Voice or Gesture in the Operating Room

Helena M. Mentis
University of Maryland, Baltimore County
Baltimore, MD, USA
mentis@umbc.edu

Kenton O'Hara
Microsoft Research Cambridge
Cambridge, UK
keohar@microsoft.com

Gerardo Gonzalez
King's College London
London, UK
gerardo.gonzalez@kcl.ac.uk

Abigail Sellen, Robert Corish, Antonio Criminisi
Microsoft Research Cambridge
Cambridge, UK
[asellen, rcorish, antcrim]@microsoft.com

Rikin Trivedi
Addenbrooke's Hospital
Cambridge, United Kingdom
rikin.trivedi@addenbrookes.nhs.uk

Pierre Theodore
UCSF Medical Center
San Francisco, CA USA
pierre.theodore@ucsfmedctr.org

CHI'15 Extended Abstracts, April 18 - 23 2015, Seoul, Republic of Korea
Copyright 2015 ACM 978-1-4503-3146-3/15/04...$15.00
http://dx.doi.org/10.1145/2702613.2702963

Abstract
This case study represents our efforts to investigate the uses of voice control versus gestural control in the OR. We present a system we expressly built to allow for both gestural or voice control at the choice of the surgeon. We explain our deployment of this system in the context of cardiothoracic surgery and present a vignette on how the system was used in the moment by the attending surgeon. We learn that, in terms of design, its not just a question of saying voice is better for one type of functionality and gesture is better for another; rather, the benefits are circumstantial. Thus, there is a case for building in redundancy in control with both voice and gesture.

Author Keywords
Voice control; gesture; operating room

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

Introduction
Both before and after the commercial introduction of the Microsoft Kinect, there has been a great deal of interest in gestural interaction in the OR [3, 4, 5, 6, 7, 9, 10, 11, 12, 13]. The premise lies with gesture being a 'touchless' mechanism for surgeons to interact

with imaging systems, displays, and controllers without breaking the sterility barrier. With this motivation, gestural interaction has proven to provide this benefit. But there are occasions too where gesture poses certain challenges for the operating clinician as well as imposing certain costs to the surgical team – namely the extra work necessary to use the system and ensure it works correctly in the OR environment. For instance, there is the work that must be done prior to using the system in learning the particular set of gestures, followed by the work that must be conducted in situ by the entire surgical team to ensure that the gestural recognition system is only detecting the intended user [10]. Additional constraints relate also to the bodily and spatial requirements of gesture based systems. For example, there are many points during a procedure when the surgeon's hands are engaged elsewhere such as when they are manipulating various medical instruments. At these points, they do not have the bodily potential to enact hand and arm based gestures for the purposes of interacting with medical imagery. In addition to this, there are many factors within the surgical setting that require the surgeon to be in particular positions and bodily orientations with respect to the patient that may inhibit their bodily performance in front of a depth and gesture sensing camera [9, 10].

With these reasons in mind alternative forms of touchless interaction are also worthy of consideration here in terms of their ability to overcome the constraints of sterility in interaction. One such modality is that of voice control, which has been suggested as potentially a more suitable interaction mechanism for the operating room. In part this may come down to issues of a more precise and unambiguous form of interaction. In other words, 'on' and 'off' is thought to embody an unambiguous command versus a wave of the arm 'up' and 'down'. There are problems with voice control as well of course, but in general, they do have a higher rate of accuracy. Such issues though are only part of the story here and much of what is of interest in the alternative modality lies in what it may offer in terms of the opportunities for control that may contrast with those of gesture based interaction. For example, voice control is not dependent upon the movement of arms and hands and thus affords certain opportunities for hands free control when the surgeon's hands may be busy. Similarly, voice control is arguably less constrained in terms of the demands of bodily positioning of the clinician, allowing the clinician greater flexibility over their positioning in theatre while interacting with images.

In this respect, voice control as a touchless modality is not something that need be regarded as singularly better or worse that gestural interaction. Rather the issue might be better characterised here as one in which voice based interaction might be better for some things while gestural interaction might be better for others. The concern then becomes one of understanding how these modalities might combine with and complement each other.

