The Potential Impact of 3D Telepresence Technology on Task Performance in Emergency Trauma Care

Hanna M. Söderholm*, Diane H. Sonnenwald*, Bruce Cairns, James E. Manning, Greg F. Welch, Henry Fuchs

*Göteborg University and University College of Borås, SE-501 90, Borås, Sweden

University of North Carolina, Chapel Hill, NC 27599, USA

{hanna.maurin, diane.sonnenwald}@hb.se, {bruce_cairns, james_e_manning}@med.unc.edu, {welch, fuchs}@cs.unc.edu

ABSTRACT

Emergency trauma is a major health problem worldwide. To evaluate the potential of emerging 3D telepresence technology for facilitating paramedic – physician collaboration while providing emergency medical trauma care we conducted a between-subjects post-test experimental lab study. During a simulated emergency situation 60 paramedics diagnosed and treated a trauma victim while working alone or in collaboration with a physician via 2D video or a 3D proxy. Analysis of paramedics' task performance shows that the fewest harmful procedures occurred in the 3D proxy condition. Paramedics in the 3D proxy condition also reported higher levels of self-efficacy. These results indicate 3D telepresence technology has potential to improve paramedics' performance of complex emergency medical tasks and improve emergency trauma health care when designed appropriately.

Categories and Subject Descriptors

H.5.3 [**HCI**]:Group and Organization Interfaces-*Evaluation/methodology* J.3 [**Computer Applications**]: Life and Medical Sciences

General Terms

Measurement, Performance, Design, Experimentation

Keywords

3D telepresence, Video-conferencing, Emergency Medicine, Evaluation, Task Performance, Self-efficacy, Collaboration

1. INTRODUCTION

Trauma, or serious physical injury, is a major health problem accountable for more productive years lost than heart disease, cancer and stroke combined [13,31]. A trauma victim's recovery often depends on how soon the victim receives appropriate medical care. In general, paramedics are the primary medical personnel to provide health care to trauma victims at the scene of an accident. They diagnose each victim and, based on their diagnosis, perform medical procedures within their legal authorization and transport the victim to nearest appropriate hospital or medical facility. In complex trauma cases when the paramedic is having difficulty diagnosing the patient, needs advice regarding a procedure that should be performed, and/or needs legal permission to perform a specific procedure, paramedics will consult with a physician at a hospital or medical facility. Today paramedic – physician consultation occurs via cell phone or radio. During this consultation, a paramedic must quickly and accurately do the following: verbally describe the victim, the accident scene and the victim's symptoms; answer the consulting physician's questions; discuss treatment options with the physician; monitor the victim's progress; and simultaneously perform complex medical procedures to save the victim's life. All of these activities must often be performed within minutes. An incorrect description and subsequent decision and/or action could result in death, or further complications and a longer recovery time for the victim. Thus, visual technologies could be beneficial in emergency healthcare by providing the consulting physician with a directly transmitted view of the patient and accident scene when patients are severely injured or when there are long transport times to nearest hospital and the patient is in need of immediate care beyond the level paramedics are authorized to provide.

Two-dimensional (2D) wireless video-conferencing has been proposed as a replacement for cell phone and radio consultation in emergency medicine [11]. Yet when used in non-emergency medical situations, medical personnel report two critical limitations of this technology: difficulty obtaining the desired camera views and a lack of depth perception [15,32,40]. We are investigating whether emerging three-dimensional (3D) telepresence technology may overcome these limitations. If so, it could provide trauma victims and other patients in rural, natural- and human-made catastrophic areas with better medical care although physicians or medical experts are not physically present. Today, research is underway to develop components of 3D telepresence technology; a lab prototype is under development but not yet completed [41]. Millions of dollars are still required to develop and implement the 3D technology. However funding agencies, such as the National Library of Medicine, would like to know today whether such a large investment of resources is merited, and avoid spending these resources if the technology has no potential to improve emergency health care. To address this issue traditional human-computer interaction experimental and field evaluation methods must be augmented with other techniques because the 3D telepresence technology does not yet exist. Thus we developed a 3D proxy that simulates the 3D technology, and compared the quality of emergency trauma care provided when paramedics consulted with a physician via the 3D proxy, via 2D video, and worked alone.

GROUP'07, November 4-7, 2007, Sanibel Island, Florida, USA.

2. PREVIOUS RESEARCH

2.1 Telemedicine

Telemedicine is defined as "the use of electronic information and communications technologies to provide and support health care when distance separates the participant." [18, p.1]. Telemedicine includes a wide array of applications, ranging from transmission systems for sharing medical data such as x-ray pictures to remote health monitoring systems for elderly and chronically ill individuals, from video systems for counseling or consultation between patients and physicians or other medical specialists to video and haptic systems to support remote surgery. The majority of the research on video and visual technologies in telemedicine focuses on consultation among a patient, a physician and/or other health care personnel, or between physicians at distributed locations in non-emergency situations [e.g., 1,7,29,30].

When individuals consult or work face-to-face they share a rich visual space. Previous research in computer supported cooperative work [e.g., 17,22,34] and theory of language [12] suggests that the visual space that is shared when working remotely using 2D video conferencing technology lacks the richness of collocation and face-to-face interaction. In particular it lacks multiple and redundant communication channels, implicit cues, and spatial co-references that are difficult. This lack of richness means it is more difficult to establish situation awareness, impairing medical consultation in telemedicine applications. For example, in telemedicine two problems with respect to 2D video-conferencing technology arise repeatedly in the literature: the difficulty associated with obtaining the desired camera views and depth perception. For example, camera view difficulties are mentioned multiple times in the final report for the U.S. National Library of Medicine's National Laboratory for the Study of Rural Telemedicine [18] and David and his colleagues [15] who used a video-conferencing system for semi- and non-urgent complaints at a short-term correctional facility. Patients also recognize this problem. In a study at Georgetown University Medical Center [33] patients reported that the physician could not always "see" how the patient was "really doing." In emergency situations today physicians must resort to secondary visual cues or verbal clarification from a remote paramedic, which both impose additional cognitive loads compared to the very natural views afforded if the consulting physician were able to "be there" with the patient.

To address the visibility problem, the use of multiple cameras and camera views has been proposed, e.g., providing a picture-in-a-picture or a large display of multiple views [36]. However switching between multiple disjoint views, as a security guard might with a surveillance system, is not very feasible in time critical health care situations. In fact Fussell and colleagues [19] found that providing users with a single scene-view actually lead to better performance of a physical task than if they were given a combination of multiple camera views. Users had problems knowing which view to focus on when being presented a combination of the view from a head-mounted camera and a scene-view of the remote location. It is not difficult to imagine that this result would also hold for task performance in stressful time critical health care situations.

