International Journal of Bio-Science and Bio-Technology
Vol.5, No.6 (2013), pp.137-146
http://dx.doi.org/10.14257/ijbsbt.2013.5.6.14

# An Efficient Virtual Patient Image Model Interview Training in Pharmacy

Mira Park[1] and Peter Summons[2]

*[1]School of Electrical Engineering and Computer Science*
*The University of Newcastle, Australia*
*[2]School of Design, Communication & Information Technology*
*The University of Newcastle, Australia*
*[1]mira.park@newcastlel.edu.au, [2]peter.summons@newcastlel.edu.au*

***Abstract***

*This paper presents the development of a virtual patient simulation by a 3D talking head and its use by pharmacy students as a training aid for patient consultation. The paper concentrates on the virtual patient modeling, its synthesis with a speech engine and facial expression interaction. The virtual patient model is developed in three stages: building a personalized 3D face model; animation of the face model; and speech driven face synthesis. The model is used in conjunction with a training artificial intelligence module that creates several scenarios in which the student oral interview ability is assessed. The final evaluation phase is a randomized controlled trial at three partner universities: The University of Newcastle, Monash University and Charles Stuart University. It shows the potential to revolutionize the way pharmacy students' training is conducted.*

***Keywords:*** *3D talking head, virtual patient, pharmacy education, text to speech*

## 1. Introduction

Many variations of computerized virtual patient systems have been developed to assist the generation of case studies. The cases are presented by a computer to a student for the student's diagnosis and provide feedback to the student. An early automated virtual patient system, developed for pharmacy student assessment at Monash University [1], was based on a very prescriptive framework with an interface that domain teachers used to easily generate many case studies for assessment. The Web-Sp [2] virtual patient system employed static pictures of patients. It developed a framework that allowed the scalable generation of new medical scenarios in domains such as medicine, pharmacy and dentistry. Other systems have encouraged student interaction by providing video clips as responses to student questions [3, 4]. Video images allow dynamic system responses from the virtual patient to convey emotional aspects, such as pain or frustration, which conveys a richer interactive experience for the student than the word content provided by static image virtual patient systems. Dynamic interaction is also achieved using virtual reality avatars, which are computer generated animations used as representations of patients [5]. One of the best examples of an avatar based

ISSN: 2233-7849 IJBSBT

virtual patient operating within a virtual reality world is the Digital Animated Avatar (DIANA), created by the University of Florida. DIANA is a female virtual character who plays the role of a patient with appendicitis, while another virtual interactive character, a male virtual character, plays the role of an observing expert [6]. A similar system has been developed in conjunction with the Pharmacy program at Keele university in the United Kingdom [7, 8], which has implemented avatars for Monash University as part of its ePharm program [1].

This paper presents the development of Virtual Pharmacy Patient (VPP) system using pharmacists as domain experts. The virtual patient acts as an effective alternative for practicing assessment, diagnosis, treatment, and interpersonal skills for pharmacy students. With the introduction of 3D interactive virtual patients, pharmacy students can increase the amount of practical case experience which they should receive before working with actual patients.

The VPP system uses efficient and low-cost techniques to construct a realistic, three-dimensional (3D) model from two two-dimensional (2D) face images that can be of any existing person. This enables the implementation of different patient representations, both in age, gender and racial characteristics. The development of the final dynamic simulation is achieved in three stages. Firstly, a personalized 3D face model is built based on the two 2D face images using Faceworx, which is a fully automatic 3D face shape and texture reconstruction framework. Secondly, using the software application Blander, the face model is animated by expression channels and complemented by visual prosody (intonation, rhythm and phrasing) channels that control the major facial features, such as mouth, eye and eyelid movements. Finally, the facial animation is combined and synchronized with the emotive synthetic speech generated by incorporating an emotion transformer into a speech engine for text to neural speech synthesizer.

Section 2 outlines the development of the 3D virtual patient head, covering both the creation of the facial image and integration of facial expressions; and Section 3 presents the integration of speech for the visual interaction of a virtual pharmacy patient simulation that. Some evaluation results of the use of the virtual pharmacy patient as a training aid for patient consultation by pharmacy students in trials over three universities are also presented in Section 4.