In our own work developing touchless interaction systems with various surgeons, these issues surfaced in many of our ongoing design discussions with them. While we had initially set out to explore gestural control as a touchless interaction modality, we were regularly asked to implement elements of the interaction using voice control instead of or in addition

to gestural control. This case study represents our efforts to engage with these issues. We present a system we expressly built to allow for both gestural or voice control at the choice of the surgeon. We explain our deployment of this system in the context of cardiothoracic surgery and present a vignette on how the system was used in the moment by the attending surgeon. We conclude with our explanation of these findings in the context of our prior work's findings.

Background Literature

Accompanying the challenges of interaction with images under conditions of sterility, a growing research field has developed to explore the social, technical and design challenges related to this space. Early systems to enable gestural control of medical imaging in surgical settings were first demonstrated by the likes of [5] and [13]. Following on from their work and with the introduction of low cost commercial depth sensors, several other systems followed suit, further expanding the capabilities and expanding the vocabulary of gestures available [3, 4, 7, 10, 11, 12]. The key concerns with these systems have often been to demonstrate the concept and the technical feasibility of enabling some form of gestural interaction control for medical imaging. As the number of systems have grown however, it is becoming increasingly important to understand and articulate some of the key user experience and design concerns that arise in the development of these systems and how particular design choices play out in practice. While we are beginning to see these kinds of reflections and analyses [e.g. 10, 11], further understanding of these interaction modalities in surgical settings is necessary.

As well as explorations of gestural control in surgery, voice based interactions have also been considered for use in the operating theatre [1, 2, 8, 14]. In part these are motivated by the challenge of sterility, but they are also are driven by the desire to increase the interaction bandwidth for controlling more elements of the system. The latter motivation points to the possible opportunities of how modalities can add interaction but there is little work that really seeks to articulate the best ways to do this. Indeed very little work has sought to combine these modalities together in surgical settings with any articulable rationale behind. O'Hara et al in their touchless system for vascular surgery did combine elements of voice and gesture in some limited ways [10, 11]. While some rationale for this combination was offered in terms of preventing problems of gesture transition, a more full exposition and evaluation of these concerns were not more fully explored. As such there is a need to further explore ways that voice and gesture might be combined in complementary ways and in ways that allow interaction through the contextually dependent preferences of the clinician.

System Description

The system we developed accepts two types of input methods that allow the user to interact with the medical images on screen: voice commands and hand gestures. Both modalities can be used to achieve all the various functionalities and manipulations available in the system though they differ in terms of their ease and suitability for particular features. Voice, for example is used to issue discrete commands such as for changing between interaction modes and to trigger specific actions to control the overall application. The use of hand gestures allows the user to manipulate the

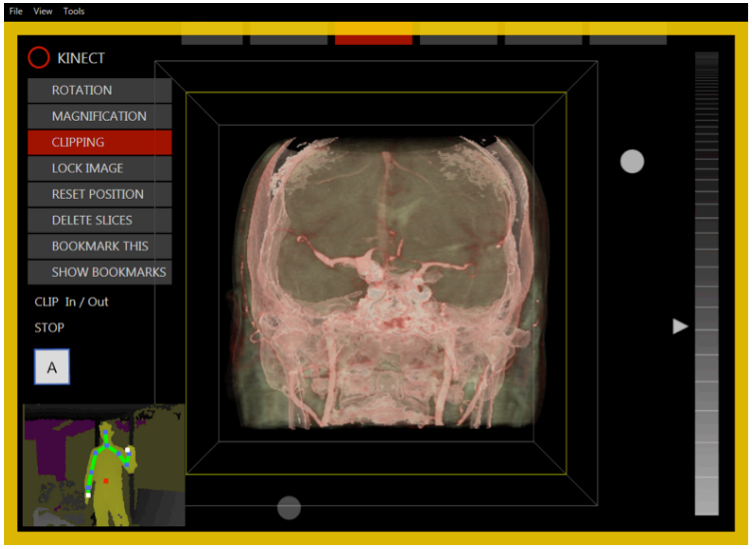

Figure 1 System Display During Clipping

images in a continuous manner and to interact with different elements of the user interface. Interface components can also be selected through the use of gestural control if necessary. Voice control cannot be used directly to control images in a continuous manner but can approximate this functionality through the use of a start-stop model of interaction. This entails issuing a discrete command to initiate a continuous action (e.g. zoom in, or clip in) followed by another discrete voice command to stop the action.