With a very large number of cameras and user head tracking, automatic switching based on view position and orientation could help. But the quantity and configuration of cameras necessary to achieve smooth and appropriate switching over a remote emergency scene, as well as the 2D video storage and bandwidth needs, would be impractical. While pan-tilt-zoom cameras can help address this problem, they require additional technical skills, impose an additional cognitive load, and require additional time to adjust (which is difficult in a trauma situation). Hauber et al [25] found that feelings of co-presence increased when improving spatiality in 2D video-conferencing by 3D views. Although when comparing the performance of a collaborative task done via 2D video versus via combined 2D video and 3D representation, 2D video was superior to the combined 2D and 3D representation view. Hauber suggests that the menu-systems and navigation tools used in their particular system seemed to increase the users' cognitive load, thus making it harder to perform the task with the added spatiality (3D view) than via 2D video [25]. The use of menu-systems and need for navigation could be extremely impractical in an emergency situation as a remote physician is fully dependent on immediately obtaining the desired view of the patient in order to be able to give advice regarding patient care. There is simply little or no time for navigation in a system or manual manipulation of equipment.

In addition to approaches for getting the desired view of a remote patient as described above, Tachakra states that "impaired depth perception is a significant problem in telemedicine" and notes that "the most important cue of depth is due to binocular disparity" [40, p. 84]. Tachakra suggests a coping strategy that involves "rotating the camera in the transverse plane about 30° at a time" [40, p. 83]. He surmises that this controlled camera rotation "enables the consultant to build a three-dimensional mental image of the object by briefly storing a range of two-dimensional views" [40, p. 83]. This is not surprising because object occlusion and motion parallax are two of the most powerful depth cues. However, it may often not be realistic to perform camera rotation as prescribed by Tachakra in time critical trauma emergencies in the field. The time taken to rotate a camera and view will reduce the time available for patient care. It could also reduce the number of on-site personnel who can provide treatment to a trauma victim when an on-site person is required to solely, or primarily, focus on rotating the camera. Furthermore, it could be physically very difficult to rotate a camera, e.g., when a victim of a car accident is lying on a hillside along the side of a road. To address these limitations research is being conducted on how 3D telepresence technology could provide depth perception and dynamic views.

2.2 3D telepresence technology

As shown in Figure 1 in the current vision for 3D telepresence technology users would be provided a seamless view of a remote scene, and the ability to virtually point to objects at the remote location [41]. This view would, ideally, dynamically change as the user moves her or his head or walks around. Users (in our case a physician) would wear a small head-mounted tracking device that tracks their head movement and corresponding eye gaze. The view of the remote scene would then dynamically change in real time corresponding to where in the scene their eyes are currently gazing. A digital laser pointer could be used and tracked locally, and projected at the remote scene to allow the user to virtually point at but not touch objects and areas in the remote location. Today, these capabilities are being explored using an array of cameras to capture the remote scene and computer algorithms to

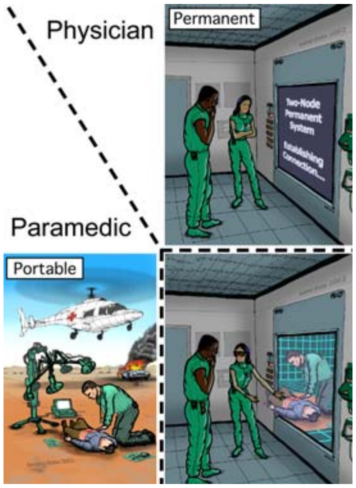

Figure 1. 3D telepresence vision

blend those camera views and present the result to remote users in real time [41]. Ideally there will be no need for navigating tools that require user input or manipulation such as zoom or change the angles of the cameras. The goal is for users to see a 3D representation of the remote scene and be able to dynamically change their viewpoint of that scene in real-time. In trauma and other emergency medical situations it could enable the physician to virtually move around the trauma victim, paramedic and equipment at a remote accident scene. It should enable the physician to quickly look at different parts of the victim's body and to look more closely at an ongoing intricate surgical procedure.

Only an early prototype system demonstrating components of 3D telepresence technology exists today [41]. However, today funding agencies wish to understand the potential of the technology. They want to avoid spending billions of dollars and intellectual resources in development and implementation costs if the technology will not improve health care in medical care situations, such as trauma emergencies. Therefore we conducted a study to investigate the potential of 3D telepresence technology in trauma related emergency healthcare.

3. EVALUATION DESIGN

3.1 Hypothesis and Approach

We propose that the use of 3D telepresence technology can improve the physician's understanding of a trauma victim and activities at a remote location, including diagnosis and treatment procedures performed by a paramedic at the remote scene. The improved understanding should enhance the consultation, and enable the paramedic to provide better medical care. In particular, our hypotheses are:

H1: Paramedics working in consultation with a physician via 3D telepresence technology will provide better medical care to trauma victims than paramedics working in consultation via 2D video or paramedics working alone.

H2: Paramedics working in consultation with a physician via 3D telepresence technology will report higher levels of self-efficacy than paramedics working in consultation via 2D video or paramedics working alone.

Evaluating technology in emergency medical situations has unique challenges, many of which can be attributed to the complex context in which trauma situations occur. When arriving at the scene of an accident, paramedics must work quickly and accurately in order to save lives, and they must do this wherever the patient is located, which could be in a ditch next to a highway at night during a heavy rain storm. Patient healthcare priorities and privacy, together with the dynamics of emergency field care make it extremely difficult, if not impossible, to collect evaluation data in the field. Yet we should take into account as many aspects of emergency medical situations as possible during an evaluation to increase the validity of our results. Thus in our evaluation we simulated a realistic emergency medical scenario, and compared paramedics' performance under three conditions – working alone, working in consultation with a physician via 2D video technology, and working in consultation with a physician via a 3D proxy. In our post-test, between-subjects design we also employed rigorously developed task performance measures in order to insure the validity and reliability of the evaluation results.