## 2. Creation of a Talking Head

### 2.1. 3D Face Reconstruction from 2D Face Images

In general, building a 3D face model customized to a particular person involves the use of an expensive 3D scanner or calibrated camera array and tedious manual labeling work, which are not affordable for most academic applications. Hence computer vision researchers have developed various techniques and algorithms to recover 3D face information (shape & texture) from one or more 2D face images, such as shape from shading (SFS) [9] and model-based bundle adjustment (MBA) [10].

 

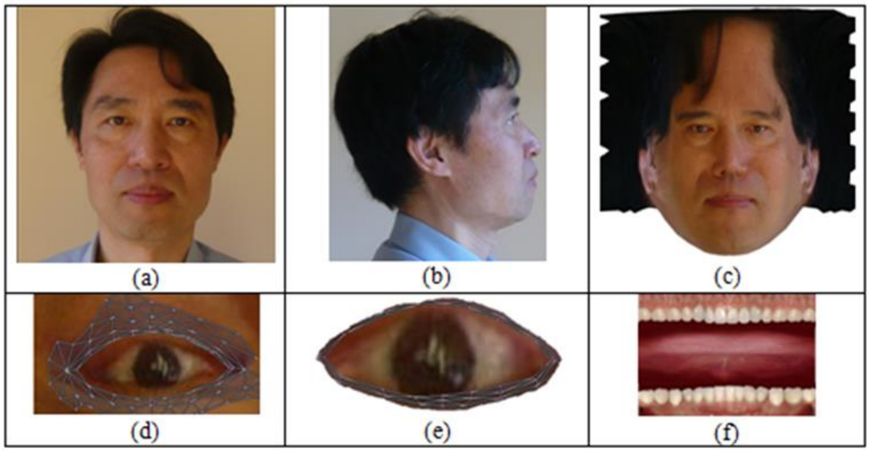


**Figure 1. Base Pictures for a Talking Head**

Faceworx [11] a software application that creates a 3D head out of two standards 2D images (front and side) of a face (Figure 1(a),1(b)) was used to construct the 3D Head. This software provides color texture along UV mapped mesh model and skin texture is modified to adjust UV coordinate. For the rendering operation on texture space, it was required to have all polygons has distinct, non-overlapped UV map (Figure 1(c, d, e)). The software demands some skill in placing reference points and marking the contours of the face; mouth, nose, ears and eyes. The amount of points in each virtual face can be changed to create a more detail and a more realistic 3D face.

The final 3D portrait can be saved and exported in the well-known OBJ format (wavefront format) for use in other 3D modelling software, such as Blender. OBJ is a geometry definition file format that was developed by Wavefront Technologies for its Advanced Visualiser animation package. The OBJ file format is a simple data format that represents 3D geometry alone — namely, the position of each vertex, the UV position of each texture coordinate vertex, normals, the faces that make each polygon (defined as a list of vertices), and texture vertices. Vertices are stored in a counter-clockwise order by default, making explicit declaration of normals unnecessary. Java3D loads the OBJ files and each OBJ file contains all the meshes of a model. The number of vertices should be the same for every expression in the 3D face model and for each expression in the separated files.

### 2.2. Facial Expression

The model that is exported from the Faceworx software is imported into the “Blender [12]” software application. Blender uses a generic 3D model with texture mapping from a set of images. The software can reconstruct a new 3D head model from a set of images, and thus generate a new facial expression. Using 3D models is suitable when the talking head actions are gross movement motions and rotations.

**2.2.1. Head Model:** The VPP models a human head through the application of a three-dimensional mesh model. A neutral face is the initial frontal face image and does not have any specific facial expressions. By warping the input image, we can morph the neutral face into various expressions. A set of points, representing facial points of interest, are marked on the neutral face as control vertices. These vertices are placed around the contour of specific features on the face, such as its eyes, nose, mouth, and chin. These control vertices are then connected to form convex polygons, such as triangles. Thus an entire facial image can be simplified and represented as a set of polygons. In the model we developed, each head contained a complete set of teeth that allows a full range of expressions and mouth shapes.

