Making Clinical Simulation Mannequins Talk

Isac Cossa[1], Guilherme Campos[2], Pedro Sá Couto[3,4] and João Lindo[4,5]
[1]*Higher School of Nautical Sciences, Avenida 10 de Novembro, Maputo, Mozambique*
[2]*Dept. of Electronics, Telecom. and Informatics (DETI) / IEETA, University of Aveiro, 3810-193 Aveiro, Portugal*
[3]*Dept. of Mathematics (DMAT) / CIDMA, University of Aveiro, 3810-193 Aveiro, Portugal*
[4]*School of Health Sciences (ESSUA) / Clinical Simul. Centre (SIMULA), University of Aveiro, 3810-193 Aveiro, Portugal*
[5]*Centre for Health Technology and Services Research (CINTESIS), 4200-450 Porto, Portugal*

Keywords: Speech Synthesis, Pre-recording, Case Study, Glasgow Coma Scale, Consciousness, Model, Markov Chain, Training, Instructor, Usability, Evaluation.

Abstract: This paper advocates the interest of applying speech synthesis to clinical simulation mannequins, by focussing on a particular case study of recognised practical interest (evaluation of consciousness level based on the Glasgow Coma Scale), chosen as a proof of concept. A response repository comprising 109 sentences was recorded and an application was developed in Microsoft® *Visual Basic*® to allow configuration of the simulation scenario, control of response generation on a low-fidelity mannequin equipped with a loudspeaker and assessment of trainee performance. The system received very positive assessment in initial user tests on a typical training setting.

1 INTRODUCTION

The work presented here corresponds to the first author's MEng dissertation project. Its main motivation was the rapid development observed in the fields of clinical simulation and speech synthesis.

Because of its multiple advantages from the standpoint of training procedure flexibility and safety, clinical simulation based on mannequins has grown and evolved very rapidly in recent decades. It now plays an essential role in health care professional training programmes in virtually every field of specialisation (Brazão et al., 2015). Commercial mannequin models range from *low-fidelity* (relatively robust and inexpensive but with hardly any automated features – mostly used at initial training stages) to *high-fidelity* (featuring complex automation but very expensive and relatively fragile – normally reserved for advanced courses). In any case, their sophistication level tends to increase, as realistic interaction with trainees is obviously the main goal. The ability to simulate verbal communication can be a very valuable contribution towards this goal. According to instructors working in medical school laboratories, trainees tend to develop a kind of emotional attachment to clinical simulation mannequins. Almost invariably, they name them and engage in pretend conversations during training sessions, as if the mannequins could follow the procedures and chat back. Furthermore, some crucially important clinical evaluation and diagnosis skills are directly dependent upon verbal communication with patients. The training settings should help practitioners develop those skills.

Speech and language processing technologies have also been the object of increasing research and development interest in the past 50 years. Speech synthesis is one of the most active fronts. Algorithms and applications are being constantly refined in terms of audio output quality (realism), computation time (to afford real-time operation) and application range (availability in multiple languages). Practical usage is growing in every field involving human-machine interaction, including simulation (Taylor, 2009).

The potential for integration of sound synthesis in clinical simulation settings is largely underexplored. Simulation mannequins incorporating speech and other vocal sound generation systems should be more widespread, even because upgrading low-fidelity mannequins for this purpose is relatively easy. In support of this position, this paper presents a pilot study on a particular health care training application of widely recognised usefulness.

Cossa, I., Campos, G., Couto, P. and Lindo, J.
Making Clinical Simulation Mannequins Talk.
DOI: 10.5220/0007571404500455
In *Proceedings of the 12th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2019)*, pages 450-455
ISBN: 978-989-758-353-7

Following introductory sections devoted to describing a typical clinical simulation scenario and very briefly reviewing available speech synthesis software, section 4 justifies and presents the chosen case study. Section 5 describes the software application developed to implement it and section 6 reports initial user evaluation tests. The final section discusses the results and considers some future work possibilities.

2 CLINICAL SIMULATION

Health care training based on clinical simulation typically involves the use of two rooms. Trainees (e.g. nursing students) take their place in one of them, normally referred to as *simulation* or *training* room, which contains the simulation equipment proper (e.g. mannequin 'patient' lying on a hospital stretcher). The instructor (e.g. nursing teacher) normally sits at the *control* or *observation* room. Figure 1 shows a typical example, with adjacent training and control rooms, separated by a glass window. Special 'intelligent' glass can be used to make visual contact unidirectional (from control room to training room only).