3.2 Trauma simulation

One key aspect of the simulation design was choosing an appropriate trauma emergency medical task for the participating paramedics to perform. The task needed to have a high level of complexity, consisting of demanding medical decision-making and technical skills. It needed to be a task paramedics were qualified to perform with measurable outcomes, and a task they would seek advice on. In addition the task needed to be socially relevant, i.e., an important medical problem. The task we selected was the diagnosis and treatment of a trauma related *difficult airway*, including performing a surgical cricothyrotomy. In a surgical cricothyrotomy an incision is made through the skin in the neck and the underlying cricothyroid membrane to allow air to pass to the lungs. Paramedics are trained to manage the difficult airway and perform a surgical cricothyrotomy, yet this is extremely challenging. A surgical cricothyrotomy is usually regarded as "the last resort" in pre-hospital care situations as it can be very risky for the patient. When we asked paramedics about performing a surgical cricothyrotomy, they reported:

> *Cric[othyrotomy] skills scare the crap out of you.*
>
> *It's something I hope I don't have to do.*

Even the most experienced physicians in airway management recognize the sense of urgency and anxiety associated with control of the difficult airway, because patients without an adequate airway will die within minutes if they do not receive appropriate treatment [23]. The inability to secure an airway is the most common cause of preventable death in pre-hospital care of injured patients [4]. A state-of-the-art computerized mannequin, the METI Human Patient Simulator™, was used to simulate the trauma patient suffering from a difficult airway. The mannequin (created by the manufacturers as a middle-aged Caucasian male) can simulate a wide range of medical conditions and responds to treatment in a life-like way. For example, its pupils dilate in response to light, its chest rises and falls when breathing, and its heart rate, breathing pattern and oxygen levels respond to drugs and medical procedures. The simulated trauma situation was a car accident in which a victim was thrown from a car and found moaning and groaning. Police had responded first on the scene and put the victim on a backboard as illustrated in Figure 2. To increase the ecological validity we added background sounds and lights that included traffic noises, a

Figure 2. Simulation setup

rotating beacon, people's voices and emergency vehicle sirens approaching and leaving the scene.

3.3 Conditions

As mentioned previously we compared paramedics' performance and perceptions of working under three conditions: a paramedic working alone, a paramedic working in consultation with a physician via 2D video technology, and a paramedic working in consultation with a physician via a 3D proxy. In the remainder of this paper these conditions are referred to as Alone, 2D and 3D proxy respectfully.

State-of-the-art, high quality 2D video was used in the 2D condition. Three views of the mannequin were provided to the remote consulting physician using digital cameras directly connected to three 21-inch high resolution monitors. One camera was a remote-controlled pan-tilt-zoom camera that the consulting physician could control. All cameras were placed in optimal positions for our particular trauma situation. That is, expert physicians determined the best locations for the three cameras to enable effective observation of the diagnosis and management of a difficult airway on the mannequin. Furthermore, the consulting physician had a full screen view of the patient monitor showing the mannequin's heart rate, blood pressure and oxygen saturation rates in real time. The consulting physician observed the patient monitor and camera views in a custom built work station. In addition, the paramedic also had a 2D video view of the remote physician during the simulation. This view was located at the end of the mannequin's stretcher.

Since the 3D telepresence technology is not yet sufficiently developed to allow us to use it even under ideal lab conditions, we designed a 3D proxy, or surrogate. For the 3D proxy, the consulting physician was physically present in the same room as the mannequin and paramedic. The consulting physician was allowed to freely move around in the room. However, the physician could not touch anything in the room and could only point to things using a laser pointer. This simulates the current vision and technical goals for 3D telepresence technology. As described earlier, the 3D technology should ideally enable the user to see a 3D representation of the remote scene and be able to dynamically change their viewpoint of that scene, mirroring the shared visual work space collaborators would have when working face-to-face. In our context it should allow the physician to virtually walk around the stretcher to get different views of the accident victim, or bend down to look more closely at the neck area. The physician should also be able to interact with the remote scene through a laser pointer that was displayed in his/her view and at the remote scene, e.g., to allow the physician to virtually point at the correct place for the paramedic to make the incision in the neck.

Social facilitation theory [8,14] suggests that an individual's performance is affected by the physical (collocated) and virtual presence of an audience, in the sense that a person being observed by an audience will perform easier tasks better and more difficult tasks worse. This effect occurs when at least one person is present. If two or more people are in the audience, however, the impact is the same as if only one person is in the audience. In our experiment a physical (collocated) audience that consisted of a researcher-observer and a virtual audience that consisted of an expert mannequin operator observing the study participant from an adjacent room were present during all experiment sessions across all conditions. Social facilitation theory indicates that the physical presence of an additional physical or virtual audience member, i.e., the consulting physician in the 3D proxy and 2D conditions, would have no additional impact on paramedics' performance. This is because all paramedics in the study had both a physical and virtual audience each consisting of at least one person no matter if they worked alone or with a physician via 2D or 3D proxy.

Two emergency care physicians acted as the consulting physicians in the 2D and 3D proxy conditions, with each physician participating in equal numbers across the two conditions. To help reduce the potential impact of any individual differences between the physicians, a physician – paramedic interaction script was used by the physicians. The script consisted of appropriate, constructive phrasing of responses to typical questions and advice to give the paramedic at certain times during the scenario. The main reason for this was to minimize individual differences between the physicians regarding tone of voice and advice given to paramedics. The script was based on actual physician – paramedic interaction observed during the pilot study and was developed in collaboration with the physicians.

3.4 Study participants

To determine the optimal number of sessions to be conducted, we reviewed the literature for similar studies and found that 10 to 20 sessions per condition was common. [e.g., 21]. Thus we had 20 participants per condition, for a total of 60. The 60 participants, 48 males and 12 females, were all certified paramedics working in southeastern US. In the US there are three certification levels for emergency medical technicians (EMTs): basic, intermediate, and paramedic. We chose the paramedic population because it is representative of individuals who perform the most advanced medical care among EMTs and their training and skill levels are somewhat consistent. For example, they are the only EMTs that are required to have training in performing a cricothyrotomy. Study participants were recruited through announcements on bulletin boards, email lists, newsgroups, personal emails, phone calls, advertisement in local papers, and through flyers posted in emergency medical services (EMS) offices, ambulance stations, and emergency rooms. They received $50 as appreciation for their participation in an experiment session which typically lasted 2 hours. The average years of total EMS work experience of the sixty paramedics who participated in our study was 11 years, with a range of 1 to 26 years. Their average paramedic work experience was 7 years, with a range of 1 to 24 years. The paramedics were randomly assigned across conditions, with equal distribution of gender and years of experience across all three conditions. Of all participants, 14 persons had previously performed a cricothyrotomy on a real patient, 7 in the Alone, 4 in the 2D, and 3 paramedics in the 3D proxy condition. In the 2D and 3D proxy conditions 5 and 6 paramedics respectively had met the physician before, a situation that mirrors daily work of paramedics. Paramedics usually have one or two specific hospitals where they take patients on a regular basis and where they know the some of the physicians in the ER. At other times, patients need to be transported to other hospitals or facilities where the paramedic does not know or has not met the physicians before.

