Interactional Order and Constructed Ways of Seeing with Touchless Imaging Systems in Surgery

KENTON O'HARA[1], GERARDO GONZALEZ[2], GRAEME PENNEY[2], ABIGAIL SELLEN[1], ROBERT CORISH[1], HELENA MENTIS[4], ANDREAS VARNAVAS[2], ANTONIO CRIMINISI[1], MARK ROUNCEFIELD[3], NEVILLE DASTUR[2], TOM CARRELL[2]

[1]Microsoft Research, Cambridge, UK
[2]Kings College London, UK
[3]Lancaster University, UK
[4]University of Maryland, USA

While surgical practices are increasingly reliant on a range of digital imaging technologies, the ability for clinicians to interact and manipulate these digital representations in the operating theatre using traditional touch based interaction devices is constrained by the need to maintain sterility. To overcome these concerns with sterility, a number of researchers are have been developing ways of enabling interaction in the operating theatre using touchless interaction techniques such as gesture and voice to allow clinicians control of the systems. While there have been important technical strides in the area, there has been little in the way of understanding the use of these touchless systems in practice. With this in mind we present a touchless system developed for use during vascular surgery. We deployed the system in the endovascular suite of a large hospital for use in the context of real procedures. We present findings from a study of the system in use focusing on how, with touchless interaction, the visual resources were embedded and made meaningful in the collaborative practices of surgery. In particular we discuss the importance of direct and dynamic control of the images by the clinicians in the context of talk and in the context of other artefact use as well as the work performed by members of the clinical team to make themselves sensable by the system. We discuss the broader implications of these findings for how we think about the design, evaluation and use of these systems.

1. Introduction

With advances in medical imaging technologies in recent decades, we have seen their widespread adoption in surgical procedures. A quick look around any modern-day operating theatre reveals the presence of an array of visual displays for accessing a wide variety of pre and intra-operative images including Computer Tomography (CT), Magnetic Resonance Imagery (MRI), fluoroscopy and various other procedure-specific imaging applications. Such imaging resources are more than simply self-contained and self-explicating visual representations that allow a clinician to "see" inside the body and make visible what would otherwise be non-visible. Rather, they are inherently constitutive of the social practices of surgery itself in the ways these visual resources are constructed, oriented to, attended to, manipulated, discussed and gestured around (Goodwin, 1994, 2000; Lynch, 1990). It is through the ways that they are embedded in the collaborative practices of surgery that such visual resources are made meaningful.

Within the operating theatre, a central feature in the organisation of such practices is the concern with maintaining strict boundaries between that which is sterile and that which is not (e.g. Katz, 1981; Johnson et al, 2011; Mentis et al, 2012; O'Hara et al, 2013). The standard input mechanisms available for controlling imaging systems in theatres (mouse, keyboard, touchscreen) are designated as non-sterile and all require contact in interaction. For those surgeons and clinicians in theatre who are scrubbed up and gloved, contact with such devices is not possible without breaking asepsis. In managing this constraint, the collaborative visual practices of surgery come to be organised in particular ways across different members of the surgical team. A well-established practice, for example, is for the scrubbed clinicians to have non-scrubbed personnel (e.g. radiographers and nurses) interact with the medical imaging equipment, under instruction, on their behalf (Graetzel et al, 2004; Johnson et al, 2011; Mentis et al, 2012; O'Hara et al, 2013). Such distributed control practices can work successfully when the imaging requirements are relatively discrete and articulable or when the surgeon is simply concerned with getting to a particular "end view". But there are times too, when such dependencies on others for control can be problematic. Team members may be engaged elsewhere and not immediately on-hand to help. Communication of particular instructions can be cumbersome and time consuming.

Perhaps more significantly though, is that at times, the analytic, interpretive and communicative work performed through these imaging resources is inextricably bound up in the actual interaction and manipulation of these images themselves. At these times, the scrubbed clinicians look to ways of achieving more direct hands-on control. For example, some clinicians will flick the surgical gown over their gloved hands and manipulate a mouse through their gown in order to achieve direct control over imaging resources without breaking asepsis (Johnson et al, 2011). Such practices are not entirely risk free but for certain clinical procedures have become accepted practice both because of the interpretive benefits of hands-on control of the images as well as certain clinical benefits of the time saving they entail (for example, less time for the patient under anaesthetic). For other procedures, where the risks may be higher, such practices are not undertaken. In these circumstances clinicians have to de-glove and rescrub in order to enable any necessary hands-on control of the imaging resources – this is a time consuming process.

