

Assessing the impact of a Clinical Audiology Simulator on first year students

Alexandre Heitz¹

Andreas Dünser¹

Christoph Bartneck¹

Jonathan Grady²

Catherine Moran²

¹ HIT Lab NZ, University of canterbury, Christchurch

Tel.: +64 (0)3 364 2349 Fax: +64 (0)3 364 2095

Email: alexandre.heitz@canterbury.ac.nz

² Department of Communication Disorders, University of Canterbury, Christchurch

Tel.: +64 (0)3 364 2431 Fax: +64 (0)3 364 2760

Email: sue.bradley@canterbury.ac.nz

Abstract

Virtual Patients (VPs) have been successful in health education to promote and foster communication. Additionally, computer simulations offer the advantage of being standardized, repeatable, and do not require as much resources as role-play simulations. The research presented in this paper offers to explore the impact of a Clinical Audiology Simulator (CAS) using virtual patients technology on first year audiology students of the University of Canterbury. We look at the CAS's effects on students' perceived level of learning, confidence, and ability to conduct adequate Pure tone audiometry as well as Clinical history taking procedures. These studies showed positive results on students' perceived level of learning, and history taking skills when using the CAS as an additional training tool. We present the findings and lessons learned from these studies as well as our plans for future experiments and software implementations. VPs have the potential to offer audiology trainees more opportunities to practice and access to a wider range of pathologies as they would with their course's traditional practical sessions.

Keywords: virtual patient, computer simulation, audiology, clinical education

1 Introduction

In healthcare education, actors trained to interact with students in role-play settings are often used as Simulated Patients (SPs). However, access to SPs for healthcare education is limited due to their cost and availability. SP training is expensive. Despite the 130 medical schools in the U.S. only 5 provide simulated patients assessment. Additionally, SPs work for a low incentive, which can be around 10 USD per hour. Considering this, SPs use in clinical scenarios to produce valid interactions for clinicians and students training can be limited (Rizzo et al. 2010).

Another training method in healthcare is Virtual Patients (VPs). VPs have the potential to improve clinical practice of healthcare trainees.

Copyright 2014, Australian Computer Society, Inc. This paper appeared at the Fifteenth Australasian User Interface Conference (AUIC 2014), Auckland, New Zealand. Conferences in Research and Practice in Information Technology (CRPIT), Vol. 150. Burkhard Wünsche and Stefan Marks, Eds. Reproduction for academic, not-for-profit purposes permitted provided this text is included.

They are computer generated patients that allow realistic training in controlled settings. One reality of healthcare practical education is that students can spend days doing clinical work to only see repetitive cases, when patients come in with the same pathologies. VP simulations can be used to provide a varied and standardized training among students (Collins & Harden 1998, Johnsen et al. 2005, Triola et al. 2006), assuring every one of them has the chance to train for low frequency events too, not just the cases that are the most common. However, at the same time students can choose to interact with the exact same VP again if, for instance, a debriefing made them realize they missed critical pieces of information. In addition, VPs also enable students to practice within safe boundaries and prepare them to interact with real patients where mistakes would be costly, as well as develop clinical reasoning skills (Round et al. 2009).

Previous research shows that VPs can effectively be implemented as an alternative to SPs in various types of scenarios to represent illnesses ranging from a standard medical exam, to eye or breast exams (Deladisma et al. 2009, Kotranza et al. 2009, Rizzo et al. 2009), cases for paramedic students (Conradi et al. 2009), continuity of care for pharmacy students (Fuhrman Jr et al. 2001), even geriatrics (Tan et al. 2010), surgery cases (Vash et al. 2007), or other psychiatric disorders such as disorderly conduct, post traumatic stress syndrome (PTSD), or other phobias (Gorrindo & Groves 2009, Kenny et al. 2008, Rizzo et al. 2010, Triola et al. 2006). This technology can also enable students to practice from their own homes (Stanton 2008). VPs have also been successfully used with patients with inadequate health literacy to explain medical documents (Bickmore et al. 2009, 2010), or to promote healthy behaviours such as adherence to medication and exercise (Bickmore & Picard 2005, Hayes-roth & Saker 2003, Ruttkay & Welbergen 2008).

On the other hand, VPs' main limitation, when compared to SPs, is frequently their vocabulary. The process of talking with a patient is often script based. Developing an interaction script for one VP is a time consuming process and often needs to be strengthened following pilot studies to make it usable. Work has been done on implementing ways to create more robust scripts efficiently for VP interactions but it is still an area that needs to be improved (Halan et al. 2010, Rizzo et al. 2010, Rossen et al. 2010). Consequently, there is little research on implementing a simulation platform using a multitude of VPs for history taking training to support students' practical experience. History taking is the procedure of interviewing

a patient and recording past events or circumstances that could be relevant to the patients' Health state.

For the purpose of our research, we implemented a computer simulation platform aimed for trainee audiologists (more information on this sample of students in the studies' sections). We propose to investigate how using this VP based simulator as an additional tool to traditional means of learning can affect students' performance in their course, and during role-play situations using SPs. We will also investigate how the VPs simulator we implemented affects students' perceived level of learning and confidence when used in addition to regular teaching.

This paper is going to present two studies. The first is a pilot study that took place during the first part of the course year and focuses on the pure tone audiometry procedure. The second study investigates clinical history taking and spreads over the second part of the year.

2 Related work

VPs have been used in a wide range of specialities to train clinical reasoning and patients' assessment in different settings. As part of a collaboration between London's universities, the Second Life virtual environment has been used as a way to generate VPs to train paramedics and allow them to explore more open ended questions in their decision making process. The researchers concluded that VPs in a virtual environment can offer realism impossible to reach in a classroom environment, allowing them to experience the consequences of the different choices they make (Conradi et al. 2009). At the University of Munich in Germany, a VP based systems allowed students to train on gynaecological sonography. The conclusions showed that training with virtual patients seems comparable to live practice with the advantage of a standardized consistent output which is not the case of real patients (Heer et al. 2004). The research from the following groups is considered the most relevant for our own, as the technologies they use are of similar level to ours, and consequently our terminology matches as well.