Figure 1: Clinical simulation setting (UNIFAE, 2017).

Normally, training rooms are fitted with audio-visual equipment (lighting, video cameras, microphones…), making it possible to record the sessions and so document trainee performance for analysis, evaluation and debriefing – a crucial element in the training process. Also, so long as the equipment is controllable remotely, multiple training rooms can be assigned to a single, non-adjacent control room.

In addition to observing and evaluating trainee performance, it is desirable that the instructor be able to control the training sequence, by specifying the condition of the 'patient' and her reactions to each possible trainee action. For this purpose, simulation mannequins tend to incorporate increasingly sophisticated hardware and software.

3 SPEECH SYNTHESIS

Speech synthesis systems generate human vocal sound in a given language from content specified in written form (text file). For this reason, they are also called *Text-to-Speech* (TTS) systems. As Figure 2 illustrates, they comprise two main blocks: the front-end translates the original text into phonetic symbols; the back-end applies digital signal processing algorithms to generate the corresponding audio files (Barros, 2012).

Figure 2: Block diagram of a TTS system.

TTS systems are useful in an increasing number of applications, such as presentation and digital book readers, disability aids and automatic messaging.

Table 1 lists some of the most popular software packages which have been developed for speech synthesis (Freewarenee, n.d.); (Freewareneesite, n.d.); (Manfio, 2012); (Araújo and Silva, 2013).

Table 1: Popular TTS software.

	Synthesis method	Operating System	Portuguese Language
Balabolka	Phonemes	Win, Linux, MacOS	Yes
Voice Dream Reader		Win, MacOS	Yes
Word Flash Reader		Win, Linux	No
Espeak	Formants	Win, Linux	Yes
Loquendo		Win, Linux, MacOS	Yes
Festvox	UniSyn / diphones	Linux, FreeBSD, Solaris	Develop. possible

An exploratory study on the usage of these software packages indicated that *Festvox* would probably be the best option in this context, due to its open architecture. It is designed to support independent development, making it possible to build new voices and libraries for different languages

(Black and Lenzo, 2014). Unfortunately, no libraries were yet available for Portuguese (the target language in this case). Developing a new library is a lengthy process, deemed unfeasible in the available period.

The software packages for which Portuguese language libraries were available are characterised by a closed architecture, which makes them unsuitable for research purposes. They seemed to be designed for relatively undemanding end-user applications: by and large, the available voices sounded quite artificial and prosody could not be altered.

In spite of their more convincing voices, even the top commercial packages lack prosody control to model emotion and excitation in the patient's speech. Moreover, libraries for different languages are sold separately, adding to their (already high) cost.

4 CASE STUDY

4.1 Selection Criteria

The development of a speech generation tool for Portuguese based on true synthesis would not fit the time frame of an MEng dissertation. The chosen alternative was the development of a proof of concept based on pre-recorded sentences. This required finding a particular case for which reasonably realistic 'dialogues' could be built from a small repository (otherwise, it too would be unfeasible) and, in spite of this limitation, retained relevance from the point of view of health care training. The choice fell on the evaluation of consciousness based on the Glasgow Coma Scale (GCS) – more specifically, its 'verbal response' parameter.

4.2 Glasgow Coma Scale

The GCS (Teasdale and Jennett, 2013) is widely used in the diagnosis and monitoring of neurologic dysfunction. It allows quantitative evaluation of consciousness levels (in a 3-15 point range) based on simple criteria (Santos et al., 2013), thus providing health care practitioners an efficient and objective means of reporting neurological status. The score is obtained by observation of the patient's behaviour, either spontaneous or in response to verbal and/or pain stimuli. Three evaluation parameters are considered: eye opening response (1-4 points), motor response (1-6 points) and verbal response (1-5 points). The focus was on this last one, as it involves dialogue with the patient. The following criteria apply:

- **Oriented Response** (consciousness level **S5):** patients oriented in space and time, who coherently answer questions like "What's your name?" or "What day is it today?", get top marks i.e. **5 points**.
- **Confused Response** (consciousness level **S4):** patients who are able to answer the same kind of questions, but do it in a confused way, get **4 points**.
- **Incoherent Response** (consciousness level **S3):** patients whose responses are not related to the questions (taking the previous example, that would be the case of answers like "I'm hungry" or "I want to go home") get **3 points**.
- **Incomprehensible Response** (consciousness level **S2):** patients who respond with moans or sounds not forming words, get **2 points**.
- **No Response** (consciousness level **S1):** patients who do not respond verbally to any stimuli get the lowest marks i.e. **1 point**.

