A Pilot Project – From Illiteracy to Computer Literacy: Teaching and Learning Using Information Technology

Mohamad Adnan Al-Alaoui, Mesrob I. Ohannessian, Ghinwa F. Choueiter, Christine Akl, T. Taline Avakian, Ismail Al-Kamal, Rony Ferzli

A Pilot Project – From Illiteracy to Computer Literacy: Teaching and Learning Using Information Technology

Mohamad Adnan Al-Alaoui[1], Mesrob I. Ohannessian[2], Ghinwa F. Choueiter[2], Christine Akl[1], T. Taline Avakian[3], Ismail Al-Kamal[1], Rony Ferzli[4]

[1]American University of Beirut, [2]Massachusetts Institute of Technology, [3]University of Cincinnati, [4]Arizona State University

Keywords: *assistive technologies for literacy, speech recognition, handwriting recognition*

Abstract:

This paper addresses the use of information and communication technologies, ICT or IT for brevity, to combat illiteracy and move participants directly from illiteracy to computer literacy. The resulting assistive technology and instructional software and hardware can be employed to speed up literacy programs and make them more attractive and effective. The approach provides an interactive, self-paced, autonomous and entertaining learning experience, eases entry and exit in and out of the program, and permits monitoring and updating progress status. The hallmark of the proposed approach is the integration of speech and handwriting recognition, as well as audio and visual aids into the flow.

1 Introduction

Reading, writing and numeracy are at the core of any literacy program. They are indeed critical to function adequately in society. Yet, increasingly, computer literacy is becoming as essential. A literacy program that does not provide computer literacy will not do justice to its recipients. Consequently, this project's fundamental perspective on literacy encompasses all such skills.

We focus on tackling the growing illiteracy rates in the Arab region, one of the highest in the world at about 37% [1], by developing assistive technology for imparting the Arabic language to illiterate adults. Of course, universal elementary education programs are the real long term solution. However, societal resistance to implement such policies and failure to qualify enough teachers and resources are hindering such efforts.

Our philosophy and novelty are to tap directly into information technologies to provide an educational model that would reach more people and move them faster to literacy, and also bridge the 'digital divide' by moving the learner from illiteracy directly to computer literacy, through the pervasive use of technology. As such, despite focusing on Arabic, the benefits of our approach are as globally applicable as illiteracy is globally pervasive.

In this paper, we first present the general framework in which we are working, by showcasing general prior work, and recent research at the ECE department of the American University of Beirut. We then proceed to present a pilot project undertaken in cooperation with the Lebanese National Committee for Literacy (NCL) and the UNESCO. We describe the methodology, and detail the technical elements of the ongoing project. We conclude by highlighting current work and our future vision for the project.

2 Background

2.1 Assistive Technologies for Literacy

The topic of using technological tools for teaching Arabic reveals work dating back to the early 1970's in the research of Victorine Abboud at the University of Texas at Austin [2]. Abboud demonstrated considerable gain in the speed and efficiency of teaching Arabic sounds to non-natives, by pacing lessons to the learner's ability. Since then, research in computer-aided instruction has steadily explored many aspects of using technology in enhancing learning. Of these, the use of speech recognition has shown significant potential, and continues to be an active research area, as surveyed in [3]. An extensive survey of general intelligent systems for tutoring can be found in [4].

Programs developed for the handicapped form another thread of particular relevance, as they can be utilized or adapted to combat illiteracy. Of course, illiteracy itself can be thought of as a handicap. A study that identified and characterized software and hardware solutions for the delivery of literacy programs [5] has pointers to programs for the dyslexic, for the deaf and partially hearing, and for the blind and partially sighted, which were deemed particularly suitable.

2.2 Projects at the ECE Department of AUB

In recent years, several final year (senior) projects and master's theses at the ECE department of the American University of Beirut have addressed topics having elements of assistive technologies. In [9] and [10], basic speech recognition via hidden Markov models and neural networks was tackled. In [11] and [12] basic handwriting recognition applied in literacy software, using touch and pen capabilities, was reported. In [13] and [14], work was done on human-machine interfaces based on motion tracking. In [15] sign language recognition and in [16] and [17] sound visualization were respectively developed for the hearing impaired. In [5], [6] and [7] the groundwork was laid out for the current research project, which we undertook and reported on initially in [8].