The University of Florida's Virtual Experiences Research Group provided our research team with the basis of our simulation platform. The group's early studies concerned the possibility of creating a system with patient-doctor interactions, focused on assessing whether current technology enabled to simulate such experiences, with sufficient immersion and fidelity. The Virtual Experiences Research group focused on implementing medical history interviews for different disciplines as well as investigating different ways to improve user-VP exchanges. Studies exploring different ways to affect interactions have also been conducted, using life size VPs thanks to projectors for instance, investigating natural interactions such as hand gesture (both real hand through infra red cameras, or virtual hands), tablet PC and audio (Ferdig et al. 2007, Kotranza et al. 2009, Johnsen et al. 2005). Real size displays, using projectors were also found to allow users to display more empathy than head mounted displays (Johnsen & Lok 2008). Consequences of the use of synthesized speech and natural recorded speech for the VPs answer have also been investigated and compared. It was concluded that if the intent was to teach what questions to ask, then both methods were as effective. However, if the purpose was to teach how to ask questions, the realism of pre-recorded speech could make up for its low flexibility (Dickerson et al. 2006).

Rather than assessing its continuous use by medical trainees, their evaluations often focused on comparing the quality of the VP to an SP using the Maas-trich assessment of Simulated Patients questionnaire, student posture, tone of voice and speech content for different exams such as a breast, eye exams or pharmaceutical exams (Deladisma et al. 2009, Kotranza et al. 2009, Rizzo et al. 2009), and studies investigating how racial and social disparities affect participants relation with VP. In the different conditions explored, research showed significant results that participants were able to act with VPs, to a similar or satisfactory level compared with simulated patients (Ferdig et al. 2007, Johnsen et al. 2005).

Another research group that researches VPs is the institute for creative technologies of the University of South California. This group primarily investigates the use of VPs with military technicians and clinicians to treat soldiers, which can be dangerous both mentally and physically. VPs are considered for PTSD patients' suicide tendencies, or even traumatic brain injuries. Significant results were found and participants were able to detect trauma, its duration, or presenting its origin. However, symptoms requiring a deep level of exchanges with the VPs to identify them were almost not detected due to the state of the scripts implemented, and current performances of speech recognition (Kenny & Parsons 2010, Kenny et al. 2008). VPs were also used as an online guide to promote access to psychological healthcare information and assist as well as encourage military personal and their family to seek care if necessary.

Looking at this researches, the main downside of the VPs-user interactions appears to be that their interaction are scripted, thus a necessary time consuming step is to implement a script with sets of clues or triggers for each of the VP's answers. This is the main limitation VPs currently suffer from and that can prevent them from being as realistic as SPs. In addition, an underdeveloped script can be frustrating for participants (Johnsen et al. 2005). Development of an interaction script for one VP is a time consuming process and often needs to be strengthen following pilot studies to make it usable. Work has been done on implementing ways to create more robust scripts efficiently for VP interactions but it is still an area that needs to be improved (Halan et al. 2010, Rizzo et al. 2010, Rossen et al. 2010). As a consequence, VP simulations used in health education are commonly focusing on a single case or VP (Bickmore & Picard 2005, Deladisma et al. 2009, Fuhrman Jr et al. 2001, Gorrindo & Groves 2009, Hayes-roth & Saker 2003, Heer et al. 2004, Kenny et al. 2008, Kotranza et al. 2009, Rizzo et al. 2010, 2009, Ruttkay & Welbergen 2008, Stanton 2008, Tan et al. 2010, Triola et al. 2006, Vash et al. 2007).

Our research is focusing on the field of Audiology. In audiology, simulations have been used with success to teach clinical skills and simulators allowing practice of procedural skills such as pure tone audiometry are available on the market (e.g. Otis Audiology simulator, Parrot Software's Audiology Clinic) for Universities to use to supplement traditional course work. There is, however, a lack of simulators incorporating such procedural skills with clinical history taking training possibilities and immersing students in realistic experiences from meeting a patient up to, and including pathology assessments. Our research aims to explore the use of a computer simulation based on a varied range of VPs, with audiology trainees as our primary target audience. We aim to answer whether this system, used as an additional training tool during the curriculum, can positively impact students' clini-

cal skills, perceived level of learning, and confidence. We reviewed students' grades as part of the Clinical Audiology Observation and Practice course to assess the impact of the pure tone audiometry component of the simulator. In addition, students were also assessed with role-play situation using SPs to evaluate their ability to conduct clinical interview and pure tone audiometry procedures following exposure to our computer simulation.

3 The Clinical Audiology Simulator

3.1 System architecture

The computer simulation used in this study is referred to as the Clinical Audiology Simulator (CAS). The CAS is used to practice procedural skills, clinical history taking and puretone audiometry, as well as decision making. This takes the form of the standard audiology range of tests including history taking, pure tone audiometry, otoscopy, and pathology diagnosis for Virtual Patients (VPs). The CAS is based on a simulation platform originally implemented by the University of Florida's Virtual Experiences Research Group (VERG 2005). The initial platform had been implemented for research study purposes and was adapted to run as a standalone application. The CAS has been implemented in C#/C++ using Visual studio 2010. The application makes use of the open source 3D library Ogre for graphics, to render the VPs as well as the room where the consultations take place within an embedded window (see Fig. 1).

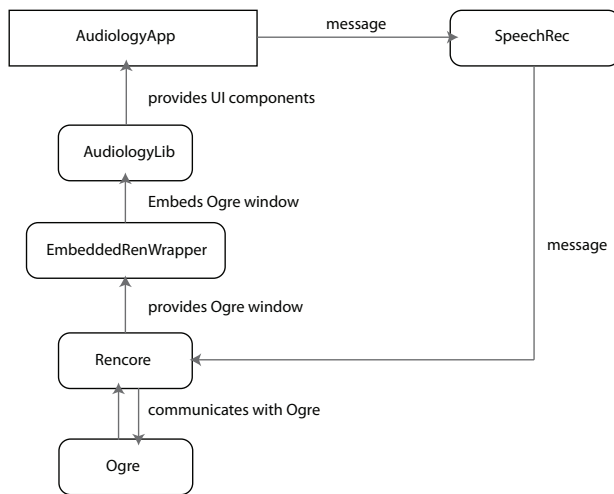


Figure 1: CAS main system architecture

3.2 Using the Clinical Audiology Simulator

When launching the CAS, a student starts with selecting one of the VPs among the different cases offered (refer to Fig. 2 for this section). The application then starts and students have access to the different features of the software, labeled 'Interview', 'Otoscopy', 'Tone test', and 'Submit Results'. The students can choose the order of the procedures but would typically start by interviewing the patient.