4.3 Response Repository

Normally, evaluation of a patient takes only five or six questions. However, taking into account all the relevant variables regarding the patient (consciousness state, gender, age, emotional state…), the evaluator (trainee) and the circumstances of the interview (date, time…), an extremely large number of responses would be required.

Based on the long GCS training experience of one of the authors, a core set of questions was defined, along with a selection of responses covering the five consciousness levels, but considering a single type of patient (adult male) and allowing only a limited degree of variability for each level. This selection of responses (109 sentences in total) were recorded in a studio and post-processed for noise filtering and amplitude normalisation.

5 APPLICATION DEVELOPMENT

5.1 Requirements

The envisaged training setting is as described in section 2. There are no particular requirements regarding the mannequin other than being equipped

with a loudspeaker to simulate vocal communication. A low-fidelity model is perfectly suitable. The 'dialogue' with the trainee should be supported by a computer application managed at the control room. This should allow the instructor to pre-define training scenarios, namely by specifying the desired behaviour of the 'patient'. For flexibility, immediate intervention options should also be available to bypass pre-defined scenarios in response to unforeseen situations. The user interface should be graphical (GUI) and intuitive.

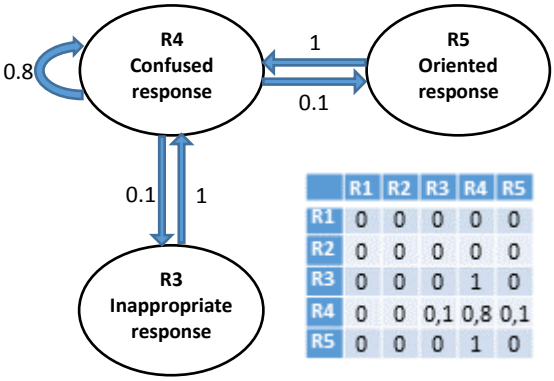
The trainee is expected to approach the mannequin (following all the protocols applicable to interaction with hospital patients) and carry out the questioning. For the reasons presented in 4.3, the set of questions must be strictly defined.

Response playback will be triggered upon question identification by the instructor. This should convey to the trainee the impression of being in conversation with a real patient. The responses, played back through the mannequin's loudspeaker, should be consistent with the scenario defined by the instructor and the corresponding questions.

The trainee is supposed to evaluate the responses obtained, record the corresponding GCS score and submit a diagnosis to the instructor. The application should assess the trainee's performance, provide feedback on it and assign a classification.

5.2 Response Generation Algorithm

In real situations, it is common to observe fluctuation in the apparent level of consciousness (normally between close GCS levels). Integrating that effect into the simulation, in addition to increasing its realism, is welcome from a training perspective, as the correct evaluation becomes variable from question to question. For that purpose, the response generation algorithm developed for the mannequin employed a probabilistic model based on *Markov chains* (Childers, 1997), as illustrated in Figure 3. The consciousness level fluctuation exhibited in the responses is governed by transition probability matrices, one per GCS level. These are configurable, allowing the instructor to define any pattern of response consciousness level, from strictly constant to completely random.



	R1	R2	R3	R4	R5
R1	0	0	0	0	0
R2	0	0	0	0	0
R3	0	0	0	1	0
R4	0	0	0,1	0,8	0,1
R5	0	0	0	1	0

Figure 3: Example of Markov chain response model for consciousness level S4 allowing sporadic responses at the levels immediately above and below, as defined by the transition probability matrix shown.

Note it would be relatively easy to build appropriate Markov chains to model more complex situations, such as recovery or slow loss of consciousness.

5.3 User Interface

The application was developed in *Visual Basic®*. As shown in Figure 4, its main window comprises three panels:

- **Definitions:** allows the instructor to enter the configurations regarding the trainee and the response generation mode, including the settings of the Markov chain models described in the previous point.
- **Interaction:** is used to manage mannequin responses through the identification of the questions posed by the trainee. In automatic mode, the algorithm chooses the response according to the consciousness level determined by the Markov chain model.
- **Evaluation:** supports the analysis of the diagnosis made by the trainee and works out her final marks as a percentage.