3 The TLIT Pilot Project

3.1 Preliminaries

The TLIT (Teaching and Learning using Information Technology) project proposal [7] envisioned an end-user product that incorporates the latest technologies in software design and Arabic speech and handwriting recognition, capable of addressing all requirements to ease the ongoing process of teaching illiterate adults how to read and write, and to also familiarize them with computer technology. Conversely, it was expected from the approach to propagate back, and spawn a family of technological systems that are better at handling tasks needed for autonomous and interactive teaching.

Upon the implementation of the project, our emphasis was on the following:

1. Interactivity: In regular teaching approaches (in classrooms), the learner rarely interacts with the teacher, except for questions. Passive audio- or video- based programs offer even less opportunity for practicing skills and adapting to new information. In a computer-based approach, teaching cannot but rely on interactivity: new skills can be learned but also be tested progressively, and revisited if need be.

2. Self-paced learning: Larger classrooms entail a great deal of compromise when it comes to pacing lectures. Individualized teaching allows self-pacing, at the expense of less discipline. Information technology can balance these aspects not only by providing the convenience and flexibility to the learner, but also by adapting to his/her strengths and weaknesses.
3. Learning through leisure: Computer-based approaches can often be designed in such a way as to engage users by providing what superficially seems to be a game, but actually strengthens intuitions and understanding.

To concretize these criteria we investigated local efforts in combating illiteracy and coordinated our efforts with that of the UNILIT project at the UNESCO in Beirut and the National Committee for Literacy (NCL) at the Lebanese Ministry of Social Affairs. The ongoing programs administered by the NCL were of particular interest. They were deemed a good framework to adopt and enable technologically, as they could ultimately provide a basis for assessment and comparison between traditional teaching and TLIT.

Two different such programs are administered: one targeting the working class who cannot pursue daily tutoring, and another more structured program in three levels, each conveyed in 160 hrs of instruction over 9 months, roughly averaging 5 hours per week. Both programs have companion textbooks, workbooks and meticulous instructor's manuals outlining pedagogical elements. In the case of the structured program, the books are per level, and are organized into lesson units. A typical lesson would introduce about six basic consonants, or constructs, with their long and short vowel forms, with fragment, word and phrase examples. As our prototype, we set to develop a computer-aided equivalent of each lesson unit.

3.2 Project Prototype

The main principle in the lesson is a pedagogically sound presentation. We do not claim, at this stage, to completely replace the instructor. However, since the aim is to have a self-learning tool, we introduce changes which effectively complement the text lessons. For example, it is advised that many examples of words containing newly introduced sounds be given, in addition to exercises in which students are asked to map pictures to words, to utter proper sounds or transcribe proper symbols. Our design incorporates such additions into the flow of the lesson, and many more can be eventually added.

The next emphasis is the hallmark of our project, i.e. is the integration of speech and handwriting recognition, as well as audio and visual aids, into the flow. For example, some of the main elements of any lesson, together with the instances in which the technical components are introduced are illustrated in Figure 1. They are:

1. Presentation: In this section, the symbols are introduced. First, the symbol is shown and is sounded in its canonical form. Then, short vowel forms of the symbol are shown and sounded, together with an animated display of examples, which are also read out loud by the machine. The same is repeated for long vowel forms. A subtle addition which can be helpful is to animate the combination of canonical letters to form merged cursive symbols.
2. Initial Reading and Writing: For each symbol's short and long vowel forms, the student is prompted to utter the sound after the machine. Then, the student is prompted to write each of these forms, following the machine's animation of the writing motion.