The interview component of the consultation takes the form of a loop where the students ask questions to the VPs which will be answered according to the VPs' scripts. The students interact with the VP, who is sitting in a room, by typing the questions he/she wishes

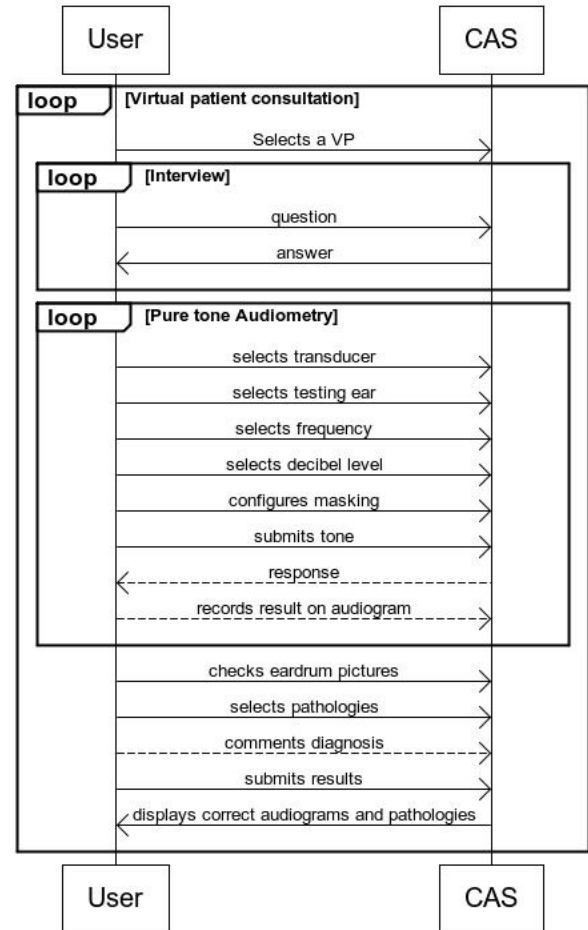


Figure 2: Using the CAS sequence diagram

to ask (see Fig. 3 Interview interface of the CAS). The VP will either answer the students' question, if it was understood, or hint the students to reformulate and/or ask another question (e.g. "Could you please rephrase that?"). The students' aim here is to collect and sort from the VPs' answers the necessary information to complete a Diagnostic Adult History Form that is commonly used in clinic when interviewing patients. The students can stop this process at any time if they consider that all the relevant information got retrieved from the VP.

A student can then check the VPs' ears if he considers it necessary. This will display two eardrum pictures from our collection of pictures retrieved on real patients, which can present additional elements to help identify the appropriate diagnosis.

The student would then follow with the Pure tone Audiometry procedure, which takes the form of another loop. This procedure is about determining the VPs' hearing thresholds. Hearing thresholds represent a patient's hearing levels, which is his ability to hear sounds properly across the range of frequencies the human ear should be able to detect. Hearing thresholds are typically displayed on audiograms, which are graphs of the hearing level in Decibel for a set of frequencies ranging from 250 Hz to 8 kHz. To determine hearing thresholds in the CAS, a student will have to first select a transducer between supra-aural, insert, and bone before deciding the ear to be tested.

This represents the type of headphone used, each having its specificities. Then, the procedure in-

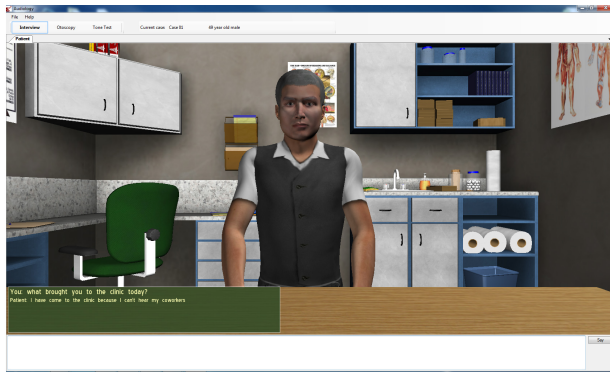


Figure 3: Interview interface of the CAS

involves submitting different intensities of sound over the tested frequencies until the VPs are able to hear them, if at all. When a tone is submitted, the VPs can either answer, or not. Once a VP answers to a particular tone the student should mark the response level on the virtual audiograms. Masking is configured for the non-test ear depending on the patient's responses, if the student considers there could have been conduction from one ear to another during tests. This process is repeated for the whole range of frequencies the students decides to test, for both ears. Fig. 4 represents the interface the students face when conducting a pure tone audiometry procedure on the CAS.

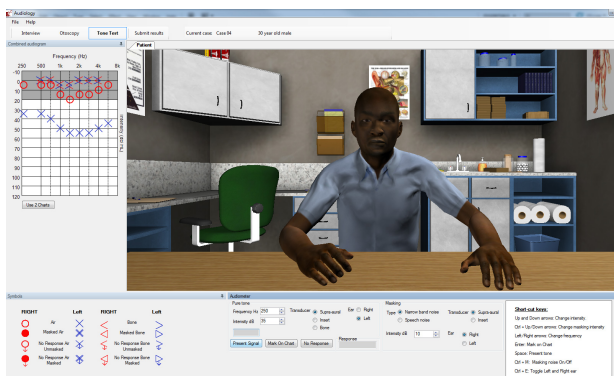


Figure 4: Tone Test interface of the CAS

Following those assessments, the student has to submit his/her results. The student has to determine the pathology(ies) associated with this patient. In addition, a student can choose to add a comment on his diagnosis decision. This information can be recorded and used for assessment. Finally, once the diagnosis is submitted the student will be given feedback in the form of the VP's actual audiogram, and its correct diagnosis.