Commands are divided in two classes: manipulation modes and single actions. Manipulation modes (e.g. rotation, magnification, clipping) allow the tracked user to firstly engage the application or to simply change modes if the user is already engaged. When a user is engaged, the position of his/her hands will be represented on screen as circular cursors. In addition, a border around the application window will be displayed. The colour of the border will change depending on the type of manipulation mode that the user is currently in.

Clipping with Voice and Gesture

In order to step through the CT Scan slices, the user could choose from voice or gestural commands. The use of the voice commands is a multi-step process. The user must issue the initialization keyword "Kinect" followed by the event command, "Clipping", and then the trigger for automatic plane clipping manipulation: "Clip in", "Clip out", or "Stop".

When using gesture, the user must again issue the initialization keyword "Kinect" followed by the event command, "Clipping", but then is able to interact with the planes by closing and opening his/her right hand and performing a push or pull gesture. In this mode, a slider bar will be displayed on the right hand side of the application window indicating the current plane depth level.

Background on Use Case

The system presented was developed in response to requests and feedback we received from neurosurgeons and vascular surgeons that have used our prior gestural system [10, 11]. For this case study, a senior cardiothoracic surgeon used the system. Three days before the system was deployed in the OR, we held a training session with the surgeon to show him how to use it. In addition, we interviewed him to determine his preconceived notions of the place for gesture and voice control in the OR as well as determine his preference for control. The following findings from that discussion provide a basis for understanding the surgeon's subsequent use of the system we developed and deployed.

In this interview he first explained his pre-conceived notion for touchless system interaction in the OR. "If it is really a dynamic tool that is interactive with a PACS system…then the software to actually allow you to go through the different films and pick them out is really

important. … Ideally, … what I really want to say is 'Patient DeMarco CT Scans, latest. … And then what I really want to do is walk through the group of images I want to see. … I need 20 images to be able to scroll through them back and forth."

When asked to further explain when these groups of images would be brought up, he continued, "I think it has to be brought up in the middle of a certain circumstance." In other words, in cardiothoracic surgery, the need to consult images is when there is a concern or problem presenting itself during the case.

Finally, when we began to discuss the use of gesture versus voice control, he specified his interest in voice control. "If I were going to try to bet on one of [the interaction mechanisms making it into the OR], I would think the voice commands." He continued by describing his experience with the Aesop robotic arm. The Automated Endoscopic System for Optimal Positioning (Aesop) was a robotic arm that maneuvered an endoscope during surgery through discrete voice commands such as up, down, left, right, in, and out.

Use Case

Three days after the above interview and training session, the system was deployed during a case of recurrent squamous cell carcinoma at the site of a previous tracheostomy. The system was placed at the side of the operating room where there was enough room for the attending surgeon and his colleagues to congregate around the display.

During the case, the surgeons began to worry that the soft tissues they were encountering were recurrent cancer. The attending surgeon approached the system towards the end of the surgery after they further investigated the soft tissue. He used voice control to start the system and then commanded "clipping" on the axial slides of the chest CT. At that point, he raised his hand and began to use gestures to move through the CT slices to show the trachea. After a moment he calls the resident over.

Surgeon1: Hey [name], let me show you something on the CT scan of this patient.
[The surgical fellow approaches and stands to the side of the display able to both see the display and speak to the attending surgeon.]
S1: Here is this guy's CT scan and there is the trachea. And you see there is a little stuff that is anterior to the trachea there?
S2: Yeah.
S1: There is a little bit of soft tissue there.
S2: Yeah I just shaved off …
S1: Oh, maybe that's what you were shaving off…
S2: Yeah the whole trachea wall was coming down at least above the fat.
S1: But the nice thing is that it looks like there is a plane behind this. Behind the Manubrium it looks like there is actually a plane there. And there is the Clavicular head. So… [nods at fellow to indicate all is OK]
S2: OK, thank you.