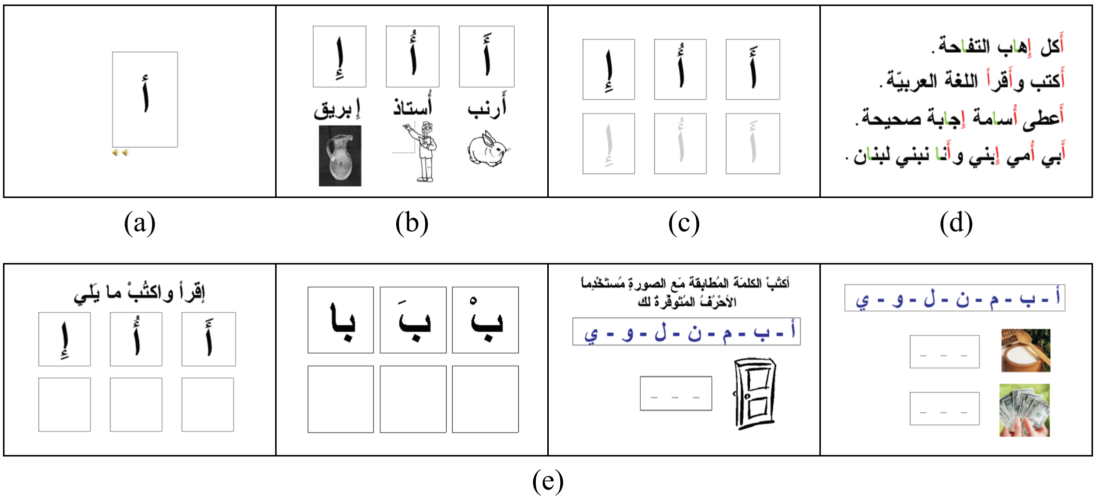


Figure 1 – Prototypical interface showcasing (a) presentation, (b) initial reading and (c) writing, (d) contextual listening, and (e) exercises.

3. Contextual Listening: A general paragraph with examples of the learned symbols is displayed and read out loud by the machine. The learned symbols are highlighted, to emphasize their use within the context of larger text.
4. Exercises: After the presentation, the initial reading and writing, and the contextual listening of a set of symbols, the lesson presents exercises to evaluate the student. Exercises do not provide the guidance given in the initial reading and writing, but use the same recognition engines. They either prompt visually and request uttered responses, or prompt with sound and request written responses. Some exercises are structured in the form of games, to engage the students more effectively.

4 Technical Components

The technical components which make TLIT unique, the speech and handwriting recognition engines, were offshoots of research and development performed at AUB, as outlined in Section 2.2. Potential commercial technologies to implement these components were initially investigated. We researched APIs for speech and handwriting recognition engines from Microsoft, Sakhr, and Aramedia. However, these commercial products either do not support Arabic, or if they do, they are not tailored to easily adapt to tutoring. For example, most of them are designed only for continuous speech recognition. Also, both speech and handwriting are trained on generic corpora, and are inflexible to retraining, whereas the project requires training on specific data, adapted to the lessons. When it comes to tutoring, it is also crucial to train the system not only to properly recognize correct spoken utterances or written gestures, but also to identify certain mistakes. Below, we delineate the details of the engines which we implemented.

4.1 Speech Recognition

For the speech component, given the simplicity of the task and the shortness of most speech utterances in the NCL lessons, we opted for Hidden Markov Model (HMM) based isolated word recognition [9], where a model is trained for each word using frame-based feature vectors, which are the commonly used Mel-Frequency Cepstral Coefficients (MFCC). The component was implemented in MATLAB and interfaced to Microsoft Visual C++. Figure 2 illustrates a simple data collection interface, which was used to generate the corpus of utterances.