The CAS was deployed on a total of seven computers to allow participants to practice during their free time. During their designated training period, participants were able to access the computers and practice on the CAS at any time of the week.

4 Evaluation

To fit the participants' curriculum in the Clinical Observation and practice course the investigation was split into two. First is the Pure tone audiometry pi-

lot study. This pilot is followed by the Clinical history taking study, which is our main evaluation.

4.1 Pure Tone audiometry pilot study

The Pure tone audiometry pilot study focuses on learning the procedure and reasoning behind determining patients' hearing threshold. This procedure is typically taught to students in the first part of the year. Our pilot took place during that period of the students' training. This pilot helped us to test the research setup, as well as measure the quality of our evaluation. The main hypotheses were:

1. Students' ability to conduct a pure tone audiometry procedure accurately will increase as a result of using the CAS in addition to traditional methods.
 - (a) Students' Perceived level of learning will improve as a result of using the CAS to practice pure tone audiometry in addition to traditional methods.
 - (b) Students' grades when assessed in Role-play, and on the simulator will increase as a result of using CAS in addition to traditional methods.

4.1.1 Participants

The University of Canterbury's Master of Audiology Degree (Maud) is spread over two years, with the bulk of theoretical teachings being focused in the first year. Entry into the Maud is very restricted and competitive and only ten to twelve students per year are accepted into the program. The primary reason for this is that audiologists' training, just like any healthcare professional, is resource intensive, particularly in term of practical training. However, within the audiology field in a small country such as New Zealand, training opportunities are even more limited. According to audiology teachers, the first year students are the ones benefiting the most from extra training opportunities as they are relatively new to the field, starting with virtually no previous domain knowledge. It would also prime them on having the set required skills to make the best of their summer placements.

A group of twelve students was recruited, ten females and two males. All participants achieved at least a Bachelor degree in tertiary education. Each student was enrolled in the Clinical Observation and Practice course which includes one day a week of observation and practice at various clinics in the Christchurch area.

Nine students reported English as being their native speaking language, two reported German, and one reported Chinese. All students are fluent in English. The twelve students declared having adequate vision and hearing to use the simulator, as well as conduct and record a medical history from a patient, as well as having adequate hearing. Five students rated their computer skills as *Above average*, six as *Average*, and one as *Below average*. However, the twelve students answered they considered having the necessary level of skills required to operate the Clinical Audiology Simulator (CAS).

4.1.2 Design

The Pure tone audiometry pilot study took place in the first semester. Participants were split into two groups, Group A and Group B. Each group was

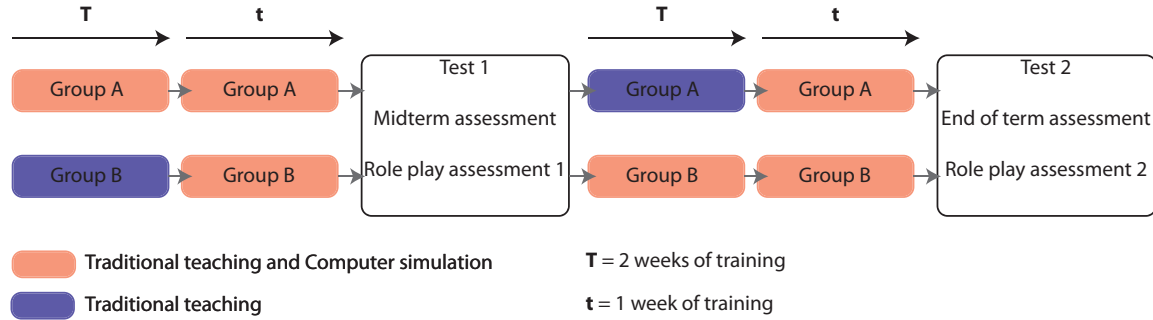


Figure 5: Pure tone audiometry study design

made of five female participants and one male participant, while the mean gpa (grade point average) of the two groups was counterbalanced (Group A *mean gpa*=7.48, *sd*=0.97, and Group B *mean gpa*=7.46, *sd*=0.87).

Following a presentation of the CAS as part of a tutorial session, participants in Group A got access to the simulator in addition to their in-class teaching for a period of two weeks. During that time, Group B was only training using traditional means of learning. After this first training phase, both groups had access to the simulator for a week in order to prepare for 'Test 1', their midterm assessment and extra role-play assessment. In addition, participants answered a questionnaire on their experience with the CAS. After these assessments, the conditions were reversed and Group B had access to the CAS for a period of two weeks in addition to their courses while Group A was only following traditional learning methods. After these two weeks, both groups had again access to the simulator in order to prepare for their end of term assessment. This assessment also got followed by a second role play assessment, and another CAS questionnaire (Fig. 5).

4.1.3 Measures

Midterm and end of term assessment grades were used as a measure for this experiment. For both these assessments, participants had to determine and report on paper the hearing thresholds of three different patients using the computer simulator. Marks were allocated depending on how close the hearing thresholds reported were to the actual threshold of the patients implemented.

Role-play assessments took place in the audiology clinic located on campus where research assistants played the role of Simulated Patients (SP) in realistic clinical settings. Participants were asked to perform a pure tone procedure on the SP and report their results. Marks were allocated to participants following typical clinical work marking criteria where in addition to actual threshold validity, points were also granted for method, pace, confidence, and explaining results.

A questionnaire was given to the participants following the role play situations. The main measure of interest was the mean of five 5 points likert-scale questions referring to participants' perceived level of learning as a result of using the CAS (Fig. 6).

4.1.4 Results

We used the Kolmogorov-Smirnov test to attest the normality of the data. For the perceived level of learn-

Name: _____ Student Questionnaire | 1

Consider your interactions with the computer simulation. Please indicate to what extent you agree/disagree with the following statements:

1 – Strongly Disagree, 2 – Disagree, 3 – Unsure, 4 – Agree, 5 – Strongly Agree

1. My ability to obtain accurate unmasked pure tone thresholds from a client has improved as a result of working with the computer simulation.

1	2	3	4	5
---	---	---	---	---
2. My ability to obtain accurate masked pure tone thresholds from a client has improved as a result of working with the computer simulation.

