

A Computer-generated Digital Patient for Oral Interview Training in Pharmacy

Mira Park¹, Peter Summons²

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
mira.park@newcastle.edu.au

Abstract. This paper outlines the development of an image model and its synthesis with a speech engine and facial interaction for a virtual patient. The simulation is for a pharmacy patient used by pharmacy students as a training aid for patient consultation at three Australian universities. The patient is simulated by a 3D talking head. The development of the final virtual patient image model is achieved in three stages: Building a personalized 3D face model; animation of the face model; and speech driven face synthesis. The image was used in conjunction with a training artificial intelligence module that created several scenarios in which the student's oral interview ability was assessed in students. Some student evaluation on the image and speech realism is also presented.

eywords: 3D face, virtual patient, pharmacy education, Oral interview.

1 Introduction

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 they receive before working with actual patients.

An early automated virtual patient system, developed for pharmacy student assessment at Monash University [5], 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 [6] virtual patient system employed static pictures of patients. Other systems have encouraged student interaction by providing video clips as responses to student questions [7, 8]. Dynamic interaction is also achieved using virtual reality avatars, which are computer generated animations used as representations of patients [9]. One of the best examples of an avatar based 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 [10]. A similar system has been developed in conjunction with the Pharmacy program at

Keele university in the United Kingdom [11, 12], which has implemented avatars for Monash University as part of its ePharm program [5].

This paper outlines the development of a virtual pharmacy patient simulation by a 3D talking head. Section 2 presents the development of the imaging and visual interaction and Section 3 the speech generation for a virtual patient simulation. Some evaluation results of its image and speech realism, when used by pharmacy students as a training aid for patient consultation, are 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) [1] and model-based bundle adjustment (MBA) [2].

Faceworx [3] a software application that creates a 3D head out of two standard 2D images (front and side) of a face (Fig 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 (Fig 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 real face.

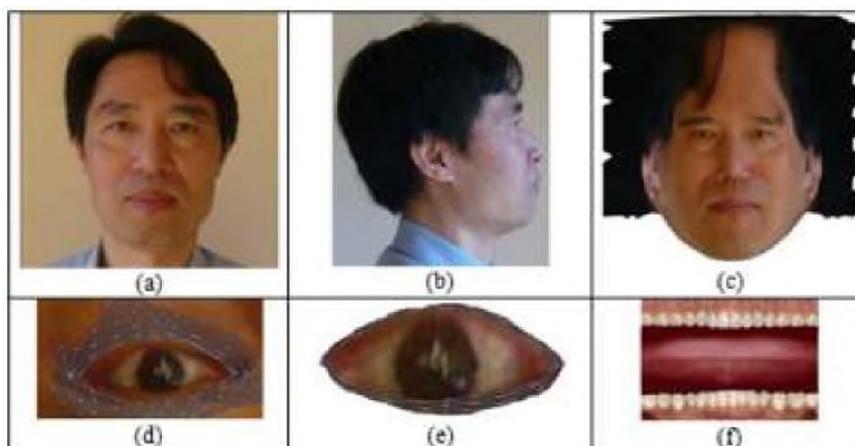


Fig. 1. Base pictures for a talking head

2.2 Facial expression

The model that is exported from the Faceworx software is imported into the "Blender" software application [4]. 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, but it takes a lot of effort to fit the 3D model for the set of images.

2.2.1 Head model

The talking head used in the trials described in Section 4 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. The colour information of the input neutral face provides a base image for the system. By warping the input image, the neutral face can be morphed into various expressions. A set of spots, representing facial points of interest, are marked on the neutral image 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.

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 (Fig 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.

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 expressions

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.

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, knowledge is required as to 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 animation typically provide a poor approximation to real speech dynamics. Text-to-Speech (TTS) functionality allows the models to speak any text dynamically, in real time and with lip-synching. Firstly, text data that has been generated by the reasoning and assessment modules, as a response by the virtual pharmacy patient system to a pharmacy student's question, is input to a text-to-speech engine. The engine compares the input text data with phonemes in a database and then sends the phonemes of the synthesized speech, as well as 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, basic facial expressions can be obtained according to the input text data that is to be spoken by the talking head model. With this information, facial animations that are synchronized with the input text data by the techniques mentioned above can be generated. 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".