3.5 Experiment session protocol

Upon arrival at our facility, each paramedic learned about the overall purpose of our research, the format of each experiment session and their rights as study participants. Each paramedic was given a study consent form to read, discuss and sign. Next, each paramedic received an introduction to the simulation, participated in our medical simulation and completed a post-test questionnaire and interview.

In the introduction to the simulation, participants were shown the video equipment and camera views, and given a hands-on tutorial to the mannequin, time to make themselves familiar with the medical equipment available for diagnosing and treating the patient, and background information about the simulated car accident and patient. The background information we provided is normally available to paramedics at accident scenes from first responders, e.g., police who first arrive at an accident scene and call for medical assistance. Each session was attended by an observer who was always collocated with the paramedic and mannequin. The mannequin technician operated the mannequin from an adjacent room and was available for questions regarding the patient simulator before and after the session, as well as during the actual simulation (through an open microphone).

The medical simulation sessions took an average of 11 minutes, with a range of 6 – 23 minutes. During each session, the paramedic needed to diagnose and treat the victim (mannequin) as discussed earlier. Each simulation session was video-recorded using four cameras that captured paramedic's actions on and surrounding the mannequin and medical monitor output (heart rate, oxygen saturation levels). The video recordings from each session were graded to evaluate the paramedics' performance, using a grading protocol developed in collaboration with two physicians as discussed in the following sections. In addition, after each session each paramedic completed a questionnaire to report their perceptions of satisfaction, self-efficacy, communication and trust with the physician (for the 2D and 3D proxy conditions). Due to space limitations this paper focuses on quality of medical care as measured by task performance and paramedics' perceptions regarding their future task performance as measured by *self-efficacy* data.

3.6 Evaluation Measures

To evaluate the quality of medical care (Hypothesis H1) we measured task performance, including task execution times, subtask performance and the occurrence of harmful interventions. In our scenario, as in many real life emergency situations, the less time used to correctly diagnose and treat a patient increases the patient's chances of a full recovery and decreases the patient's recovery time. There is also a standard emergency medical protocol, i.e., a progression of subtasks, that medical professionals are advised to follow when treating a patient who presents symptoms as in our scenario. Furthermore if a paramedic performs a harmful intervention or procedure while treating a patient, the patient may die or suffer unnecessary complications.

To measure self-efficacy (Hypothesis H2), we created a questionnaire that each paramedic completed after participating in a session.

3.6.1 Task performance

Although management of a difficult airway, including performing a cricothyrotomy, is taught to paramedics and medical students throughout the world, a standard grading protocol does not exist in the published literature. To develop a grading protocol, we researched medical education literature [e.g., 2,10,28] and previous research on performance assessment in simulated medical scenarios [e.g., 9,16,20,39]. The most common performance aspects to measure include combinations of: time for problem solving and decision making; technical, cognitive and behavioral skills; and number of appropriate/inappropriate procedures. The measures are often based on a recognized treatment algorithm, or in compliance with principles determined by medical researchers as good practice. Our main challenge was to construct an assessment protocol that could measure task performance comprehensively, based on the information that was possible to accurately estimate from the video recordings of each session. Our initial grading protocol was based on the ASA Practice Guidelines for Management of the Difficult Airway developed by American Society for Anesthesiologists [3]. The protocol was then discussed in-depth with two local emergency medical physicians who have decades of experience in managing difficult airways and performing surgical crics. The result was a grading protocol that captured the *performance time of key events, subtask execution*, and *harmful interventions*.

Timing of events (Table 1) consists of recording the time (mm:ss) for the most important events and steps in managing and treating a difficult airway.

Table 1. Timing of events

Event	Definition
Breathing stops	When the patient's chest stops moving
Decision to perform cricothyrotomy	When the paramedic either takes out any of the cricrothyrotomy equipment or the plastic bag with the equipment
Airway access – incision begins	When the paramedic begins to make the incision into the airway either with a needle or scalpel. I.E., when the needle or scalpel touches the skin.
Surgical access	Only if needle was used initially: When the paramedic begins to make the surgical incision into the airway. I.E., when the scalpel touches the skin.
Tube insertion	When a tube is fully inserted through the incision into the airway (and stays there so attaching the bag valve mask is possible)
Patient stable	When the patient's chest movements start again AND the O2 blood saturation reaches 90% or higher.

Table 2. Subtask execution

Subtask	Definition
Manual mask ventilation	If the paramedic manually ventilates the patient or not
Intubation attempts	Each time the paramedic uses the laryngoscope AND puts a tube into the patient's mouth to get it into the airway counts as one time.
Cricothyrotomy	Performing a cricothyrotomy

Subtask execution (Table 2) includes recording if certain tasks are executed or not. These tasks are not suitable for being recorded in the timing format, since they either occur repeatedly during the treatment process, or/and are better represented for interpretation in a quantified form.

Harmful interventions (Table 3) refers to treatments or procedures that are potentially harmful for the patient by increasing risks for complications or unwanted side-effects and by delaying the time it takes to the victim's breathing restored.

Table 3. Harmful interventions

Intervention	Definition
Nasal intubation	Performing nasal intubation
Chest decompression	Performing chest decompression
Locate membrane	If the paramedic is feeling the throat for locating the cricothyroid membrane before cutting
Incorrect incision	If incision is done incorrectly, either in the wrong location and/or made too large
Tube slips out	If the tube slips out of the incision after bag valve mask is attached and has to be reinserted

After the expert physicians approved the grading protocol, two researchers graded a subset of sessions independently using the protocol. They compared their results, refining the definitions used in the protocol. This procedure was repeated and inter-coder reliability reached 98%, well above standard accepted levels [35]. The protocol was then reviewed by the expert physicians to insure the definitions used were correct. A researcher then graded the remaining sessions. After this grading was done, a second researcher independently graded a random selection of graded sessions to further insure the grading was accurate.