1	2	3	4	5
---	---	---	---	---
3. My ability to confidently recognise and explain the type and degree of hearing loss from an audiogram has improved as a result of working with the computer simulation.

1	2	3	4	5
---	---	---	---	---
4. My confidence to perform audiometric testing on real clients in the future has increased as a result of working with the computer simulation.

1	2	3	4	5
---	---	---	---	---
5. I learned a new skill as a result of working with the computer simulation.

1	2	3	4	5
---	---	---	---	---

Figure 6: Extract of Pure tone pilot questionnaire

ing scores, the results were non significant for the two groups at Time 1, and at Time 2, which shows normality. For the learning outcomes, similarly, the results of the test show that the data assumes normality for the two learning outcomes (role play grade and simulator grade), for both groups, at Time 1 and at Time 2.

For the perceived level of learning scores, we used a nonparametric independent sample *u*-test. This analysis showed significant differences for perceived level of learning between the two groups at Time 1 ($U=1.50$, $p=.007$), but not at Time 2 ($U=15.50$, $p=.684$).

For the learning outcomes, we used t-tests to analyse the differences in role-play assessments and simulator assessments between the two groups at Time 1, at Time 2, and on the difference off scores between Time 2 and Time 1. No statistical differences were found between the two groups for these measures.

The means and standard deviations for these variables are presented in Table 1.

4.1.5 Discussion

Preliminary data was analyzed to explore if the additional exposure to our VP based simulator could increase audiology students' perceived level of learning and performance. Students were assessed using both the CAS and with a SP based role-play sce-

Table 1: Descriptive statistics for perceived level of learning, role play assessment grades, and simulator assessment grades at Time 1, and at Time 2

	Group of the student	Time 1		Time 2	
		Mean	Std. dev.	Mean	Std. dev.
Perceived level of learning	Group A	3.57	0.50	3.70	0.55
	Group B	1.80	0.90	3.60	0.49
	Total	2.68	1.16	3.65	0.50
Role play grade	Group A	74.31	6.32	78.94	17.33
	Group B	68.29	14.55	80.56	9.42
	Total	71.30	11.15	79.75	13.33
Simulator grade	Group A	82.22	11.26	88.62	7.31
	Group B	84.92	7.08	90.71	6.02
	Total	83.57	9.08	89.66	6.47

nario. First, as expected, the students did improve their score independently of their learning group on perceived level of learning, Simulator assessment, and Role-play assessment over the course of the experiment.

At the start of the experiment we hypothesized that students' perceived level of learning would increase following extended exposure to the CAS. The results support our hypothesis as Group A showed at Time 1 significantly higher perceived level of learning, following their dedicated period of exposure to the CAS. Following group B's exposure, this difference is not significant anymore which could indicate that they caught up with their peers (Fig. 7). Statistical equivalence has however not been calculated.

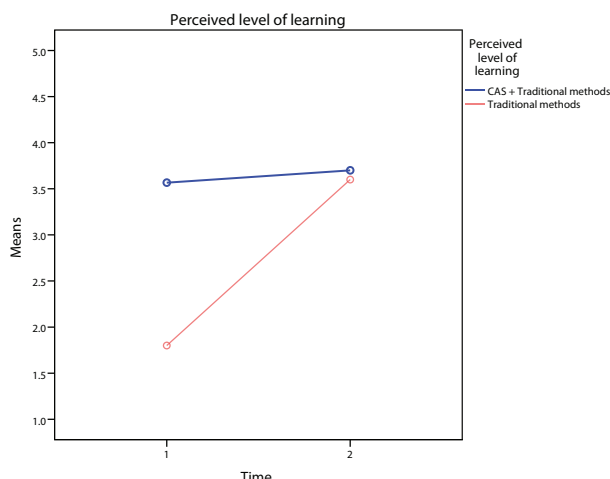


Figure 7: Perceived level of learning scores in function of time

The second hypothesis was that students would get better grades when assessed following their exposure to the CAS. The results, however, show that while there is an increase in students grades when assessed over time, for both types of assessments, both groups score increase similarly.

4.1.6 Limitations

The main limitation of this study is that both groups received one week of training with the CAS right before each assessments. This was one of the restriction of testing students as part of the Clinical Audiology course, and one of the course requirements; every student had to have access to the CAS before those course assessments. However, this allowed us to use

both students' midterm and end of term assessments as measures.

This week of common training, however, seemed to have leveled potential differences between the two groups. A few students reported during informal discussions that they got the most of their training done during the week prior to each assessment, whether in group A, or group B. Additionally, it is possible that students generally spent more time training on the CAS to prepare for the end of term assessments than for midterm's. We concluded that there are two ways to remedy this problem.

First is to remove the common week of training both groups had. Second, is to designate specific training times for students in both groups, and have them book specific time slots to ensure that the student group exposed to the CAS receives a sufficient amount of additional training, and control CAS exposure times between the two groups. This would also mean to abandon mid-term assessments and end of term assessment as measures. Indeed as mentioned earlier, to guarantee fairness for the students, they should all have access to the CAS before any of their course assessments. This means that the study needs to run before their midterm, while leaving enough time for the control group to practice as well following the study, in preparation of their exam.

In addition, it seems that an increase in students' perceived level of learning does not correlate with an increase in results. This measure will be removed as well from the following studies which will focus on the transfer of skill assessed with role-play situations.

4.2 Clinical history taking study

This study focused on assessing students on one of the main clinical exams our system supports, clinical history taking. Clinical history taking is centred around learning how to interview patients to retrieve key information that help determining an accurate diagnosis and follow up procedures. Assessments are conducted using role-play with an SP in order to test for transfer of skills from use of the CAS. While SPs are not real patients they allow for a simulated experience which is the closest to real practice. In addition, SPs enable to standardize the assessment, presenting a similar patient to each participant.

This experiment was conducted using the same prototype of the CAS. While still testing the functionality of our system, this experiment also aimed to answer the following hypotheses:

1. Students' confidence when conducting a clinical history interview will improve as a result of using the CAS, compared to students who were only given traditional instruction.
2. Students' ability to retrieve information in clinical history situation will improve as a result of using the CAS, compared to students who were only given traditional instruction.
3. Students' ability to accurately report information in clinical history taking will improve following exposure to the CAS, compared to students who were only given traditional instruction.
4. As a result of using the CAS in addition to traditional instruction, students will take clinical history more efficiently and will require less interaction with patients to retrieve relevant information, when compared to students who were only given traditional instruction.