4 Results

The Virtual Pharmacy Patient (VPP) simulation was used in a randomized controlled trial over three days [13] by students from three Australian universities (The University of Newcastle, Monash University and Charles Stuart University). Students were assessed on interview skills 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 Virtual Pharmacy Patient (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. 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. At the end of each session the VPP gave feedback to the students about their performance. The students who consented to be in the pilot study and who used the Virtual Pharmacy Patient (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 Sturt University, 3 of 110 eligible students - however, Charles Sturt University did not have enough students to constitute a control study group and so their results were excluded. Students evaluated their experience with the VPP on several different levels, the software, the appearance, the learning outcomes, etc [13,14,15]. Twenty-two (73%) of the thirty students who used the VPP answered questions relating to their interaction with the virtual patient in the final survey.

The differences between students who have worked in pharmacy for more than year to those that have worked for less than a year or not worked indicated that they felt the VPP helped them 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 virtual patient...	Agree/Strongly agree	
	n	%
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 1. Agreement with statements about the reality of the virtual patient.

5 Conclusions

This paper presented the visual and vocal aspects of an automated system that allowed pharmacy students to interact with a computer-generated digital patient. Using the computer-generated digital patient allows students to explore a realistic patient consultation and to practice interpersonal skills before working with actual patients.

Problem areas of development were outlined in regards to future application of the simulation. Solutions to these were indicated through the use of commercial software, resulting in the creation of more facial expressions and hence more realistic morphing, and in the more realistic embedding of speech due to the increased number of expressions and resultant smaller morphing time between expressions.

References

1. R. Zhang, P.S. Tsai, J.E. Cryer, and M. Shah, Shape from shading: a survey. *IEEE Transaction on PAMI* 21 (1999) 690-706.
2. Y. Shan, Z.C. Liu, and Z.Y. Zhang, Model-based Bundle adjustment with application to face modeling. *International Conference on Computer Vision 2* (2001) 644-651.
3. <http://looxis-faceworx.softwareandgamees.com/>
4. www.blender.org.
5. ePharm, Monash University, www.pharm.monash.edu.au/education/epharm/virtualpatient.html
6. Zary, N., Johnson, G., Goberg, J., Fors, U.G., Development, Implementation and Pilot Evaluation of a Web-based Virtual Patient Case Simulation Environment — Web-Sp, *BMC Medical Education* 6(10):1-17 (2006)
7. Bergin R., Fors U. Interactive Simulated Patient: an Advanced Tool fro student-activated learning in Medicine and Healthcare, *Computers and Education*, 40(4):361-376 (2003)
8. Farrar, S., The Arizona Virtual Patient: Using Question-answering Technology to Enhance Dialogue Processing, *Proceedings of HLT2002 Second International Conference on Human language Technology research*, 222-225 (2002)
9. Cavazza M., Simo A., A virtual patient based on qualitative simulation, *Proceeding of IUI03: international conference on intelligent user interface*, 19-25 (2003)
10. Lok B., Ferdig R.E., Rajj A., Johnsen K., Dickerson R., Coutts J., Stevens A., Lind S., Applying virtual reality in medical communication education: current findings and potential teaching and learning benefits of immersive virtual patient, *Virtual Reality*, 10:185-195 (2006)
11. Keele university: Virtual Patient Demonstration, www.keele.ac.uk/schools/pharm/explore/vp.htm
12. PJ Online, Avatars help Keele students hone skills, *The pharmaceutical journal* 280(1):249, www.pharmj.com/pdf/articles/pj_20080301_atavars.pdf (2008)
13. Newby D.A., Jin J.S., Summons P.F., Athauda R.I., Park M., Schneider J.J., et al. Development of a computer-generated digital patient for teaching and assessment in pharmacy: Final Report, Australian Learning and Teaching Council, Surry Hills, 75 (2011)
14. Summons P., Newby D., Athauda R., Park M., Modelling a Simulated Pharmacy Patient, *ISC'2011, 9th Annual International Industrial Simulation Conference ISC'2011*, 68-72 (2011)
15. Summons P., Newby D., Athauda R., Park M., Pranata L., Shaw P., Jin J., Xu Y., Design Strategy for a Scalable Virtual Pharmacy Patient, *Australasian Conference on Information Systems*, 96-110 (2009)