifty, sixty, seventy, eighty, ninety, hundred, thousand, million, billion.

D. Reading extra language words

In a text, all the words are not necessary given by the language dictionary because of names of objects or technical names that can be used. Two kinds of solution can be implemented when during the resolution of the word in the audio warehouse the ATRS doesn't find the correspondent audio file. The first solution useful for the words in capital letter is to spell the word character by character. This means that the resolution is done for every character in the audio warehouse; not for the entire word. For example, "HTML" will be pronounced "H-T-M-L". The second solution is to put the text under test after producing the course so that the script can early detect the additional words to be put in the language dictionary and thus, in the audio warehouse. This will let the dictionary becoming richer. Anyway, it is assumed that after producing a course, the author must perform the ATRS script and listen to the reading of the text to detect anomalies. It is also possible for the author to adjust the text reading parameters to make his message understandable at his convenience. The access of the text reading parameters given as input to the ATRS script can also be let at the convenience of the learner.

E. Integrating new other language subtleties

Dependent to the robustness of the ATRS, the DOUNG can offer many courses in many languages, regardless to the

evolution of the content of the course. This is made possible because of the separation between a course and its content provided in text format in one side, and in the other side the audio warehouse containing audio files of the basic words of a language. The ATRS is used to read any course provide in the language. This contribution is different to the case by which a particular course is recorded in audio format as in the case of the course provide with the multimedia captors and recorded in an audio file. In that case, when the content of the course changes another audio file must be recorded. The audio warehouse offers flexibility to teachers without the constraints to record any course or to record again when the content of the course changes. The ATRS must be designed as an evolving algorithm because every language has its own subtleties. Thus, it is paramount important to let the possibility to make the ATRS system becoming more robust among versions so that it can integrate other particular languages pronunciation of words. The following figure 2 shows the operating mechanism of the ATRS script.

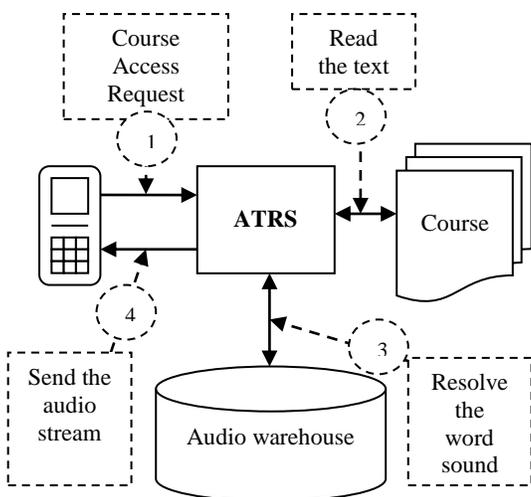


Fig. 2. Interactions between the learner, the ATRS and the courses

VI. IMPROVING THE CLOUD COMPUTING ARCHITECTURE OF THE DOUNG

The DOUNG services are designed to operate on a cloud computing architecture in a completely transparent manner to the learner. The cloud computing is integrated to achieve a minimum level of quality of service when the learner is monitoring the courses. The Software as a Service (SaaS) part of the cloud computing in the DOUNG architecture allows to subdivide the applications in elementary modules freely assembled. A web service is suggested to be used for assembling a course in image format, video format, and fixed text format or by extracting information from a database. For an advanced use of that component of the cloud computing, the ATRS script will accompany a web page by providing the audio format of the transferred text on the learner's device. The same initial advantage is maintained, that is to fit the case of GSM supports more suitable for the audio transmission. This additional ATRS script provides help for the cell devices by offering a solution to the problem inherent to their limited capacity of installing and running applications.

The Data as a Service (DaaS) of the cloud computing used by the DOUNG consists initially of available courses in a warehouse. The warehouse is implemented on different servers to resolve the limited capacity problem of data storage of the cellular devices. The contribution of the ATRS system brings to the improvement of the new version of the DaaS part of the services offered by the DOUNG. The courses warehouse in HTML format (text, image, video) is having as neighbor, the audio warehouse that is used to generate the voice format of courses in different languages. The same transparent location is maintained within a collaborative exchange between the DOUNG servers and all GSM partners.