4.2.1 Participants

The same participants who took part in the pure tone audiometry pilot study were recruited for this study.

4.2.2 Design

Participants for this study were split into two equal groups according to the results of a pre test role play situation. Following discussions with the course coordinator and other audiology expert, Students' grade point average had been determined to not be a valid representation of clinical interview taking abilities, thus could not be considered as a means to group students for this study. Indeed, clinical history taking skills somehow differ from other procedures as students need to ask questions accordingly, all the while interpreting each of the patients' answers to contribute towards a diagnosis.

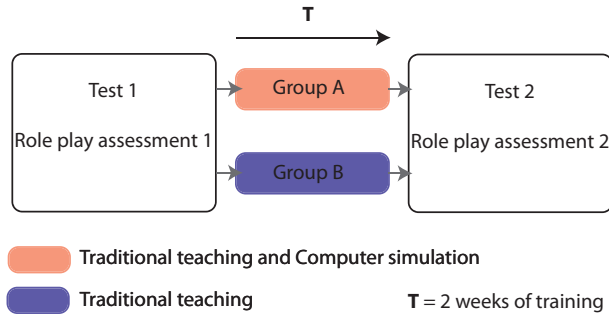


Figure 8: Clinical history taking study design

Following the pre-test (Test 1), participants were split into two groups, CAS and no CAS balanced based on the accuracy, confidence, and efficiency results of this test (see the following section for more details on these measures). Students in the CAS group then practised for a period of two weeks in addition to traditional teaching while students in the no CAS group only followed the traditional instruction. Students practising on the CAS agreed to train at least two hours. At the completion of this training phase both groups undertook a second role play assessment (Test 2), with a similar assessment method to Test 1 (see Fig. 8).

4.2.3 Measures

Interview taking skills are not typically graded as part of the first year audiology students' curriculum. Our main focus was to measure students' abilities during role play situations. While quantifying the improvement of the students is outside the scope of this paper, we assess whether the CAS had a positive impact on students interview taking skills. The role play simulations took place in a realistic setting, with two audiology experts acting as SPs; one for assessment 1, the other for assessment 2. During those experiences students were assessed on accuracy, where students had to retrieve information from the VPs and translate them onto standard history sheets, confidence and efficiency. Assessments were undertaken by two audiology experts to ensure reliability.

Verbal Accuracy: The verbal accuracy score is based on the number of answers participants were able to retrieve from the SP while talking during the interview. This measure was assessed using a transcript of the interaction between participants and

Table 2: Descriptive statistics for perceived level of learning, role play assessment grades, and simulator assessment grades at Time 1, and at Time 2

	Group of the student	Time 1		Time 2	
		Mean	Std. dev.	Mean	Std. dev.
Confidence	Group A	3.78	0.73	4.86	0.77
	Group B	3.63	0.99	4.33	1.25
	Total	3.71	0.81	4.62	1.00
Verbal accuracy	Group A	42.64	8.91	72.48	8.75
	Group B	42.33	8.92	60.93	7.05
	Total	42.49	8.46	67.23	9.73
Written accuracy	Group A	65.65	16.85	81.09	13.17
	Group B	59.03	21.42	76.60	14.86
	Total	62.64	18.37	79.05	13.44
Nbr of questions	Group A	24.67	8.04	44.50	6.09
	Group B	22.40	6.35	38.00	6.52
	Total	23.64	7.06	41.55	6.86
Efficiency	Group A	1.33	0.21	1.43	0.14
	Group B	1.24	0.30	1.45	0.13
	Total	1.29	0.25	1.44	0.13

SPs. Transcripts were recorded during student-SP interactions by one of the audiology experts. It is presented as a percentage.

Written Accuracy: The written accuracy score points to the number of adequately reported pieces of information on a history diagnosis form used in clinics to report information critical to diagnose the patients. Students were asked to fill in the diagnosis form while interacting with the SP, as they do during clinical exams. Clinical forms were then checked by both audiology experts and marked. This measure is presented as a percentage.

Confidence: For each role play situation, participants rated pre and post simulation confidence in their performance using three seven point Likert scale items.

Number of questions: For each role play situation, this measure is the number of questions a student asked the SP before considering the clinical history complete.

Efficiency: The efficiency score is calculated by using the verbal accuracy raw scores divided by the number of questions asked to an SP.

4.2.4 Results

We used a multivariate analysis of variance (MANOVA) to analyze the results of the clinical history taking study. Group was used as a between factor and the Time 2 confidence, verbal accuracy, written accuracy, number of questions asked (to SP), and efficiency measures were used as dependent variables. Time 1 confidence, verbal accuracy, written accuracy, number of questions asked (to SP), and efficiency measures were used as covariates. The means and standard deviation for these variables are presented in Table 2.

The MANOVA shows that there were no significant differences between students in the CAS and no CAS group at Time 2 for any of the dependent variables tested.

Students who received additional training with the CAS had a somewhat higher increase in verbal accuracy between Time 1 ($m=42.64$, $sd=8.91$) and Time 2 ($m=72.48$, $sd=8.75$), compared to students following only traditional instruction, Time 1 ($m=42.33$, $sd=8.92$); Time 2 ($m=60.93$, $sd=7.05$). However this difference is not significant.

4.2.5 Discussion

This study tested the first prototype of the CAS and explored how the additional use of the CAS to traditional methods of teaching impacts on students' confidence, accuracy and efficiency when practising clinical history taking.

We hypothesized that students' confidence level in their performance would increase following a dedicated training period with the CAS as an additional practice tool. The results show that both groups had a similar increase in confidence. We could find no evidence that practicing with the CAS increases confidence levels.

For accuracy measures, we hypothesized that both verbal and written accuracy would increase for students in the CAS group. Results show that verbal accuracy, which represents the number of answers students were able to retrieve from the SP, increased for all students and no significant differences between groups were found. Also for written accuracy, which is the percentage of adequately reported information from the verbal accuracy scores, we could not find significant differences between groups. Students seem to perform better than in the initial assessment whether or not they had access to the simulator.