Orientation towards sterility here is not simply a question of *adopting* these strategies and techniques to manipulate medical images. Rather, it is in the need to adopt such strategies and techniques that imaging resources come to be mobilised in particular ways in the context of practice. This distinction is an important one since in the context of any surgical procedure there is not a singular prescribed set of image views and manipulations to be performed. Rather, they are a set of visual resources that the clinicians assemble and orient to in an occasioned manner. The way they are mobilised in particular instances is subject to all sorts of judgements and clinical factors. For example: what is the level of uncertainty about a particular interpretation; what kinds of imaging will help resolve it; what might help my colleague understand my interpretation and persuade them of a particular course of action; is it worth the extra level of radiation to create a better image; is it worth the extra injection of contrast dye; is it worth the extra time involved; are there other resources available to allow the clinician to act and proceed in a clinically suitable manner etc. These ongoing judgments about the utility of any particular image manipulation are bound up in the work necessary to achieve a particular course of clinical action. The point here then is that an orientation to sterility is not simply adopting sterile ways of doing a prescribed set of imaging tasks. Rather the sterile strategies change the ways that the images come to be mobilised in particular contexts of practice.

In response to the concerns of clinician control over image manipulation and navigation in the context of sterile environments, a growing number of research groups have begun to explore the interaction with medical imaging systems without touching, for example by tracking gestures of the clinicians or using other touchless modalities such as voice control. We see the beginnings of such work emerging in the middle part of the last decade with, for example, Graetzel et al's gesture controlled mouse functionality (Graetzel et al, 2004) and Wachs et al's gesturally richer Gestix system (Stern et al, 2008; Wachs et al, 2006). In more recent years, the introduction of lower cost sensing and tracking technologies, has enabled several other research groups and commercial organisations to develop additional touchless imaging systems and further explore the design space in their development (e.g. Ebert et al, 2012, 2013; Gallo et al, 2011; Jacob et al, 2012; Kipshagen et al, 2009; Ruppert et al, 2012; Strickland, et al, 2013, Tan et al, 2011).

With the growth of the field, the concerns are no longer simply with the mere demonstrations of technical feasibility of such solutions. Indeed these have been amply demonstrated by the systems outlined above. Rather, it is becoming increasingly important that we understand how such systems and their particular design instantiations come to bear on the organisation of imaging and surgical practices within the operating theatre. Curiously, while we have seen a growth in the number of systems developed, there has been very little, if any, work that has attempted to articulate how the social practices of surgery and seeing become organised around these touchless systems. Rather, such systems have typically been assessed and judged simply with respect to the "alleviation" of the frustrations outlined above; in their argued ability to cut out any de-gloving and rescrubbing time; in their ability to overcome the communication frustrations and dependencies of image control by proxy under instruction. What is missing in such accounts is how these new touchless ways of manipulating medical images allow them to be mobilised in new ways in the context of collaborative surgical practices and how collaborative practices of surgery may shift to accommodate such technologies. With these concerns in mind, we present in this paper, an ethnographic study of a Kinect-based system that we developed to enable the touchless control of medical images during

3

vascular surgery. The system has been deployed and used for real clinical procedures in the endovascular suite of a large UK hospital. Our aim here is to move beyond the existing demonstrations of technical feasibility to further understand the collaborative practices emerging in its use in context. Before moving on to the study, we move on to discuss a range of related work that we will use to ground our understanding and subsequent discussion.