The Platform as a Service (PaaS) part of the DOUNG services provides a support for processing. In the case of the ATRS, the PaaS is improved by the offered option to the learner to run the script of the translation of a course from its text format to its equivalent in audio format. That audio format can be easily conveyed through the GSM network. By considering the limitation of the text traffic in the GSM network, the audio format is assumed to provide an appropriate framework for the learner to follow his courses. The improved PaaS service by the integration of the remote ATRS system will help to overcome the problem of the limited processing capabilities inherent to mobile devices.

The Infrastructure as a Service (IaaS) is provided in [10] with the interfacing between the web service and other scripts. The example of CGI scripts is given which are written in low level programming languages and which execution extends to files access. The case of ATRS script is illustrative of the implementation of that part of the DOUNG service for the need of reading the text file, the need of recognizing words and selecting the appropriate audio file in the voice warehouse. The illustration is extended to the running of the audio file and its transmission through GSM channel.

VII. CONCLUSION

The e-learning, the m-learning, the VPN, and the cloud computing are techniques used in an integrated architecture using Internet and GSM for the advent of the DOUNG with the goal to cover large world areas. The initial model on which the DOUNG is operating let appears some complementary work of re-recording lessons according to the language of learners or according to some changes that occurs in recorded sessions. To overcome these limitations, the new model integrates the use of an audio warehouse with its ATRS exploitation script. An audio warehouse is an audio dictionary of a chosen language, recorded in a particular folder with a fair organization associating each language to a specific folder. The ATRS script is designed to read a course produced in the HTML format by extracting words, and accessing to the language audio warehouse. It then gives the correct pronunciation of the text. The system conveys the voice stream through the network. That format is more suitable for the GSM, while the cloud-computing gives facility to cell-devices used by to follow a course according to his own budget and time table. An incoming work to be conduct is to integrate many languages constraints in the ATRS. The second work is related to the adding of the language translation module. Then, the

performances of the whole system must be evaluated under simulations with specific criterions.

REFERENCES

- [1] LeCavé, A.D. Salamin, "Mobile Learning : Les avantages du papier virtuel," FI 1, 3 Février 2004, pp.7-9.
- [2] H. Jeng, J. Uttaro, L. Jalil, B. Decraene, Y. Rekhter, R. Aggarwal, "Virtual Hub-and-Spoke in BGP/MPLS VPNs", October 2013, RFC 7024
- [3] Y. Cai, E. Rosen, IJ. Wijnands, "IPv6 Multicast VPN (MVPN) Support Using PIM Control Plane and Selective Provider Multicast Service Interface (S-PMSI)", RFC 6516, February 2012
- [4] B. Davie, F. Le Faucheur, A. Narayanan, " Support for the Resource Reservation Protocol (RSVP) in Layer 3 VPNs ", RFC 6016, October 2010
- [5] E. Rosen, Y. Cai, I. Wijnands, "Cisco Systems' Solution for Multicast in BGP/MPLS IP VPNs", RFC 6037, October 2010
- [6] S. Frankel, S. Krishnan, "IP Security (IPsec) and Internet Key Exchange (IKE) Document Roadmap", RFC 6071, February 2011
- [7] K. Lam, S. Mansfield, E. Gray, "Network Management Requirements for MPLS-based Transport Networks", RFC5951, September 2010
- [8] A. Freier, P. Karlton, P. Kocher, "The Secure Sockets Layer (SSL) Protocol version 3.0", RFC 6101, August 2011
- [9] T. Murphy, P. Rieth, J. Stevens, "IBM's iSeries Telnet Enhancements", RFC 4777, November 2006
- [10] Issoufou Tiado, H. Saliyah-Hassane, "Cloud-Computing based architecture for the advent of a New Generation of Digital Open Universities in m-learning", ICEER 2013 conference Proceedings, pp572-579, Marrakesh – Morocco, July 2013