Our final hypothesis was that efficiency would improve for students who used the CAS in addition to traditional methods. We characterize efficiency with two measures: the number of questions students ask the SP in role play, and the relation between the number of questions asked and answers retrieved from the VP. The results did not show a significant effect of CAS training on the number of questions asked to SPs in role play. It seems that over time, students generally started by asking only a small amount of questions to the SP during the assessments, then experimented with asking the SP more questions during the second assessment. Our main efficiency measure, the number of questions asked divided by verbal accuracy, however, decreased between the first and the second assessment. This can be attributed to students experimenting with the number of questions they could ask SPs, as mentioned above. No significant differences between groups could be found to support our hypothesis.

Limitations

The main limitation concerning the clinical history study is the VPs' scripting itself, which was still early stage in this project. While students did have positive results using the VPs, some reported having difficulties in retrieving specific information from them. This was due to the way they asked questions to the VPs, that was not adequately recognized by the system. We believe that by improving patients' interaction-scripts, we could obtain more positive results in further studies. It is also important to mention that scripts' implementation and robustness are a crucial element in VP based systems used in those studies (Halan et al. 2010, Rizzo et al. 2010, Rossen et al. 2010). Despite not having captured data on students' frustration, it is clear that a student is unlikely to spend a large amount of time practising if he finds the task daunting due to low responsiveness from the VPs.

Another method that could be followed in further studies could be the approach designed by Rossen et al. (2010), which presents a system aiming to improve virtual agents scripts implementation. It starts by recording interactions between multiple humans taking turn interacting with a VP, before formulating an

aggregate encompassing all the previous interactions within a single script. This process should be iterative, and undertaken alongside field experts (Rizzo et al. 2009). It is also a process that takes considerable time commitment in order to capture the wide range of conversational stimuli available (Halan et al. 2010).

We suggest that interaction scripts should start being tested as early on as possible when implementing a VP simulation system, to allow enough time for sufficient refinement.

A further limitation of this study was the lack of control over students' commitment to training using the CAS, and the size of the sample. The Master of Audiology is already an intensive degree and asking students to train in their free time, while dealing with an already full schedule and assessments due for the different courses they were enrolled in proved difficult. Students admitted after the study ended, during an informal meeting, that they did not feel as though they had enough practice time on the CAS. This could explain the lack of significant differences between groups for this study. Another concern was the limited number of participants available due to the small number of students enrolled in this degree.

5 Conclusion

Our current studies have found no clear statistical evidence on the effectiveness of using a VP simulator as an additional training tool for audiology trainees. However, the data suggests that there was some improvement in some measures after using the CAS. Those measures were confidence, perceived level of learning, and ability to retrieve information from SPs in a clinical role situation. Additionally, we identified areas that would benefit from being explored more deeply.

One of the main limitations of this study was the lack of control over students' commitment to training using the CAS. The Master of Audiology is already an intensive degree and asking students to train in their free time, while dealing with an already full schedule and assessments due for the different courses they were enrolled in proved difficult. Students admitted during our final group reflection meeting that they did not feel as though they had enough practice time on the CAS, especially during the study for Pure tone audiometry. This could explain the lack of significant differences between groups for this study. A further limitation concerns the clinical history study and is the VPs' scripting itself. While students did have positive results using the VPs, some reported having difficulties in retrieving specific information from them. This was due to the way they asked questions to the VPs that was not adequately recognized by the system. We believe that by improving patients' interaction-scripts, we could probably obtain more positive results in further studies. Another concern was the limited number of participants available due to the small number of students enrolled in this degree.

While we need more in depth studies to evaluate the effectiveness of the CAS as part of the Clinical Audiology course, at this stage we conclude that training with the CAS did not impact negatively on the students. Taking this into account, using the CAS as a supplementary tool for audiology trainees has a number of practical advantages. Students were able to practice at their own pace during their free time, as the computer room where the CAS was installed could be accessed 24/7. Then, within the few hours

of training to which students' were exposed, they accessed 25 cases covering most of the pathologies they encounter during their professional career. This would not have happened during any of their clinical placements as patients are limited and pathologies can be redundant because students could conduct similar diagnoses throughout the day. Finally, the number of patients and opportunities for actual hands-on experience can vary greatly from one clinic to another. This results in students having very different learning experiences. The CAS, on the other hand, standardizes the training, offering the same opportunities to each student.

6 Future work

The next step of this research is to conduct a study using a similar design, while taking into consideration the lessons learned from the studies presented in this paper. Improvements will be made to the CAS according to participants' feedback gathered in the present studies.

Additionally, we plan to control more thoroughly students' practice sessions by having them book training slots before hand for both studies. The CAS will also record practice time and cases seen by students. This should allow us to confirm that students are actually training for a sufficient amount of time, and also focusing on the task at hand. We also want to investigate how many and what type of patients students meet in clinic as part of the traditional training methods. We will be investigating number of patients, the range of pathologies, and the time spent in actual hands-on work with patients. Additional work will be undertaken with field experts to diversify the VPs' recognisable vocabulary and ability to respond to a varied range of questions. Students will also be required to complete a typical Adult History form for each VP assessed in order to practice reporting relevant pieces of information from patients' interviews. Other improvements will include but are not limited to the following: adding different ways of displaying students' results in order for participants to engage more in reasoning while examining their answers, integrating a more varied range of 3D models for the VP characters, and implementing additional VPs to train with.

Finally, we will aim to explore the effect of an additional formative feedback component when using the CAS, providing students with hints while they are conducting their VP assessments. This study will be conducted on a larger sample of participants recruited from Speech and Language Therapy students to allow for more statistical power.