During the above interchange, the surgeon has his 'hand on' the image. He is slowly clipping back and forth over the spot they were dissecting. He continues to talk to the surgical team while moving slightly back and forth – providing a gestalt of the tracheal area through clipping.

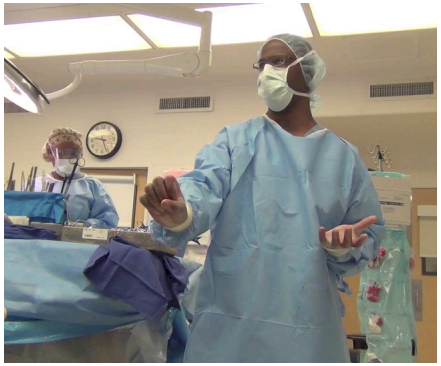

Figure 2 Attending surgeon clipping CT Scan with right hand while discussing with his colleague and gesticulating with his left hand.

Throughout this entire discussion as well as before and after, he used the voice control to initialize the system and choose the clipping manipulation mode, but as soon as he needed to engage continuous interaction such as clipping, he switched to hand control. His primary motivation to use the system was in order to discuss with his colleagues the structure they had just been working on that they had a concern may be further cancer. His continual interaction with the images during his discussion was for the benefit of his collaborators, not him.

It is important to note that the researcher who was present reminded the attending surgeon that he could use the verbal commands to clip in and out. However, the attending's assessment at that moment was that the OR was too noisy and so he preferred the "hand control".

In a follow-up interview, the surgeon commended how good the system was in the OR. He explained that he appreciated being able to show his colleagues what they were concerned about. He also explained that, beforehand, he was afraid his use of the system was going to be him waving his hands, looking silly. But now, after using it in the context of an OR procedure with a need to talk about the images, he actually saw the benefit of being able to gesture in the OR.

Discussion

Our aim in this case study is not simply a straightforward concern with evaluating our system in and of itself. Rather in developing the system we have been looking to open up discussion and further our understanding of the different types of touchless modalities (voice and gesture) that may offer a different set of opportunities to clinicians under different contextual circumstances of the procedures. Through our early design discussions with clinicians, through direct responses to the system by clinicians, through the ways that they actually use the system, and through the broader discussion opened up through use, different facets of these modalities come to the fore in interesting ways.

First, what strikes us as noteworthy is the suggested use of voice to specify a particular set of images to bring up. Here the suggested potential for voice manipulation can be found in the ability to easily specify a direct and deep link into the PACS system (e.g. "Patient DeMarco CT Scans, latest") something that would be cumbersome through equivalent gestural manipulation. Gesture then might be the preferred way to follow up once the appropriate images are in place.

Second is that voice commands appear to offer important control opportunities when other interaction channels may be limited or constrained – the AESOP system being a case in point here where the manipulations of the system need to be made in the context of other interactional requirements with the robotic system. Furthermore, in the case of the AESOP system, what is being performed here is the production of a suitable but static view. There is not a need here for an ongoing continuous control of the system once the desired view is in place.

A third point is that away from the patient table, gestural command was suitable as a primary manipulation modality. In part, this relates to the fact that these medical image interactions were being

performed at a time when the clinician's hands were available. But perhaps more significant here is the fact that the image manipulations were being dynamically performed in the context of collaborative discussion rather than just for the benefit of the individual clinician. In this respect, it was important to have both continuous control as well as the ability to speak freely about the image with the colleague. As such, gestural control was the adopted modality.