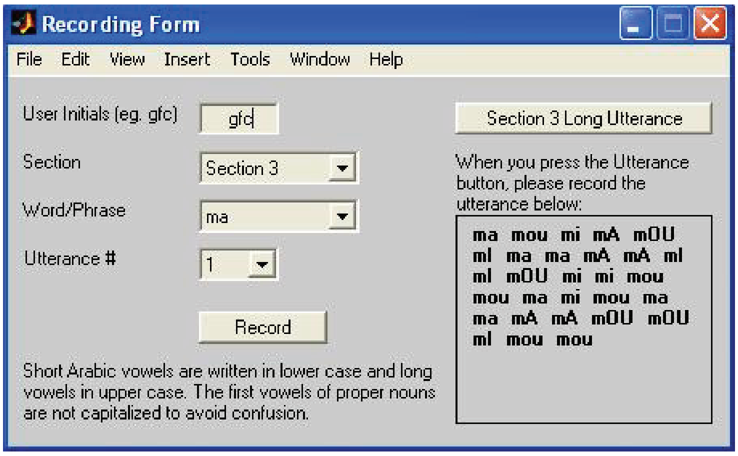


Figure 2 – Speech recording interface for corpus generation.

Audio is collected at 16 kHz. A front-end processor first pre-emphasizes the waveform and then generates overlapping frames of speech using 10ms Hamming windows every 4ms. These parameters can possibly be adjusted, to optimize performance. Frames are converted to 12-dimensional MFCC vectors, using the LPC method, and their temporal derivatives are appended to them to form the final observation vector. These are then quantized to 32 codebook vectors, to reduce the problem into a finite one. The HMMs are set to 6 states, and trained using a batched Baum-Welch algorithm [18], [19]. For recognition, maximum likelihood is performed, with simple confidence thresholding based on the difference between the two likeliest models.

We experimented with many adaptations of the basic engine to better suit the task at hand. For example, we tried to train separate HMMs corresponding to different types of mispronunciation. As mentioned above, by having such built-in models for errors the system can potentially provide practical guidance to the subjects, by giving more elaborate feedback than simple validation or rejection.

Finally, a promising alternative approach to the speech recognition engine is one employing artificial neural networks rather than HMMs. In particular, neural networks trained with the Al-Alaoui cloning algorithm [20] have shown considerable generalization capabilities [10].

4.2 Handwriting Recognition

For handwriting, we developed a multi-gesture recognizer, which uses the single-gesture recognition via feed-forward multi-layer neural networks as in [11], but interprets the input meaningfully by comparing relative distances, sizes, and slants of individual gestures to a library of predefined prototype-models. This component was directly implemented in Microsoft Visual C++.

The acquisition of handwriting is performed via a touch screen or pen, on a tablet PC. Upon inputting a letter, fragment, word, or sentence, the input is identified as a bundle of distinct connected gestures. Individual gestures correspond to connected fragments of Arabic handwriting, such as various cursive forms. Each gesture is initially represented as an ordered set of point coordinates. After proper interpolative downsampling to N points, where N can be adjusted for optimality, a gesture is then represented by the angle of the vector difference of consecutive points, with respect to a reference direction. This insures shift- and scale-invariance.

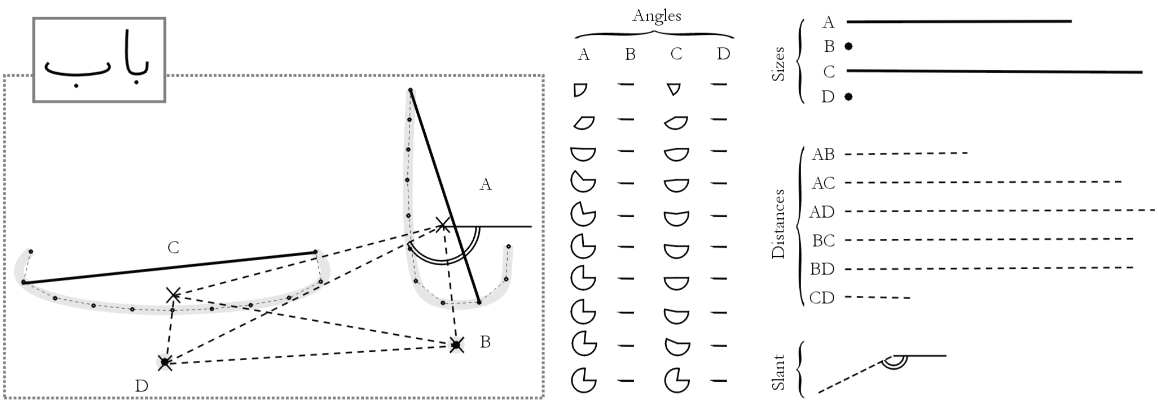


Figure 3 – Complete representation of handwritten input.