2. Related Work

As we alluded to in the introduction, efforts to develop touchless control of medical imaging systems have been in evidence since the middle part of the last decade. One of the earliest examples here was the contactless mouse system developed by Gratezel et al (Graetzel et al, 2004). Gratezel's system employed camera tracked hand gestures and was used simply to control basic mouse functionality such as cursor movement, pointing and clicking. While such a system served the purpose of overcoming the constraints of sterility, it was fairly limited in its utilisation of gestural capabilities. We began to see more sophisticated use of air-based gestures for the control of medical imaging technology in Wachs et al's Gestix system (Stern et al, 2008; Wachs et al, 2006). Instead of emulating mouse functionality, the Gestix system sought to introduce more bespoke gesture-based control for functionality such as navigation, zooming and rotation. Building on the core ideas of these initial systems, more recent years have seen a growth in the number of touchless systems being developed to support the touchless control of medical images. Much of this growth can be attributed to introduction of low cost sensor and development kits such as the Kinect, which have lowered several barriers to entry to the development of such systems (financial costs, development complexity, the need to wear trackable markers to overcome some of the inherent challenges of full depth skeleton capture from purely camera based systems). One of the first systems to take advantage of the infra-red depth sensing capabilities of the Kinect sensor was the GestSure system deployed in Sunnybrook Hospital in Toronto (Strickland et al, 2013). Again this system took the approach of gesture-controlled mouse emulation, which while constrained, was motivated by a number of important pragmatic considerations. In the first instance, this choice was made to enable compatibility with existing medical software systems, a potentially important concern in the adoption of these technologies. Furthermore, such an approach was considered to provide value in terms of ease of use and learnability as well as reliability gains arising from the more reliably distinctive gestures possible with smaller gesture sets.

In contrast, other systems utilising the Kinect, have sought to offer a richer set of image manipulation possibilities by developing much larger gesture sets. Compatibility issues in these systems have been oriented to by interfacing with various standardised open source DICOM image viewers and PACS (Picture Archive and Communication System) systems, such as MITO and OsiriX. Notable examples here have included the systems developed by Gallo and colleagues (2011), Ebert and colleagues (2012, 2013), Ruppert et al (2012) and Tan et al (2011). Enabling these richer image manipulation possibilities is an important development but it also raises a number of interesting challenges for such systems. One such concern is that of *expressive richness*, namely, how to map an increasingly larger set of functional possibilities coherently onto a reliably distinctive gesture vocabulary. Related to this is the issue of inadvertent triggering arising in *gestural transition* – whereby movements to transition between gestures share

common kinaesthetic components with actual defined gestures in the set. The systems outlined above have adopted various approaches in attempt to deal with these concerns such as the use of modes to distinguish gestures, the use of composite multi handed gestures and different input modalities such as speech. For example, in both Gallo et al's (2011) and Ebert et al's (2012) systems the gesture sets are made up of both one handed and two handed gestures in various different combinations. More recent versions of the Ebert system increase the expressive richness with algorithms for finger level gesture detection (Ebert et al, 2013). The systems of Tan et al (2011) and Ruppert et al (2012) use composite bimanual gestures in which the non-dominant hand is used to denote a particular function while movement of the dominant hand enables the continuous adjustment of image parameters.

The other intriguing possibility for these functionally richer systems is the use of different input modalities such as speech. This has been explored, for example in Ebert's work and our own work. While there remain some acknowledged challenges with voice recognition, it can nevertheless be useful for the initiation of discrete commands though is not suitable for continuous parameter adjustment for which gestures are better suited. Indeed as a hands free mechanism, voice may offer some potential benefits within surgical settings but what remains an important consideration is how best to combine and distribute functionality across gestural and voice modalities.

What is apparent from looking at the systems is that embedded amongst the common motivating concerns of the work is a varied set of design specifics in the way these concerns are realised. While some of these systems have been deployed in actual clinical cases, there is little in the way of articulation of what these specific technologies mean for the "practices of seeing" in surgery. As such, while we can ascertain to an extent how these particular system approaches may be motivated by the pragmatics of control, what is more difficult to ascertain is the extent to which these system are motivated by the broader set of socio-technical concerns associated with their context of use. What we need, then, is a deeper understanding of how these different kinds of systems and their design choices impact on the collaborative organisation of visual and surgical practices within these settings.