3.6.2 Participants perceptions: Self-efficacy

Self-efficacy refers to the perceived capability to perform a certain task and is considered a powerful determinant of how well a task will be performed in the future [6]. Previous experiences, especially successful ones, are the strongest source of influence on self-efficacy, but also observing other peoples' experiences can increase one's self-efficacy [6]. Thus a paramedic's perceptions of self-efficacy after participating in our simulation can help predict the paramedic's future performance in managing a difficult airway and performing a cricothyrotomy. If after consulting with a physician in the 3D proxy condition paramedics report higher levels of self-efficacy than the other paramedics, the theory of self-efficacy predicts that the paramedics with the higher levels of self-efficacy will actually perform better in the future, improving patient outcomes in the future. Since we did not find a self-efficacy questionnaire for difficult airway management or for performing a cricothyrotomy, we developed questions based on Bandura's recommendations [5] and in consultation with emergency medical physicians. Bandura [5] recommends that self-efficacy questions at the most detailed measurement level should be related to a specific performance under a specific set of conditions. On the next level, the self-efficacy measurement should be related to a group of performances within the same activity domain under a group of conditions sharing common properties. Following these suggestions we developed questionnaire items about participants' self-efficacy to manage a difficult airway. The questions ranged from basic airway management tasks, such as manual mask ventilation (putting a breathing mask on a victim), to more complex tasks, such as securing the opening made in the cricothyrotomy membrane when performing a cricothyrotomy. The questions also referred to both physical and cognitive tasks, such as deciding on alternative treatment strategies. Thus the self-efficacy questionnaire also provides insight into cognitive task performance that can not be measured through observation.

4. RESULTS

4.1 Ecological validity

To investigate the participants' perspective regarding the ecological validity of our simulation, i.e., how closely the simulation mirrors real world conditions, we included several questions in the post-questionnaire about fidelity of the simulation and participants' engagement during the simulation. Smith and Gaba [38] define a simulation as "the artificial replication of sufficient elements of a real-world domain to achieve a specified goal" [38, p.1] and the level of accuracy of which the simulation reproduces the domain can be referred to as fidelity. The highest possible fidelity attained would be a simulation so precise that the participants in the simulation not could distinguish it from the real thing [38]. All participants reported that they thought the simulation was realistic, they were intensely absorbed in the activity, and that they fully concentrated on the scenario (Table 4.) An analysis of variance (ANOVA) yielded no significant differences between responses due to condition, suggesting that face validity was equally high across conditions.

Table 4. Ecological validity results

Question*	Result	
	Mean	**SD**
The simulation was realistic	5.80	1.246
I was absorbed intensely in the activity	6.05	0.899
I concentrated fully on the activity	6.20	0.971

** Response Scale: 1 (strongly disagree) to 7 (strongly agree)*

4.2 Task execution times

When managing a patient with breathing difficulties it is crucial to minimize the time when the patient is without oxygen in order to avoid damage to the brain. Thus, the timing of events (Figure 3) were analysed by first calculating times for total task performance (T0), time before the paramedic made the decision to perform the cricothyrotomy (T1), and time for the cricothyrotomy procedure (T2) which, if performed correctly, should provide the patients with oxygen again. The T0, T1 and T2 times also reflects decision making ability – when to take the step to actually perform a cricothyrotomy. The technical skills – such as finding

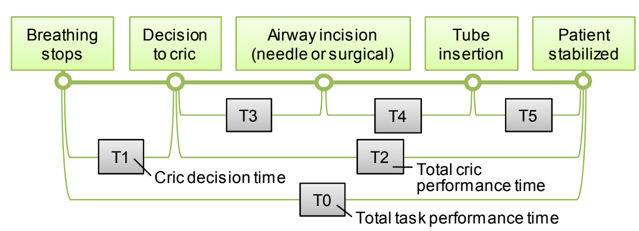

Figure 3. Task execution time calculation

the right place for the incision, inserting the breathing tube through the incision into the airway is reflected in T3 and T4. All these steps have to be done as quickly as possible in order to minimize the time when the patient is without oxygen, although performed incorrectly could have fatal consequences.

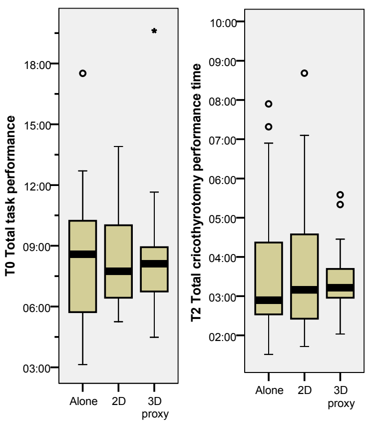

Figure 4. Variance differences in task performance time

We compared average task performance times across conditions using ANOVA. No statistical significant differences were found between task performance times due to condition. Still, it is interesting to observe that for all tasks (except for T4 "airway incision – tube insertion") the 3D proxy condition tend to show less variation in task performance times than the other conditions. Figure 4 illustrates these differences between the conditions for total task and cricothyrotomy performance. Although there were no significant differences with respect to performance times between conditions, other factors such as years of professional experience, i.e., years of paramedic experience and years of total EMS experience (defined as years of experience as paramedic EMT, and basic or intermediate EMT) influence task performance times for the Alone and 2D conditions but not for the 3D proxy condition.

Table 4. Correlations between Years of Professional Experience (YPE) and Task Execution Times

Task execution times	YPE	Condition		
		Alone	2D	3D Proxy
T0 Total task performance	Paramedic	-.633**	—	—
	Total EMS	-.519*	-.590**	—
T1 Breathing stops - cricothyrotomy decision	Paramedic	—	—	—
	Total EMS	—	-.549*	—
T2 Total cricothyrotomy performance	Paramedic	-.777**	—	—
	Total EMS	-.483*	—	—
T3 Cricothyrotomy decision – airway incision	Paramedic	—	—	—
	Total EMS	-.510*	-.534*	—
T4 Airway incision- tube insertion	Paramedic	—	—	—
	Total EMS	—	-.489*	-.453*

*** Correlation is statistically significant at $p \leq 0.01$*
** Correlation is statistically significant at $p \leq 0.05$*

The results from a Spearman bivariate correlation analysis (Table 4) show multiple statistically significant negative correlations between task execution time and years of professional paramedic and total EMS experience for the Alone and 2D conditions. That is, the fewer the years of professional experience, the longer it took to perform the task. In comparison, there is only one negative correlation in the 3D proxy condition. This suggests that impact of professional experience on task performance may decrease with the use of 3D telepresence technology.

4.3 Subtask Performance

Table 5 shows the mean and range of manual mask ventilation, intubation attempts made per session and cricothyrotomy performance in each condition. An ANOVA showed no statistically significant differences across the conditions for the execution of manual mask ventilation or intubation attempts. The standard accepted protocol for managing a difficult airway includes performing each of these subtasks several times before performing a cricothyrotomy. The lack of statistically significant results could indicate that paramedics followed this standard protocol irrespective of the condition.