If classification of individual gestures is to be performed over M classes corresponding to M distinct cursive forms, a 3-layered neural network is employed, using sigmoidal activation functions, with $2N$ synapses at the input, $2N$ hidden neurons, and M output axons. The synapses are not subjected directly to the N angles of the gesture, but are fed the sine and cosine values of each angle (whence the $2N$ instead of N). Training is done via standard back-propagation either by artificial data, generated by distorting a prototype gesture, or using a training corpus. Current research is investigating the use of the aforementioned Al-Alaoui cloning algorithm for training [20]. Recognition is done by evaluating the network, and performing simple confidence thresholding on the difference between the two highest classes.

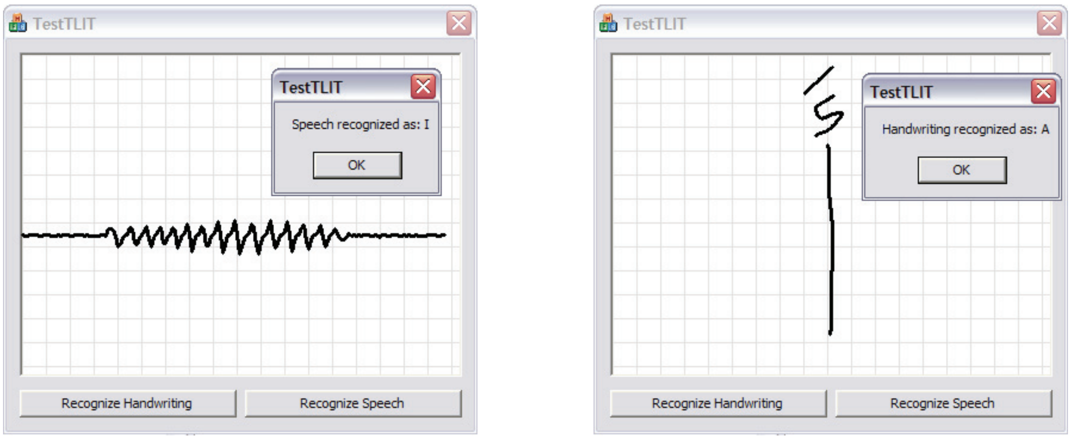
Once each gesture is individually recognized, as detailed above, the relative distances, sizes and slants of the gestures are computed. The complete representation of the input is illustrated in Figure 3. The final stage of recognition consists of comparison with a library of predefined prototype-models. For each model, this involves first checking that all gesture types are correct, and then evaluating a distance between the model's typical relative position, scaling and slant, and that of the input. If the match is close, within some tolerance, then recognition is declared. It is unlikely that more than one model matches, but in that case priority is given to the closer model. If all models fail, then the input is rejected.

For a more detailed description of technical and architectural details, the reader is invited to refer to [8]. It is worth noting that this multi-gesture recognition approach would, within a reasonable amount of additional work, provide useful feedback to the user. As in the speech engine, common mistakes can be assigned to separate models, and detected as such. Moreover, the representation allows correcting the user's own handwriting, by rescaling and realigning his/her input based on the correct word characterization. Such visual feedback is akin to a teacher's tutoring, and thus a potentially valuable addition to the system.

5 Ongoing and Future Work

Figure 4 shows the interface of a testbed for the speech and handwriting engines. Aside from a basic adaptation of the engines into the prototype lesson described in Section 3.2, some of the elements we are working on to enhance the learning experience are as follows:

1. Direct Computer Literacy: Although the project familiarizes the user with computers via their pervasive use, the software can include components which directly train the user in the fundamentals of computer literacy.