While we lack any ethnographic studies of these technologies in practice, our concerns are informed by a number of key studies of imaging practices in surgical settings that have set out specifically to understand the implications for touchless technologies in the control of medical images (Johnson et al, 2011; Mentis et al, 2012; O'Hara, 2013). The Johnson et al study examines the practices of interventional radiology, a closely related domain to that of the vascular surgery work of the current paper. The study raises a number of important issues for our understanding of touchless systems and the organisation of action in imaging practices in surgery. Key here is their discussion of image production and use as a collaborative concern of the larger surgical team. Of interest is how such collaborative organisation is bound up with the orientation to sterility and the particular constraints of more traditional touch-based technologies. They highlight the particular challenges of manipulating images under instruction and clinicians' attempts to work around these in seeking more direct "hands-on" control. But they also note that this collaborative organisation of imaging is in part a pragmatic concern whereby various imaging production and reference practices occur while one or more of the surgeon's hands are manipulating surgical tools and catheter wires. Such concerns become significant in how we understand the impact of new touchless systems and the particular ways they can be used in the context of other surgical activities. Of further import to our concerns in their study is how imaging practices are not just a visual

5

concern in which they are viewed and understood. Rather, they argue that the imaging resources are also sites for the production of talk, gesture and discussion by the clinical team as they collaboratively interpret, persuade and plan appropriate courses of action. What is points to then is a need to understand how touchless and gesture based systems come to be situated in the production of this talk, both in the way they potentially enable and constrain these activities.

The related study by Mentis et al (2012) is also concerned with the organisation of imaging practices in surgical settings, this time within neurosurgery. While it again explores the orientation to sterility and implications for touchless medical imaging technologies, it raises key questions about touchless control as a spatial concern. To this end, it focuses on the notion of *interaction proxemics*, namely the spatial consequences of touchless technology. The study highlights a number of key concepts relating to the ways that particular interaction technologies spatially organise imaging practices, namely *control proxemics*, *perception proxemics* and *deixis proxemics* – that is, where one can be positioned to be in order to manipulate images, from where they can be viewed and from where they can be pointed to and gestured over. There is an important interplay here with other features of the surgical setting that place particular demands on the spatial organisation of the surgical team. This interplay then has consequences for how the imaging resources can be drawn into the practice; when they can be viewed and manipulated and by whom. For our concerns in this paper, it is key that we understand how the new spatial dynamics of touchless technologies play out in the organisation of imaging practices during surgery. Of further note in their work is the collaborative management of *line of sight* between the surgeons and the imaging systems. In their observations, significant work was done by the surgical team to monitor the gaze of the surgeons so as to time their actions to avoid breaking the line of sight between surgeon and monitor at key points in the surgery. Such concerns have important consequences for our understanding of touchless interaction in these settings in how they introduce new line-of-sight issues between the clinicians and the sensing cameras of the Kinect.

In addition to these specific studies of surgical practice relating to the potential for touchless imaging in surgery, are some more general social studies of surgical practice that help inform our work (Cassell, 1987; Fox, 1992; Goffman, 1961, 1983; Hindmarsh and Pilnick, 2002, 2007; Hirschauer, 1991; Katz, 1981; Lammer, 2002; Wilson, 1954). Of particular significance are those where there is an analytic concern with asepsis (e.g. Katz, 1981) and the social organisation of teamwork (e.g. Hindmarsh and Pilnick, 2002, 2007; Wilson, 1954). Drawing on arguments by anthropologists such as Mary Douglas (1966), Katz articulates the ritual nature of practices in the operating theatre in relation to the boundaries between different "*realms of cleanliness*" – sterile and contaminated – which involve the constrained and carefully choreographed movements among the clinicians when working within confined spaces in the theatre. Central to the management of boundaries between sterile and non-sterile is an organised distribution of labour between scrubbed and non-scrubbed personnel within different spatial zones of the theatre. This distribution of labour and hierarchic organisation is further considered in the work of Hirschauer who refers to the larger surgical team as the "surgeon-body". The surgeon-body is argued to extend the surgeon with complementary and "additional 'right' and 'left' hands'" and coordinates itself through words and gestures, and the anticipation of actions by the clinical team.

Such professional anticipation of action in the theatre is a key theme within the organisation of practice that is crucial to some of the concerns we discuss in our later

fieldwork. This is further detailed in the studies of Hindmarsh and Pilnick (2002, 2007) which highlight how team members recognise the *"trajectory of action"* to make inferences about the clinicians' future actions and intentions based on their current actions. Referring to this "tacit order of teamwork", Hindmarsh and Pilnick articulate how the embodied gestures, talk and action of teamwork are not only understood, but how they are produced and made visible in relation to the spatial and material arrangement of people, objects and artefacts in the theatre. Such issues are elaborated further in the work of Svenssen et al (2007). Here the concern is with the handling of instruments in the theatre and more specifically the ways in which instrument exchange is organised in a timely manner. This seemingly mundane activity of instrument exchange is a highly complex and contingent accomplishment that relies on a practiced familiarity and engagement with the tasks at hand. Through this, clinical team members can be aware of the conduct of others and its implications for their prospective coordination. They are able to anticipate particular actions and configure their exchange behaviour in relation to the contingent trajectories of action. Such arguments, then, may point us to a different set of considerations in the way we view and understand the new contingencies of gesture-based interaction mechanisms in the operating theatre and how they are accommodated in these coordinated trajectories of action.