**2.2.2. Face Mesh Fitting:** As indicated, the first stage in the development of the talking head system fits a generic three-dimensional face mesh to a model's face image (Figure 1(c)). After the front and side image are inputted, a generic 3D mesh is applied to the face image. A boundary box is used to approximate the head size of the resulting image, and a user can manually adjust control points to fit with feature points, for example, the eyes, nose and lips on the image. To generate a realistic model's face, a generic face model is manually adjusted to the model's face image. The generic face model has all of the control vertices for facial expressions defined by a 3D movement of grid points to modify geometry. A frontal face image is input to the system and then corresponding control vertices are manually moved to a reasonable position on the face image using a mouse. A synthesized face emerges by mapping the blended texture, generated by the model's frontal image, and the profile image onto the modified personal face model. The rotation angle of the face model can be controlled in a preview window to achieve the best fit to a face image that has been captured from any arbitrary angle. The model is a mesh of 3D points controlled by a set of conformation and expression parameters. The control vertices are adjusted via the relative location of facial feature points such as the eye and lip corners. Changing these vertices can re-shape a base model to create new heads.

**2.2.3. Mouth Shape Generation:** Many mouth shapes of alphabets are quite similar to each other, and all mouth shapes of alphabets can be simulated by combining basic mouth shapes. In our model, 10 basic mouth shapes are adopted, such as basic shapes formed by ai, cdg, e, fv, l, mbp, o, u and wq.

**2.2.4. Synthesizing Facial Expression:** As mentioned above, various facial expressions can be synthesized by mapping part of the original texture to specific polygons that are defined by control vertices. Putting the texture mapped mesh model and the background together, the resulting image scene looks just like the original face with some specific facial expressions. The first step to animate the facial expressions is to define the key frames that make up the major facial expression feature changes. The neutral face without any facial expressions can be thought as a key frame that contains a neutral facial expression and this is the base image that is varied to produce specific facial expressions.

**2.2.5. Generic Teeth Model:** The VPP model has one teeth model (Figure 1(f)) over the lips. The one inner mouth model has both the upper and lower teeth appearing as flat rows behind the lips. The generic inner mouth model can be resized according to the dimensions of the mouth size in the neutral face image. This would be more realistic if all inner mouth models were employed so that they could be moved separately – for instance the upper teeth size and shape could change to be more of a curve when smiling.

## 3. Speech Driven Face Synthesis

After the 3D face mesh is adjusted, it can be used to animate facial expressions driven by speech. To synthesize animations of facial expressions synchronized with speech data, we must know which phonemes appear in the input data. In addition, the start and stop time of a certain phoneme should be obtained to synchronize the mouth shapes with speech wave data.

 

For example, assume that the system is required to speak the sentence, “How are you?”, the system invokes a speech engine and finds that from StartTime to TimeA is silence; TimeA to TimeB should be the interval taken to speak “How”; TimeB to TimeC should be the interval taken to speak “are”; and TimeC to EndTime should be the interval taken to speak “you”. The system then translates these results into neutral (from time 0 to TimeA), How (from TimeA to TimeB), are (from TimeB to TimeC), you (from TimeC to EndTime) and appropriate key frames are fetched from the expression pool to represent these lip movements.

Speech is usually treated in a different way to the animation of facial expressions because simple keyframe-based approaches to the animation typically provide a poor approximation to real speech dynamics. Text-to-Speech (TTS) functionality allows the models to speak any text dynamically with lip-synching in real time.

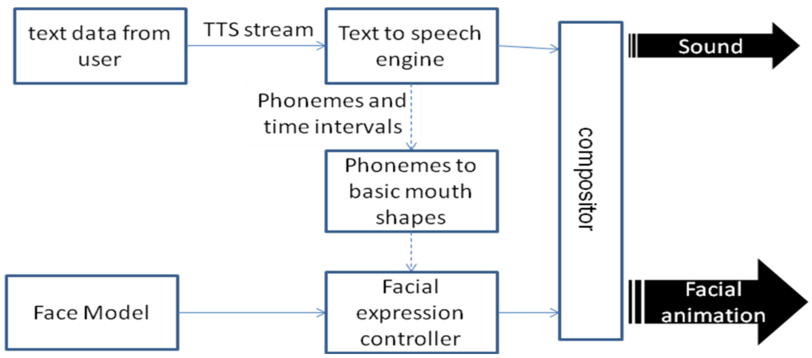


**Figure 2. Architecture for Synchronization of Facial Expressions and Speech**

Figure 2 is the flow diagram of our synchronization of facial expressions and speech. Firstly, text data, which is input to TTS engine, has been generated by the reasoning and assessment modules as a response by the VPP system to a pharmacy student’s question. The engine compares the input text data with phonemes in a database and then sends the phonemes of the synthesized speech and interval information for each phoneme to a face expression controller. The expression controller translates the phonemes and timing information into face expression parameters (mouth shapes). Thus, we can get basic facial expressions according to the input text data. With this information, facial animations are synchronized with the input text data. For example, a word “how” pronounced as /hau/ is converted to be /h/ +/au/ and the corresponding mouth shape is from “h” then gradually morphed to “au”.