In conclusion, what is apparent through the process of designing, developing and testing this system are the different ways that voice and gesture may come to the fore for different purposes and circumstances throughout a procedure. What this highlights in our understanding of these different modalities is that there is not a clear and straightforward breakdown of how voice and gesture should be allocated to particular functionalities within the interface. In terms of design its not just a question of saying voice is better for this type of functionality and gesture is better for that type of functionality. Rather, there benefits are circumstantial. In this respect there seems to be an additional case for building in elements of redundancy and overlap within these voice and gesture controlled systems. As we have done in this system, we have enabled functionalities to be achieved both through voice and through gesture in ways that can enable a more flexible combination of their benefits as the contextual circumstances of the procedure require.

References

1. Allaf, M Jackman, S., Schulam, P., Cadeddu, J., Lee, B., Moore, R. and Kavoussi, L. (1998) Laparoscopic visual field: Voice vs. foot pedal interfaces for control of the AESOP robot. In *Surgical Endoscopy*, 12(12) pp. 1415–1418.

2. Carpintero, E., Perez, C., Morales, R., Garcia, N., Candela, A. and Azorin, J. (2010) Development of a robotic scrub nurse for the operating theatre. In *Biomedical Robotics and Biomechatronics (BioRob*) 2010, pp. 504 –509

3. Ebert, L., Hatch, G., Ampanozi, G., Thali, M., and Ross, S. (2011) You Can't Touch This: Touch-free Navigation Through Radiological Images. In *Surgical Innovation,19(3), pp. 301-307*.

4. Gallo, L. Alessio Pierluigi Placitelli, A.P. Mario Ciampi, M. (2011) Controller-free exploration of medical image data: experiencing the Kinect. In *Proceedings of 24th International Symposium on Computer-Based Medical Systems (CBMS),* Bristol, UK.

5. Graetzel, C., Fong, T., Grange, S., Baur, C. (2004) A Non-Contact Mouse for Surgeon-Computer Interaction. In *Technology and Health Care,12.*

6. Johnson, R., O'Hara, K., Sellen, A., Cousins, C., & Criminisi, A. (2011). Exploring the potential for touchless interaction in image-guided interventional radiology. In *Proc. of CHI 2011*, Vancouver, Canada, pp. 3323-3332.

7. Kipshagen, T., Graw, M., Tronnier, V.,.Bonsanto, M. and Hofmann, U. (2009) Touch- and marker-free interaction with medical software. In *Proc. of WC '09,* pp75–78.

8. Kochan, A. (2005) Scalpel please, robot: Penelope's debut in the operating room. In *Industrial Robot, 32(6)*, pp. 449–451, 2005.

9. Mentis, H., O'Hara, K., Sellen, A. and Trivedi, R. (2012) Interaction Proxemics and Image Use in Neurosurgery. In *Proceedings of CHI 2012*, Austin, Texas.

10. O'Hara, K., Gonzalez, G., Penney, G., Sellen, A. Corish, R., Sellen, A. et al (2014) Interactional Order and Constructed Ways of Seeing with Touchless

Imaging Systems in Surgery. In *Journal of Computer Supported Cooperative Work, 3(23).*

11. O'Hara, K., Mentis, H. Sellen, A. et al (2014) Touchless Interaction in Surgical Settings. In *Communications of the ACM*, 57 (01), pp70-77.

12. Ruppert, G., Amorim, P., Moares, T., and Silva, J. (2011) Touchless Gesture User Interface for 3D Visualization using Kinect Platform and Open-Source Frameworks. In *Proc. of the 5th International Conference on Advanced Research in Virtual and Rapid Prototyping*, Leiria, Portugal, pp 215–219.

13. Stern, H., Wachs, J. and Edan. Y. (2008) Optimal Consensus Intuitive Hand Gesture Vocabulary Design. In *Proc. of ICSC '08*, Washington, DC, USA, 96-103.

14. Vara-Thorbeck, C., Murioz, V., Toscano, R., Gomez, J., Fernandez, J.,Felices, M. and Garcia-Cerezo, A. (2001) A new robotic endoscope manipulator a preliminary trial to evaluate the performance of a voice-operated industrial robot and a human assistant in several simulated and real endoscopic operations. In *Surgical Endoscopy*, 15, pp. 924–927.