6 TESTING

Initial user tests took place at one of the laboratories of the University of Aveiro's clinical simulation centre (SIMULA), where a low-fidelity mannequin was equipped with a miniature amplifier-loudspeaker set. This was connected by *bluetooth* to the control room computer, in which the application presented in the previous section had been installed.

The tests involved seven nursing students acting as trainees and their teacher in the instructor role. The

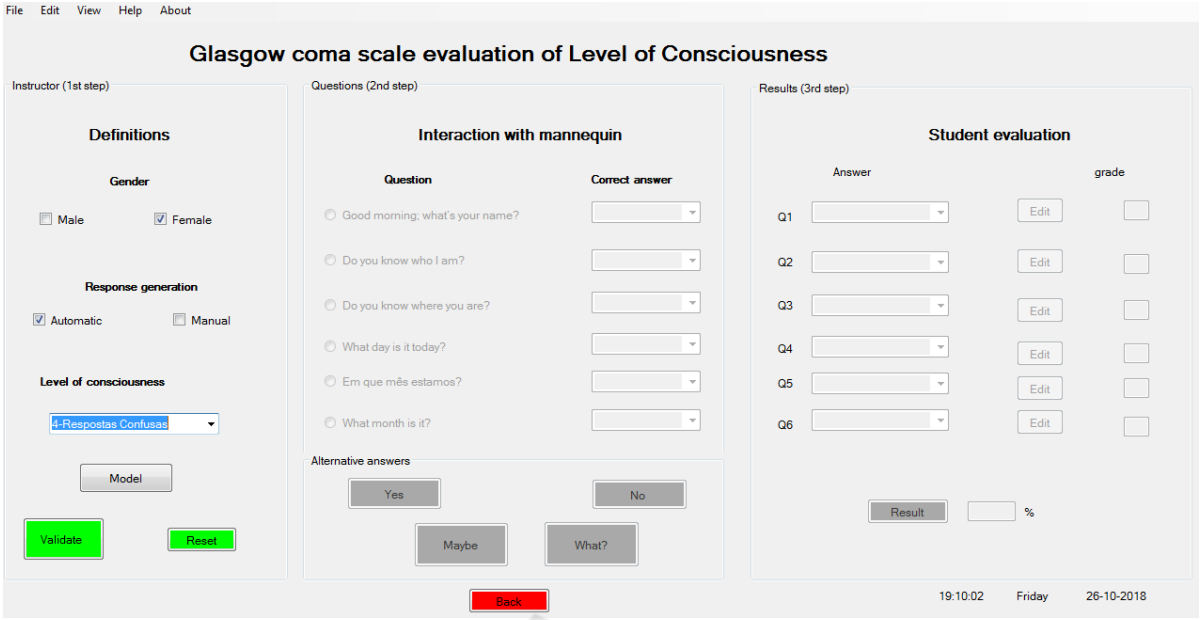


Figure 4: Application main page.

students carried out, and handed to the instructor, their GCS consciousness level evaluations based on a questionnaire for which a response repository had been previously recorded, as explained in section 4.3.

At the end of the test session, the students provided feedback by filling an evaluation form based on the System Usability Scale (SUS) proposed by John Brooke. They were asked to indicate, on a range from 1 ("I totally disagree") to 5 ("I totally agree"), to what extent they agreed with the following statements:

1. *This experience has stimulated my interest for clinical simulation.*
2. *The simulation made it easier to learn the Glasgow Coma Scale.*
3. *The simulation helped me develop my ability to assess consciousness level using the GCS.*
4. *It was a pleasant and valuable experience.*
5. *Simulation mannequins equipped with speech synthesis would be useful in several other situations in our training course.*

The resulting scores, shown in Table 2, are close (low standard deviation) and high (total agreement in most cases – averages between 4.57 and 5). It should be mentioned that the lowest scores came from a student who had missed the initial part of the session (all the others attended the whole session).

Table 2: Assessment of the application by the trainees.

	Trainee						
Statement	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1	5	5	5	4	5	5	3
2	5	5	5	5	5	5	5
3	5	5	5	5	5	5	4
4	5	5	5	4	5	5	5
5	5	5	5	5	5	5	5

7 DISCUSSION. FUTURE WORK

Although only small-scale initial tests could be carried out, their results support the position that incorporating vocal capabilities in simulation mannequins can bring significant benefits in health care training. The opinions and suggestions put forward by the trainees in the evaluation form (an area was reserved for that purpose) showed enthusiasm with the experience, and interest in repeating it with more time and a larger number of questions. One of the suggestions was to generate increasingly elaborate mannequin answers in the course of the questioning. Since the GCS evaluation protocol does not provide an opportunity for simulating longer, more demanding dialogues, this line of suggestions (which confirm the score obtained for statement 5) encourage the exploration of other application case studies. The instructor who took part in the tests specifically suggested the area of primary health care.