Table 5. Subtask performance results

Condition	Manual Mask Ventilation		Intubation Attempts*		Cricothyrotomy Performed	
	Yes	No	Mean	Range	Yes	No
Alone	19	1	2.50	0-19	17	3
2D	20	0	2.95	0-7	20	0
3D Proxy	19	1	1.85	0-6	20	0

** Number of intubation attempts performed by each paramedic*

However, there were statistically significant results with respect to the cricothyrotomy subtask. Three paramedics in the Alone condition did not perform a cricothyrotomy. A Pearson chi-square test shows this to be statistically significant (p = 0.043). In our simulation, as in many real life situations, the trauma victim dies when a cricothyrotomy is not performed when it is needed.

4.4 Harmful Interventions

Table 6. Total Number of Harmful Interventions Performed

Harmful Intervention	Condition		
	Alone	2D	3D Proxy
Nasal intubation	1	0	1
Chest decompression	4	0	0
Not locating the cricothyroid membrane	3	1	0
Improper incision	3	1	0
Airway tube slippage	0	4	1
Totals	*11*	*6*	*2*

As shown in Table 6 two paramedics, one in the Alone and one in the 3D proxy condition, attempted to perform a nasal intubation. These results are not statistically significant, although performing a nasal intubation in real life can lead to serious medical complications for the victim in our scenario. Four paramedics in the Alone condition but none in the 2D or 3D proxy conditions attempted to perform a chest decompression. This result is statistically significant using a Pearson chi-square test (p = 0.014). In our scenario there were no indications for attempting a chest decompression. Furthermore, doing it in real life can lead to medical complications for the patient, and it is unnecessary in the sense that it takes up additional time when the patient is without oxygen. Locating the cricothyroid membrane is important because if not performed an incorrect incision may be made, prolonging the time until oxygen reaches the lungs or in the worst case, leading to a severed artery or vocal cords. As illustrated in Table 6, three paramedics in the alone condition and one in the 2D condition did not locate the membrane before making an incision. However, all paramedics in the 3D proxy condition located the membrane. These differences while important from a patient's perspective are not statistically significant.

Three paramedics working alone and one in the 2D condition performed improper incisions, either cutting in the wrong location and/or making the incision too large. This result is not statistically significant using a Pearson chi-square test and it mirrors the result regarding locating the cricothyroid membrane. As shown in Table 6, all paramedics who performed a cricothyrotomy in the Alone condition and all in the 3D proxy condition inserted the airway tube correctly without subsequent slippage. In the 2D condition, the airway slipped out in 4 instances. A Pearson chi-square test shows these differences approach standard significant level (p = 0.076).

In sum the highest number of harmful interventions occurred when paramedics worked alone. Six harmful interventions occurred when a paramedic worked in consultation with a physician via 2D video technology. In comparison only two harmful interventions occurred in the 3D proxy condition. This result is statistically significant using a Pearson chi-square test (p = 0.012).

4.5 Self-efficacy

There are two categories of self-efficacy items, referring to 1) basic airway management tasks, and 2) cricothyrotomy tasks. Basic airway management tasks are those tasks performed frequently by paramedics to insure their patients are getting oxygen into their lungs. Cricothyrotomy tasks are used less frequently, i.e., when basic airway management is not sufficient. The mean for each question per condition are shown in Table 7. ANOVA tests were performed to determine if there are statistically significant differences between conditions.

Table 7. Self-efficacy results

	Question (response scale 1, strongly disagree, to 7, strongly agree)	Mean: Alone	2D	3D proxy
Basic Airway Management	*I can quickly…*			
	- diagnose a difficult airway	*5.95*	*5.75*	*6.15*
	- manually ventilate a patient	***6.80****	***6.35****	*6.55*
	- observe problems with manual mask ventilation	*6.30*	*6.00*	*6.35*
	- decide on alternative strategy when manual mask ventilation is unsuccessful	*6.20*	*5.75*	*6.11*
	- perform initial intubation	*6.30*	*6.00*	*6.11*
	- observe intubation problems	*6.55*[a]	*6.10*[a]	*6.35*
	- decide on alternative strategy when intubation is unsuccessful	*6.30*[a]	*5.75*[a]	*6.10*
Cricothyrotomy	*I can quickly…*			
	- decide when to perform a surgical cricothyrotomy	*5.60*	***5.25****	***5.95****
	- find the location of the cricothyroid membrane (CTM)	*6.05*	***5.55****	***6.35****
	- palpate the CTM	*6.25*	***5.80****	***6.50****
	- secure the opening made in the CTM	*6.05*	***5.63****	***6.35****
	I am confident that I can do a surgical cricothyrotomy	*5.70*	***5.35****	***6.16****

** Difference between groups is statistically significant at $p \leq 0.05$.*
[a] Difference between groups is statistically significant $p \leq 0.08$.

The results are somewhat surprising. For all basic airway management tasks paramedics in the 2D condition reported *lower* levels of self-efficacy than paramedics working alone or in the 3D proxy condition. For one task, manually ventilating patients, the differences between paramedics working alone and in the 2D condition are statistically significant ($p \leq 0.05$). For two other tasks, observing intubation problems and deciding on an alternative strategy when intubation fails, the differences are statistically significant at $p \leq 0.08$.

For all cricothyrotomy tasks, paramedics in the 2D condition again reported the lowest levels of self-efficacy. Paramedics in the 3D proxy condition reported the highest levels of self-efficacy. The differences in self-efficacy between the 2D and 3D proxy conditions are statistically significant at the $p \leq 0.05$ level.

Negative correlations between self-efficacy and years of professional experience were found in the Alone and 2D conditions (all statistically significant at $p \leq 0.05$). That is, the less work experience the paramedics in these two conditions had, the lower they rated their ability to treat the patient in the simulated scenario. There were no correlations between work experience and self-efficacy in the 3D proxy condition.

5. LIMITATIONS

A limitation of this study is its post-test, between-subjects design. Alternative evaluation experimental designs include a repeated measures and Solomon four-group design [35,37]. In a repeated measures, or within-subjects, design each study participant performs the same task under all conditions. This eliminates any potential bias due to individual differences. However, it introduces a bias due to order of conditions. Study participants typically perform a task better a second time irrespective of the condition. A solution is to use different, yet ideally equivalent, tasks in each condition. However, learning can still occur in the first session and transfers to and impact performance in subsequent conditions. A Solomon four-group design in which study participants perform different and, ideally, equivalent tasks under each condition in different order is needed to isolate such learning effects. Greater resources, in terms of numbers of study participants, time and materials, are needed to conduct a Solomon four-group evaluation design. Instead we used a stratified random sampling approach and a sufficiently large number of study participants to negate potential bias due to individual differences.