   The embodied nature of these practices in surgical teamwork leads us to a final point of discussion concerning our analytic orientation to the treatment of medical images, which is informed in particular by seminal texts of Goodwin (1994, 2000) and Lynch (1990). In Goodwin's (1994, 2000) discussion of professional ways of seeing - what he terms *professional vision* - he argues how it is important to concern ourselves not simply with representations and other visual phenomena as coherent and self-contained entities in their own right. Rather, what is crucial to our understanding is an articulation of the ways that images and representations are situated in practice and the part they play in the production of meaningful action within particular settings. That is, how are they constructed, manipulated, attended to, pointed at and used by participants as constitutive of practice. In this respect our concerns are not just with the visual as a spatially represented concern but also with the temporal qualities bound up in the action unfolding around them. In a similar vein, Lynch (1990) argues that we need to shift attention in our analysis of visual representations away from simplistic and individualistic cognitive foundations and concerns with retinal and mental images. Rather, our focus needs to be on the "externalised retina" of the settings in which the scientific image is impressed and the range of practices through which meaningful objects and relationships in the image are collaboratively made visible for analysis. Of particular significance for our concerns is that the talk and practices occur across multiple sites of visual representation (e.g. a photograph and diagram) which continually and iteratively re-shape what is being seen in each representation. This is particularly significant when we extend this to surgical settings for which multiple sites of visual representations and the patient's body are brought together and made meaningful as a directional chain of representations.

   More specific medical examples of such accounts of the dynamic, embodied and interactionally-organised processes of image reading can be found in the work of Hartswood et al (2003) and Alač (2008). Hartswood et al's study (2003) explores the repertoire of representational manipulations and techniques used in diagnostic radiology to make professionally relevant things visible and interpretable. For example, the clinicians reading the mammograms may measure feature with rulers, pens or their hands, compare views, align scans, trace along represented tissues strands as well as annotate and mark up scans. Hartswood argues that such features are not simply improvised

workarounds in the performance of scan reading but rather are an integral part of the ecology of practice that is constitutive of reading mammograms. Of further significance is that these practices of reading are more than just for the facilitation of individualised forms of reading and interpretation. Rather such practices are inherently social and part of the routine ways through which features, interpretations and uncertainties are made visible and accountable to others and allowing the intersubjective calibration of diagnosis with others. In Alač's work, the focus is on the analysis of fMRI (functional Magnetic Resonance Imaging) images. Alač discusses the embodied nature of fMRI reading where gestures, body orientation and gaze of the practitioners turn the physical space into a site of meaning production. In this respect the multiple semiotic fields of the digital representation (fMRI images) and the one inhabited by the material bodies are brought together in the production of meaning. Of particular note here is how this coordination of the two fields has a "salient temporal dimension". An example here is the strategy of rapid alteration of serially organised images to create an animation like appearance of motion in which differences in the images are perceived and understood as a dynamic whole – the way of manipulating the image is the way of seeing. A second issue is the enactment of motion in the space of the material body in the form of gestures around the digital image. The dynamic coupling of gestures to images on screen enacts imagined transformations of the image. For example, the readers may make squashing or shearing actions in front of the scan image, or they might mime rotation of an imagined 3D version of the brain above the 2D fMRI image. By aligning these embodied actions with talk, the scientists are able to make relevant features of the images visible and accountable to their colleagues.