Speech synthesis would bring obvious advantages, since a virtually infinite set of dialogue simulation scenarios could be programmed. That is the main goal. However, given that prosody manipulation is the main difficulty of speech synthesis systems and clinical simulation applications are especially demanding from that point of view, this work suggests that, in some application niches, systems based on pre-recorded repositories may be interesting in their own right, to the extent that they provide more realistic audio (at least as regards prosody). This hypothetical interest of pre-recorded speech does not hinder the evolution to true synthesis – on the contrary, it may constitute additional motivation for it.

In the current operation mode, the instructor must identify the question, which demands constant attention to the dialogue between trainee and mannequin. It would be interesting to automate this process and dispense (at least to a certain extent) with instructor intervention. This would involve equipping the mannequin with a microphone and simulating hearing capability through the incorporation of a question identification system based on speech recognition.

ACKNOWLEDGMENTS

The authors wish to express their gratitude to António Veiga, from the Dept. of Communication and Arts (DeCA) of the University of Aveiro, for his kind assistance with the use of DeCA's recording studio to build the response repository and also to the ESSUA nursing students who agreed to take part in the user evaluation tests.

This publication was supported by National Funds assigned through the Foundation for Science and Technology (FCT) to IEETA, a research unit of the University of Aveiro, in the context of the project UID/CEC/00127/2019.

REFERENCES

Brazão, M. L. Nóbrega, S. Correia, J. P., Silva, A., S., Santos, D. and Monteiro, M. H. (2015). 'Simulação Clínica: Uma Forma de Inovar em Saúde (Clinical Simulation: A Way to Innovate in Health)', *Medicina Interna, vol. 22/3 Jul-Sept: 146-155*.

Taylor, P. (2009). Text-to-Speech Synthesis. *Cambridge University Press*.

UNIFAE - Centro Universitário das Faculdades Associadas de Ensino (n.d.). Laboratório de Simulação Avançada. Retrieved November 2017 from http://www.fae.br/portal/laboratorio-de-simulacao-avancada/

Barros, M.J.A.S. (2012). Estudo Comparativo e Técnicas de Geração de Sinal para a Síntese da Fala. MSc dissertation. *Faculty of Engineering of the U. of Porto*.

Freewarenee (n.d.). Sintetizadores de Fala Reconhecimento de Voz – Leitor de Texto. Retrieved August 2017 from http://freewarenee.weebly.com/sintetizadores-de-fala.html

Freewareneesite (n.d.). Software e Recursos Livres Necessidades Especiais – Sintetizador. Retrieved August 2017 from https:// freewareneesite.wordpress.com/sintetizador/

Manfio, E.R. (2012). 'Como funcionam alguns fonemas no aplicativo Balabolka (How some phonemes in application Balabolka works)', *Via Litterae, vol 4/2 Jul/Dec: 191-204*.

Araújo, A.L.S.O. and Silva, J.S.O. (2013). 'Educação e tecnologia: alternativas de aplicativos facilitadores à expressão oral para portadores de necessidades especiais', in *5º Simpósio Hipertexto e Tecnologias na Educação*. Retrieved November 2018 from http://www.nehte.com.br/simposio/anais/Anais-Hipertexto-2013

Black, A. W. and Lenzo, K. A. (2014). Building Synthetic Voices (For FestVox 2.7 Edition). Retrieved November 2018 from http://www.festvox.org/bsv/bsv.pdf

Teasdale, G. and Jennett, B. (1974). 'Assessment of coma and impaired consciousness. *A practical scale'. Lancet, Jul 13;2(7872):81-4*

Santos, W. C., Vancini-Campanharo, C. R., Lopes, M.C.B.T., Okuno, M.F.P. and Batista, R.E.A. (2016). 'Assessment of nurse's knowledge about Glasgow coma scale at a university hospital'. *Einstein (São Paulo), vol 14/2, Apr-Jun: 213–218*.

Childers, D. G. (1997). Probability and Random Processes: Using Matlab with Applications to Continuous and Discrete Time Systems. *Florida: Richard D Irwin*.