TTS capabilities refer to its ability to play back text as a spoken voice. The VPP used two kinds of TTS engines. This was due to the different operating systems and platforms on which the evaluation was carried out in the three universities. The first TTS engine is a generic one that supports Microsoft API from Microsoft SAPI 5.1 (Microsoft Speech Application Interface version 4.0); the other TTS engine is the Java Speech API (JSAPI). JSAPI was used for the male virtual patient model’s voice, and SAPI was used or the female patient’s voice.

JSAPI allows Java applications to incorporate speech technology into their user interface. It defines a cross-platform API to support command and control recognizers, dictation systems and speech synthesizers. JSAPI can be used by FreeTTS which is an open source speech synthesizer written entirely in the Java programming language.

SAPI offers a general framework for building speech synthesis systems. SAPI contains many interfaces and classes for managing speech. For TTS, the base class is SpVoice. The C++ Win32 console application produced one Windows Dynamic link library (DLL) file, which allows a Java program to access TTS functionality provided by the SAPI.

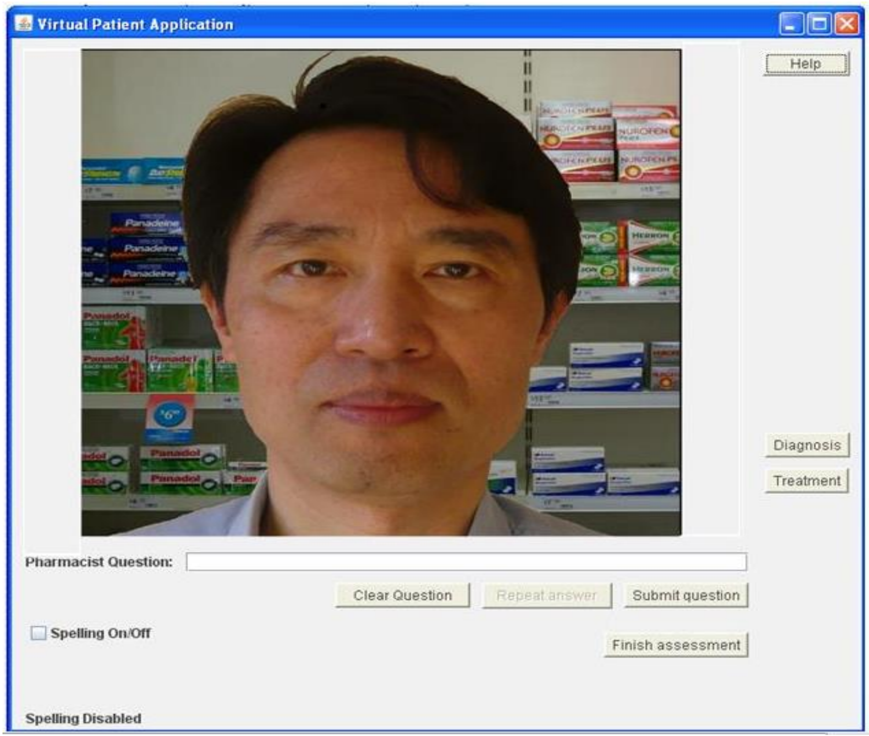


**Figure 3. Virtual Patient System User Interface**

## 4. Results

For the trial evaluation study (Fig 3), students from three universities (The University of Newcastle, Monash University and Charles Stuart University) were assessed in a randomized controlled trial over three days [13]. They were assessed on skills such as their coverage, convergence and style of investigative questions that they used to diagnose three clinical scenarios: a cough, Gastro-oesophageal reflux disease (GORD), and constipation. Each condition was presented with 3 levels of severity: mild, moderate and severe. At each assessment session the VPP system presented a student with all 3 conditions at randomly selected severities before re-assessing the same conditions at different levels of severity at an assessment session held on another day.