A second limitation is the use of a 3D proxy in the evaluation instead of actual 3D telepresence technology. It is possible that the actual physical presence of the consulting physician in the 3D proxy condition had a positive impact on the results that will disappear in 3D telepresence technology. Even a virtual 3D hologram of a remote consulting physician shown to a paramedic (as shown in science fiction films) may not be as effective as a face-to-face consultation. Furthermore, if 3D telepresence technology introduces time delays during transmission and tracking the physician's view, task performance may be negatively impacted. However, if we are to evaluate the potential of a technology that is extremely resource-intensive to develop, some technology proxies or surrogates will be necessary. Our approach to using a proxy is similar to Wizard of Oz studies, in which a human simulates the role of an ideal computer [e.g., 27]. In both approaches simulation is used to evaluate the potential of technology not yet developed and an ideal computer system is assumed. Further research evaluating the usability and usefulness of 3D telepresence technology is needed when prototypes are available.

6. DISCUSSION

Results indicate there is partial support for Hypothesis 1: fewer subtask errors and harmful interventions were performed in the 3D proxy condition than in the Alone or 2D condition. However task performance time did not significantly differ across conditions. Three paramedics in the alone condition did not perform a cricothyrotomy and a total of 19 harmful interventions were committed across all conditions. A total of 11 harmful interventions were done in the Alone condition, 6 harmful interventions were performed when paramedics consulted with a physician via 2D video, whereas only 2 harmful interventions were performed when the consultation occurred via the 3D proxy. Thus a paramedic working alone is the least desired condition and the 3D proxy condition is more desired from a patient healthcare perspective.

Although no statistically significant differences with respect to task performance times across groups emerged from the data analysis, the results suggest 3D telepresence technology may reduce variation in task performance time across all levels of experience among paramedics. Furthermore, only one (out of five) task performance time in the 3D conditions was influenced by years of professional experience. In comparison, three and four task performance times were influenced by years of professional experience when paramedics worked alone or in consultation via 2D video, respectively. Recall that paramedics were assigned to conditions based on their years of professional experience, such that there was an equal distribution of years of professional experience across all three conditions. Thus 3D technology might have the potential to reduce differences in diagnosis and treatment caused by differences in years of professional experience. For patients this could mean that they would receive the same high level of care regardless a paramedic's years of professional experience.

The results show support for Hypothesis 2. As discussed earlier perceptions regarding of self-efficacy predict future task performance [6]. Paramedics consulting via the 3D proxy reported the highest levels of self-efficacy. This suggests that the 3D telepresence technology might have a positive impact on future task performance.

In contrast, the paramedics consulting via 2D video reported the lowest levels of self-efficacy – even lower than paramedics working alone. These results, although not statistically significant, suggest that an important area for future research. We should investigate whether the use of 2D video-conferencing technology for emergency medical care actually harms paramedics' future task performance. Could the physician – paramedic interaction during a 2D video consultation session – where the physician and paramedic must exchange basic information regarding the patient's condition and actions undertaken by the paramedic – erode a paramedic's confidence and his or her future performance? For example, during the 2D sessions we saw a physician asking a paramedic while trying to ascertain what was not working well: "Did you make the hole [surgical incision] big enough?" The physicians just could not see this important detail with the 2D cameras. However from a paramedic's perspective the question is frustrating because no paramedic would deliberately make an incision too small. Perhaps physician – paramedic interaction that is perceived as counter-productive from the paramedic's standpoint has a long term negative impact on task performance.

It is important to note that we saw several features of our 3D proxy frequently utilized during the 3D proxy sessions. For example, physicians used the laser pointer frequently to identify the location and size of the required incision and to point to specific pieces of medical equipment that the paramedic needed to use. The paramedics paid attention to the physician's pointing. As Clarke [12] and other research have indicated the ability to point to physical objects facilitates mutual understanding and task completion.

A second important feature is the ability to dynamically change views. We often saw physicians changing their viewpoint – bending down to get a side-angle view, standing up on tiptoe and bending over the victim and paramedic hands. The physicians did not need to ask paramedics to move so the physician could see the victim better. The paramedic was free to focus on the medical task at hand, and did not need to worry about the physician's view. In contrast, we seldom observed a paramedic looking directly at physician. The paramedic's visual focal point was the patient. The paramedics concentrated on doing what needed to be done to quickly save the patient. Thus producing a visual or holographic image of a consulting physician at the paramedic's location may not be so important in emergency medical care.

In summary, 3D telepresence technology shows potential to improve patient healthcare in complex emergency medical situations. Technology features, such as remote laser pointing and dynamic views, should be included in any future implementations of 3D telepresence technology for emergency healthcare.

7. ACKNOWLEDGMENTS

Our sincere thanks to: the study participants; Jim Mahaney for his expert technical assistance; Gloria Mark for her helpful feedback; and the team developing the 3D technology. This research is supported by the National Library of Medicine; contract N01-LM-3-3514, 3D Telepresence for Medical Consultation: Extending Medical Expertise Throughout, Between and Beyond Hospitals.

8. REFERENCES

[1] Alem, L., Hansen, S., and Li, J. Evaluating clinicians' experience in a telemedicine application: a presence perspective. In *Proc. OZCHI '06 Conf.,* vol. 206, ACM Press, NY, 2006, 47-54.

[2] Ali, J., Cohen, R., Adam, R., Gana, T., Pierre, I., Bedaysie, H., Ali, E., West, U., and Winn, J. Teaching effectiveness of the advanced trauma life support program as demonstrated by an objective structured clinical examination for practicing physicians. *World J Surg*., *20,8* (1996), 1121-6.

[3] American Society of Anesthesiologists Task Force on Difficult Airway Management. Practice guidelines for management of the difficult airway. *Anesthesiology*, *98* (2003), 1269-1277.

[4] Bair, A., Panacek, E., Wisner, D., Bales, R., and Sakes, J. Cricothyrotomy. *J Emerg Med*., *24, 2* (2003), 151-156.

[5] Bandura, A. *Guide for constructing self-efficacy scales. (Rev.)* Available from Frank Pajares, Emory University, 2001.