(a) (b)

Figure 4 – Interface of the technical component testbed, illustrating (a) speech and (b) handwriting recognition.

2. Everyday Scenarios: In line with the main philosophy of this project, which promotes enhancing the lives of the learners, one can envision the integration of various real scenarios as part of the lesson materials. For example, a hospital visit or an airport setting can be instructive scenarios for teaching numbers.
3. Smart assessment and adaptation: Technology can be used to enhance learning by means beyond speech and handwriting recognition. For example, the system can use the history of the performance of the learner in exercises and drill exams to maintain a user profile. This profile can then be used either by the monitoring instructor, or by the system itself to guide the learner, and control the presentation of subsequent topics.
4. Instructions: To make the software accessible to illiterates, it is imperative to have instructions spoken out loud. Including the textual versions of these instructions, and highlighting them, can provide the learner with further exposure to the learned material.
5. Appealing Graphics: The more appealing the presentation of the material is, the more engaging it can become to the learner. As such, we plan to obtain professional graphic design assistance.
6. Freelance Tools: Speech and handwriting can be integrated with the structured lesson system in more creative ways. Games and tools that can be explored at the learner's discretion can be useful. For example, one can envision a "dictionary" feature, where a small-vocabulary system can be queried by voice, to obtain the spelling and illustration of a word.

We conclude by noting that no amount of technology and creativity can replace experience in pedagogy. In that light, for the project to truly impact the speed and wider adoption of illiteracy-combating measures, technology and pedagogy should be interwoven, with research and development guided by educators who have worked first hand on the problem. Indeed, our follow up efforts with the Lebanese National Committee for Literacy are in this direction. Additionally, once the project is completed, we expect to assess its impact and compare it quantitatively with the traditional methodology, by simultaneously monitoring the progress of two groups, using the TLIT and traditional approaches respectively. We believe that, with such cooperation, assessment, and successive refinement, the initiative can indeed succeed in empowering illiterate adults faster and in greater numbers.

Acknowledgements

The authors greatly appreciate the support and cooperation of Ms. Amal Charara of the Lebanese National Committee for Literacy at the Lebanese Ministry of Social Affairs, and Ms. Nour Dajani of the UNESCO. Gratitude also goes to George Kadifa, President and CEO of Coreo Inc. and to Ibrahim Hajj, Dean of the Faculty of Engineering and Architecture at the American University of Beirut, for their support and encouragement. This research was partially supported by grants from the Rathmann Family Foundation, UNESCO, and URB (University Research Board) of AUB.

References:

[1] EFA Monitoring Report, Literacy for Life, Arab States, UNESCO, 2006.

[2] V. Abboud and V. Bunderson, "A Computer-Assisted Instruction Program in the Arabic Writing System", The University of Texas at Austin, Computer-Assisted Instruction Laboratory, Technical Report No. 4, 1971.

[3] F. Ehsani and E. Knodt, "Speech Technology in Computer-Aided Language Learning: Strengths and Limitations of a New Call Paradigm", Language Learning and Technology, Vol.2, No. 1, pp 45-60, July 1995.

[4] "Intelligent Tutoring Systems", Association for the Advancement of Artificial Intelligence. (http://www.aaai.org/AITopics/html/tutor.html)

[5] M. A. Al-Alaoui, "From Illiteracy to Computer Literacy: Teaching and Learning Using Information Technology (TLIT)", ESCWA Report, Beirut, July 31, 2003.

[6] M. A. Al-Alaoui, "Using Information Technology to Combat Illiteracy", Proceedings CIBIT 2004, First International Conference on Current Issues in Business and Information Technology, Haigazian University, Beirut, Lebanon, May 21-22, 2004.

[7] M. A. Al-Alaoui, "From Illiteracy to Computer Literacy: Teaching and Learning Using Information Technology (TLIT)", Pilot Project Proposal, FEA, AUB, 2005.