References

- Bickmore, T. W., Pfeifer, L. M., Byron, D., Forsythe, S., Henault, L. E., Jack, B. W., Silliman, R. & Paasche-Orlow, M. K. (2010), 'Usability of conversational agents by patients with inadequate health literacy: evidence from two clinical trials.', *Journal of Health Communication* **15** Suppl 2, 197–210.
- Bickmore, T. W., Pfeifer, L. M. & Paasche-Orlow, M. K. (2009), 'Using computer agents to explain medical documents to patients with low health literacy.', *Patient Education and Counseling* **75**(3), 315–320.
- Bickmore, T. W. & Picard, R. W. (2005), 'Establishing and maintaining long-term human-computer relationships', *ACM Transactions on Computer-Human Interaction* **12**(2), 293–327.
- Collins, R. P. & Harden, R. M. (1998), 'AMEE Medical Education Guide No. 13: real patients, simulated patients and simulators in clinical examinations', *Medical Teacher* **20**(6), 508–521.
- Conradi, E., Kavia, S., Burden, D., Rice, A., Woodham, L., Beaumont, C., Savin-Baden, M. & Poulton, T. (2009), 'Virtual patients in a virtual world: Training paramedic students for practice.', *Medical Teacher* **31**(8), 713–720.
- Deladisma, A. M., Gupta, M., Kotranza, A., Bittner, J. G., Imam, T., Swinson, D., Gucwa, A., Nesbit, R., Lok, B., Pugh, C. & Lind, D. S. (2009), 'A pilot study to integrate an immersive virtual patient with a breast complaint and breast examination simulator into a surgery clerkship.', *American journal of surgery* **197**(1), 102–106.
- Dickerson, R., Johnsen, K., Raij, A., Lok, B., Stevens, A., Bernard, T. & Lind, D. S. (2006), 'Virtual patients: assessment of synthesized versus recorded speech.', *Studies In Health Technology And Informatics* **119**, 114–119.
- Ferdig, R. E., Coutts, J., DiPietro, J., Lok, B. & Davis, N. (2007), 'Innovative technologies for multicultural education needs', *Multicultural Education Technology Journal* **1**(1), 47–63.
- Fuhrman Jr, L. C., Buff, W. E., Eaddy, M. & Dolar, M. (2001), 'Utilization of an integrated interactive virtual patient database in a Web-based environment for teaching continuity of care', *American Journal of Pharmaceutical Education* **65**(3), 271.
- Gorrindo, T. & Groves, J. E. (2009), 'Computer Simulation and Virtual Reality in the Diagnosis and Treatment of Psychiatric Disorders', *Academic Psychiatry*.
- Halan, S., Rossen, B., Cendan, J. & Lok, B. (2010), 'High Score ! - Motivation Strategies for User Participation in Virtual Human Development', *Seven* pp. 482–488.
- Hayes-roth, B. & Saker, R. (2003), 'Next-Generation Automated Health Behavior Coaches', *Symposium A Quarterly Journal In Modern Foreign Literatures* pp. 28–34.
- Heer, I. M., Middendorf, K., Müller-Egloff, S., Dugas, M. & Strauss, A. (2004), 'Ultrasound training: the virtual patient', *Ultrasound in Obstetrics and Gynecology* **24**(4), 440–444.
- Johnsen, K., Dickerson, R., Raij, A. & Lok, B. (2005), 'Experiences in using immersive virtual characters to educate medical communication skills', *IEEE Proceedings VR 2005 Virtual Reality 2005* pp. 1–8.
- Johnsen, K. & Lok, B. (2008), 'An Evaluation of Immersive Displays for Virtual Human Experiences', in 'Virtual Reality Conference, 2008. VR '08. IEEE', pp. 133–136.
- Kenny, P. G. & Parsons, T. D. (2010), 'Virtual Patients for Virtual Sick Call Medical Training', in 'Training, Simulation, and Education Conference'.
- Kenny, P. G., Parsons, T. D., Rizzo, A., Reger, G. M., Caroly, P., Pato, M., Sugar, J. & StGeorge, C. (2008), 'Virtual Patients for Future Leaders'.

- Kotranza, A., Johnsen, K., Cendan, J., Miller, B., Lind, D. S. & Lok, B. (2009), Virtual multi-tools for hand and tool-based interaction with life-size virtual human agents, *in* '2009 IEEE Symposium on 3D User Interfaces', IEEE, pp. 23–30.
- Rizzo, A., Parsons, T. D., Buckwalter, J. G. & Kenny, P. G. (2010), 'A New Generation of Intelligent Virtual Patients for Clinical Training pdf free ebook download from projects.ict.usc.edu'.
- Rizzo, A., Patrick, K. & Parsons, T. D. (2009), 'JVRB - Journal of Virtual Reality and Broadcasting - Intelligent Virtual Patients for Training Clinical Skills'.
- Rossen, B., Cendan, J. & Lok, B. (2010), 'Using Virtual Humans to Bootstrap the Creation of Other Virtual Humans', pp. 392–398.
- Round, J., Conradi, E. & Poulton, T. (2009), 'Training staff to create simple interactive virtual patients: the impact on a medical and healthcare institution.', *Medical Teacher* **31**(8), 764–769.
- Ruttkay, Z. & Welbergen, H. V. (2008), Elbows Higher ! Performing , Observing and Correcting Exercises by a, *in* 'Proceeding IVA '08 Proceedings of the 8th international conference on Intelligent Virtual Agents', pp. 409–416.
- Stanton, T. (2008), 'Aspiring Medical Assistants Can Now Provide Care for Virtual Patients via Interactive Simulation'.
- Tan, Z. S., Mulhausen, P. L., Smith, S. R. & Ruiz, J. G. (2010), 'Virtual Patients in Geriatric Education', *Gerontology & Geriatrics Education* **31**(2), 163–173.
- Triola, M., Feldman, H., Kalet, A. L., Zabar, S., Kachur, E. K., Gillespie, C., Anderson, M., Griesser, C. & Lipkin, M. (2006), 'A randomized trial of teaching clinical skills using virtual and live standardized patients.', *Journal of general internal medicine* **21**(5), 424–9.
- Vash, J. H., Yunesian, M., Shariati, M., Keshvari, A. & Harirchi, I. (2007), 'Virtual patients in undergraduate surgery education: a randomized controlled study.', *ANZ journal of surgery* **77**(1-2), 54–9.
- VERG (2005), 'Virtual Experiences Research Group - University of Florida'.
URL: <http://verg.cise.ufl.edu/>