The students who consented to be in the trial study and who used the VPP system in the trial comprised of: The University of Newcastle, 15 of 83 eligible students; Monash University, 15 of 220 eligible students; and Charles Stuart University, 3 of 110 eligible students. Students evaluated their experience with the VPP on several different levels, the software, the appearance, the learning outcomes, etc. Twenty-two students used the VPP system answered questions relating to their interaction with the virtual patient in the final survey.

The students felt the VPP system helped them to identify areas of their communication that they could work on (100% vs 56% agreeing/strongly agreeing), and that using the virtual patient will improve their confidence with real patients (90% vs 56% agreeing/strongly agreeing).

With regards to the appearance of the virtual patient, respondents were generally negative, in particular with regards to the voice. Some of the results for the physical visual appearance and the voice of the VPP are reproduced in Table 1.

The responses for domestic versus international students (see Table 2) indicated that domestic students were overall more positive about the virtual patient than the international students.

Respondents indicated that the limited number of facial expressions and visual clues from the VPP system was a drawback to realistic interaction. As indicated in Newby et al [13], this made it difficult for students to judge the effect of their questioning, that is, whether their questions were appropriate or not:

- *"I think maybe if it had more facial expressions it would be good" Charles Sturt student*
- *"There was no emotion in it that you would get in a normal patient" Monash student*
- *"it's not going to teach you to read patients reactions to how to you, like to how you ask the questions, whether it thinks that's an inappropriate question or the way you phrase that was not very polite..." Newcastle student*
- *"...you don't have that sort of feedback response, you can't see ok, that I should have gone through a different path there…" Newcastle student*

**Table 1. Agreement with Statements about the Reality of the Virtual Patient**

| **The virtual patient…** | **Agree/Strongly agree** | |
|---|---|---|
| | **(n=22)** | **%** |
| appeared authentic | 9 | 41 |
| acted like a real patient | 10 | 46 |
| appearance fitted the role | 14 | 64 |
| simulated physical complaints unrealistically | 8 | 36 |
| answered questions in a natural manner | 6 | 27 |
| voice had a good pitch | 10 | 46 |
| voice was difficult to understand | 12 | 55 |
| | | |
| I felt I was making decisions as a pharmacist would make in real life | 15 | 68 |
| I felt I was the pharmacist looking after this patient | 13 | 59 |

**Table 2. Agreement with Statements about the Reality of the Virtual Patient**

| The virtual patient… | Agree/Strongly agree (n,%) | |
|---|---|---|
| | Domestic (n=13) | International (n=8) |
| appeared authentic | 8 (62) | 1 (13) |
| acted like a real patient | 8 (62) | 2 (25) |
| appearance fitted the role | 11 (85) | 2 (25) |
| simulated physical complaints unrealistically | 4 (31) | 3 (38) |
| answered questions in a natural manner | 5 (38) | 1 (13) |
| voice had a good pitch | 7 (54) | 3 (38) |
| voice was difficult to understand | 7 (54) | 4 (50) |
| | | |
| I felt I was making decisions as a pharmacist would make in real life | 11 (85) | 3 (38) |
| I felt I was the pharmacist looking after this patient | 9 (69) | 3 (38) |

## 5. Discussion

This paper presented the visual and vocal aspects of an automated system that allowed pharmacy students to interact with a VPP. The VPP allows students to explore the full patient consultation and to let them practice interpersonal skills before working with actual patients. The open source code of the VPP system can be downloaded from resweb.newcastle.edu.au/VirtualPatient/private/uploads on the authors' permission.

As indicated in the Section 4, respondent to the appearance and the voice of the virtual patient need to be improved. As indicated in Section 2.2.4, the freeware software (Blender) used for the facial modelling required a lot of effort and the number of facial expressions were limited due to the time needed to construct the expressions. In the trial only three expressions were used: neutral, smile and laugh. There is another problem to be solved. The facial expressions are changed by morphing between expressions, but VPP system does not facilitate the change of expression while the VPP is speaking. The expressions are changed at the start and the end of speaking. Both of these factors contributed to the unnatural appearance of the facial responses of the VPP system, particularly when it was responding to student questions.