[8] M. A. Al-Alaoui, M. I. Ohannessian, G. Choueiter, C. Akl, and T. Avakian, "From Illiteracy to Computer Literacy: Teaching and Learning Using Information Technology", Progress Report, ECE Department, AUB, 2006.

[9] Y. S. Naous, G. F. Choueiter, M. I. Ohannessian, and M. A. Al-Alaoui, "Speech Recognition for Voice Based Control", Proceedings of IEEE ISSPIT 2002, Marrakesh, Morocco, pp. 527-531, December 18-21, 2002

[10] M. M. El Choubassi, H. E. El Khoury, C. E. Jabra Alagha, J. A. Skaf and M. A. Al-Alaoui, "Arabic Speech Recognition Using Recurrent Neural Networks", Proceedings of IEEE ISSPIT 2003, Darmstadt, Germany, December 14-17, 2003.

[11] M. A. Al-Alaoui, C. Akl, R. El-Kady, and R. Nassif, "Adult Literacy Using Information Technology", Proceedings of IMCL 2007, Amman, Jordan, April 18-20, 2007.

[12] M. A. Al-Alaoui, C. Akl, R. El-Kady, and R. Nassif, "Adult Literacy using Information Technology", International Journal of Emerging Technologies in Learning" - iJET, Vol. 2, No. 2, 2007.

[13] M. Abou Nehme, W. Khoury, B. Yameen, and M. A. Al-Alaoui, "Real Time Color Based Motion Detection and Tracking", Proceedings of IEEE ISSPIT 2003, Darmstadt, Germany, December 14-17, 2003.

[14] L. El-Afifi, M. Karaki, J. Korban, and M. A. Al-Alaoui, "'Hands-Free Interface'-A Fast and Accurate Tracking Procedure for Real Time Human Computer Interaction", Proceedings of IEEE ISSPIT 2004, Rome, Italy, December 18-21, 2004.

[15] J. Naoum-Sawaya, M. Slim, S. Khawam, and M. A. Al-Alaoui, "Dynamic System Design for American Sign Language Recognition", Proceedings of ISCCSP 2006, Marakesh, Morocco, 13-15 March, 2006.

[16] J. Azar, H. Abou Saleh, and M. A. Al-Alaoui, "Sound Visualization for the Hearing Impaired", Proceedings of ICL2006, Villach, Austria, September 27-29, 2006.

[17] J. Azar, H. Abou Saleh, and M. A. Al-Alaoui, "Sound Visualization for the Hearing Impaired", International Journal of Emerging Technologies in Learning - iJET, Vol. 2, No. 1, 2007.

[18] L. Rabiner and B. Juang, *Fundamentals of Speech Recognition*, PTR Prentice Hall, NJ, 1993.

[19] L. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", Proceedings of the IEEE, Vol. 77, No. 2, February 1989.

[20] M. A. Al-Alaoui, R. Mouci, M. M. Mansour, and R. Ferzli, "A Cloning Approach to Classifier Training", IEEE Transactions on Systems, Man and Cybernetics, Part A, Vol. 32, No. 6, pp. 746-752, November 2002.

Authors:

M. Adnan Al-Alaoui [1], Professor, *adnan@aub.edu.lb*

Mesrob I. Ohannessian [2], *mesrob@mit.edu*

Ghinwa F. Choueiter [2], *ghinwa@mit.edu*

Christine Akl [1], *akl_christine@hotmail.com*

T. Taline Avakian [3], *talinea@gmail.com*

Ismail Al-Kamal [1], *ima16@aub.edu.lb*

Rony Ferzli [4], *rony.ferzli@asu.edu*

[1] American University of Beirut
Department of Electrical and Computer Engineering
Beirut, Lebanon

[2] Massachusetts Institute of Technology
Department of Electrical Engineering and Computer Science
Cambridge, Massachusetts, USA

[3] University of Cincinnati
Counseling Program
Cincinnati, Ohio, USA

[4] Arizona State University
Department of Electrical Engineering
Tempe, Arizona, USA