[6] Bandura, A. *Self-Efficacy*. H. Freeman, New York, 1997.

[7] Bardram, J., Bossen, C., and Thomsen, A. Designing for transformations in collaboration. In *Proc. GROUP'05 Conf.* ACM Press, NY, 2005, 294-303.

[8] Bradner, E., and Mark, G. Social presence with video and application sharing. In *Proc. GROUP '01 Conf.* ACM Press, NY, 2001, 154-161.

[9] Byrne, A., and Greaves, J. Assessment instruments used during anaesthetic simulation: Review of published studies. *Br. J. Anaesth,, 86* (2001), 445-450.

[10] Chapman, D., Rhee, K., Marx, J., Honigman, B., Panacek, E., Martinez, D., Brofeldt, B., and Cavanaugh, S. Open thoracotomy procedural competency. *Ann. Emerg. Med, 28,6* (1996), 641-647.

[11] Chu, Y., Huang, X., and Ganz, A. Wista: A wireless transmission system for disaster patient care. In *Proc. BROADNETS*, Omnipress, 2005, 118–122.

[12] Clarke, H. *Using Language*. Cambridge University Press, Cambridge, UK, 1996.

[13] Coates, T., and Goode, A. Towards improving prehospital trauma care. *Lancet, 357, 9274* (2001), 2070.

[14] Cottrell, N., Wack, D., Sekerak, G., and Rittle, R. Social facilitation of dominant responses by the presence of an audience and the mere presence of others. *J. Pers. Soc. Psycho., 9, 3* (1968), 245-250.

[15] David, M., Ellis, G., Mayrose, P., Dietrich, M., Jehle V., Ronald, M., Moscati, M., Guillermo, M., and Pierluisi, J. A telemedicine model for emergency care in a short-term correctional facility, *Telemed. J. E-Health*, *7,2* (2001), 87–92.

[16] Devitt, J-H., Kurrek, M-M., Cohen, M-M., Fish, K., Fish, P., Noel, A-G., and Szalai, J-P. Testing internal consistency and construct validity during evaluation of performance in a patient simulator. *Anaesth-Analg.*, *86* (1998), 1160-4.

[17] Dourish, P., Adler, A., Bellotti, V., and Henderson, A. Your place or mine? Learning from long-term use of audio-video communication. *CSCW, 5, 1* (1996), 33-62.

[18] Field, M. (Ed.). *Telemedicine: A guide to assessing telecommunications for health care*. National Academy Press, Washington, DC; 1996.

[19] Fussell, S. R., Setlock, L. D., and Kraut, R. E. Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks. In *Proc. CHI '03 Conf.* ACM Press, NY, 2003, 513-520.

[20] Gaba, D., Howard, S., Fish, K., Smith, B., and Sowb, Y. Simulation-based training in anesthesia crisis resource management. S*imulation & Gaming*, *32, 2* (2001), 175-193.

[21] Gale, C. The effects of gaze awareness on dialog in a video-based collaborative manipulation task. In *Proc. CHI '98 Conf.* ACM Press, NY, 1998, 345-346.

[22] Gaver, W. The affordances of media spaces for collaboration. In *Proc. CSCW'92 Conf.* ACM Press, NY, 1992, 17-24.

[23] Gawande. A. *Complications*. Henry Holt, 2001.

[24] Gutwin, C., and Roseman, M. Usability study of workspace awareness widgets. In *Proc. CHI '96 Conf.* ACM Press, NY, 1996, 214-215.

[25] Hauber, J., Regenbrecht, H., Billinghurst, M., and Cockburn, A. Spatiality in videoconferencing: Trade-offs between efficiency and social presence. In *Proc. CSCW '06 Conf.* ACM Press, NY, 2006, 413-422.

[26] Haynes, S., Purao, S., and Skattebo, A. Situating evaluation in scenarios of use. In *Proc. of CSCW'04 Conf.* ACM Press, NY, 2004, 92-101

[27] Höysniemi, J., Hämäläinen, P. and Turkki, L. Wizard of Oz prototyping of computer vision based action games for children. In *Proc. Conf. on Interaction Design and Children*. ACM Press, NY, 2004, 27-34.

[28] Johnson, D., Macias, D., Dunlap, A, Hauswald, M., and Doezema, D. A new approach to teaching prehospital trauma care to paramedic students. *Ann Emerg Med.*, *33, 1* (1999), 51-5.

[29] Li, J., Wilson, L., Stapleton, S., and Cregan, P. Design of an advanced telemedicine system for emergency care. In *Proc. OZCHI '06 Conf.* ACM Press, NY, 2006, 413-416.

[30] Mbarika, V. Is telemedicine the panacea for Sub-Saharan Africa's medical nightmare? *CACM*, *47, 7* (2004), 21-24.

[31] Meyer, A. Death and disability from injury: A global challenge. *J Trauma*, *44, 1* (1998), 1-12.

[32] Michael, M. and Kienzle, G. *Rural-academic integration. Technical Report*, National Laboratory for the Study of Rural Telemedicine, 2000.

[33] Mun, S. Project Phoenix: *Scrutinizing a telemedicine testbed. Final Project Report*, National Library of Medicine, Contract #N01-LM-6-3544, 2000.

[34] Olson, G., and Olson, J. Distance matters. *Human Computer Interaction*, *15, 2-3* (2002), 139-178.

[35] Robson, C. *Real World Research*. Blackwell, MA, 2002.

[36] Sellen, A. Speech patterns in video-mediated conversations. In *Proc. CHI'92 Conf.* ACM Press, NY, 192, 49-59.

[37] Shadish, W., Cook, T., and Campbell, D.T. *Experimental and quasi-experimental designs for generalized causal inference.* Houghton Mifflin, Boston, 2002.

[38] Smith, B., and Gaba, D. Simulators. In C. Lake, C. Blitt, R.Hines (Eds.) *Clinical Monitoring*. Saunders, NY, 2000, 26-44.

[39] Stringer, K.R., Bajenov, S., and Yentis, M. Training in airway management. *Anaesthesia*, *57, 10* (2002), 967–983.

[40] Tachakra, S. Depth perception in telemedical consultations. *Telemed J E-Health*, *7, 2* (2001), 77–85.

[41] Welch, G., Sonnenwald, D.H., Mayer-Patel, K., Yang, R., State, A., Towles, H., Cairns, B., and Fuchs, H. Remote 3D medical consultation. In *Proc. BROADMED Conf.* 2005, Omnipress, Boston, 2005, 103-110.