In an updated version of the 3D face model that is generated by the commercial software FaceGen, much quicker development is possible by utilizing through default expressions and visually adjusting parameters in real time to generate multiple facial expressions. They can then be exported to the virtual patient software so that the software can morph between exported expressions to create expression transitions, such as neutral to smile and smile to angry. The inner mouth of the model is separated into three parts: the upper teeth, the lower teeth and the tongue. The upper teeth model is moved according to the control vertex at the philtrum, and the lower one is moved according to the control vertex at the chin. This inner mouth model solves an unnatural view problem when the lips move to talk or to smile. This also allows the breaking up of the VPP speech output to smaller segments to link the morphing of facial expressions to the duration of the spoken output, resulting in a more natural morphing of mouth expressions between these smaller speech segments. The content of the spoken output is thus made richer and capable of providing emotional cues by incorporating expression changes and head movements within the speech output.

 

To cater for many students interacting with the patient at the same time and in the same room, students talked with the virtual patient by typing questions into an interface and the virtual patient responded verbally to students via earphones. Typing questions rather than speaking directly to the VPP also reduced the ambiguities that might have been introduced by voice recognition systems.

The VPP is used in conjunction with a training artificial intelligence module [14, 15, 16] that creates several scenarios in which the student oral interview ability is assessed.

## References

[1] Pharm, Monash University, www.pharm.monash.edu.au/education/epharm/virtualpatient.html.
[2] N. Zary, G. Johnson, J. Goberg and U. G. Fors, “Development, Implementation and Pilot Evaluation of a Web-based Virtual Patient Case Simulation Environment – Web-Sp”, BMC Medical Education, vol. 6, no. 10, **(2006)**, pp. 1-17.
[3] R. Bergin and U. Fors, “Interactive Simulated Patient: an Advanced Tool for student-activated learning in Medicine and Healthcare”, Computers and Education, vol. 40, no. 4, **(2003)**, pp. 361-376.
[4] S. Farrar, “The Arizona Virtual Patient: Using Question-answering Technology to Enhance Dialogue Processing”, Proceedings of HLT2002 Second International Conference on Human language Technology research, **(2002)**, pp. 222-225.
[5] M. Cavazza and A. Simo, “A virtual patient based on qualitative simulation”, Proceeding of IUI’03: international conference on intelligent user interface, **(2003)**, pp. 19-25.
[6] B. Lok, R. E.Ferdig, A. Raij, K. Johnsen, R. Dickerson, J. Coutts, A. Stevens and S. Lind, “Applying virtual reality in medical communication education: current findings and potential teaching and learning benefits of immersive virtual patient”, Virtual Reality, vol. 10, **(2006)**, pp. 185-195.
[7] Keele university: Virtual Patient Demonstration, www.keele.ac.uk/schools/pharm/explore/vp.htm.
[8] P. J. Online, “Avatars help Keele students hone skills”, The pharmaceutical journal, vol. 280, no. 1, **(2008)**, pp. 249, www.pharmj.com/pdf/articles/pj_20080301_atavars.pdf.
[9] R. Zhang, P. S. Tsai, J. E. Cryer and M. Shah, “Shape from shading: a survey”, IEEE Transaction on PAMI, vol. 21, **(1999)**, pp. 690-706.
[10] Y. Shan, Z. C. Liu and Z. Y. Zhang, “Model-based Bundle adjustment with application to face modeling”, International Conference on Computer Vision, vol. 2, **(2001)**, pp. 644-651.
[11] http://looxis-faceworx.softwareandgamees.com.
[12] www.blender.org.
[13] D. A. Newby, J. S. Jin, P. F. Summons, R. I. Athauda, M. Park and J. J. Schneider, “Development of a computer-generated digital patient for teaching and assessment in pharmacy: Final Report”, Australian Learning and Teaching Council, Surry Hills, vol. 75, **(2011)**.
[14] P. Summons, D. Newby, R. Athauda and M. Park, “Modelling a Simulated Pharmacy Patient”, $9^{th}$ Annual International Industrial Simulation Conference ISC’201, **(2011)**, pp. 68-72.
[15] P. Summons, D. Newby, R. Athauda, M. Park, L. Pranata, P. Shaw, J. Jin and Y. Xu, “Design Strategy for a Scalable Virtual Pharmacy Patient”, Australasian Conference on Information System, **(2009)**, pp. 96-110.
[16] M. Park and P. Summons, “A Computer-generated Digital Patient for Oral Interview Training in Pharmacy”, Artificial Intelligence and Application, **(2013)**.

 