Drawing on the insights and theoretical orientations presented in the literature, we present a Kinect based system for the touchless control of a medical imaging technology in the particular context of vascular surgery. Our primary concerns here are not with particular claims of technical novelty over and above the systems outlined above. While there are certain novel features of the system design that warrant discussion in relation to the clinical practices of vascular surgery, the system also shares a number of core motivations and characteristics with the aforementioned systems. Our primary concern in this paper is with understanding the significance of these systems in the context of clinical practices. To this end, we report on an ethnographic study of the system being used during in actual surgical procedures. In particular, we discuss how the touchless system comes to bear on the clinicians' professional ways of seeing and the ways in which it becomes constitutive of surgical practice. We begin with a look at the clinical setting for our work both in terms of the environment and key elements of the vascular procedures in question. Having set this scene we move on to present the system in more detail, both in terms of its role within complex aneurysm repair procedures and in terms of a design rationale for the touchless features of the system within this clinical context. We then move on to present the findings from the study and discuss the implications in relation to the broader use of touchless imaging systems in surgery.
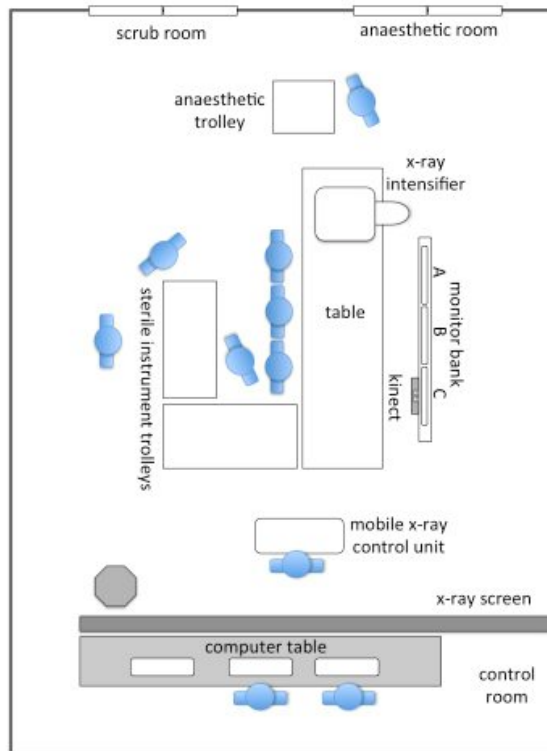
3. Vascular Surgery and the Endovascular Suite

Prior to the development of the system we undertook a series of observations in the Endovascular Suite of a large National Health Service (NHS) hospital in the UK. Working closely with the vascular surgeons we were given access to view 3 complex aneurysm repair procedures with a view to understanding the collaborative organisation

of their particular work practices. In light of the arrangements of the clinical team and equipment, the most suitable vantage points for the observations were along the back wall (bottom of Figure 1b) of the Theatre and behind the X-Ray screen. We would position ourselves as necessary to obtain suitable views for understanding key points of interest at particular moments during the procedures. During the observations, we had the opportunity to interact with the surgical team (surgeons, radiologists, radiographers and operating assistants) providing us with additional explanations of the pathologies and procedures being undertaken, details of interactions with particular equipment, interpretations of images, and explanations of talk and actions. As well as taking field notes, we collected audio and video recordings of the procedures. Within the contextual understanding of the *in situ* observations, the recordings enabled a more detailed, reflective and systematic analysis of the unfolding actions of the team than was possible during in-the-moment observations.

Within the procedures, the focal point of the suite (see Figure 1) is the x-ray table on which the patient lies. Other medical equipment and furniture are organised around the patient table: an x-ray image intensifier, 2 sterile instrument trolleys, the anaesthetic trolley, mobile image intensifier control unit, and radiation screen. Various shelves and storage facilities are positioned around the room to house a variety of vascular surgery equipment such as wires, catheters, and stents of various dimensions. Above the x-ray table is a bank of 5 monitors attached to a mobile lever arm allowing them to be positioned in terms or height orientation and distance from the clinicians who stand on the opposite side of the table. Three of these monitors are positioned in a row and oriented to the clinicians. The left-most screen (screen A on Figure 1(b)) is used to display the live fluoroscopy images, while the middle screen (screen B on Figure 1(b)) displays higher resolution reference images ("roadmaps") from angiographic runs. The right-most screen (screen C on Figure 1(b)) is used to display a 3D rendering of the aorta (built up from preoperative CT scans) overlaid on top of the live fluoroscopy images. Above the row of 3 screens are two additional screens, one showing the position and angle of the x-ray scanner and the other a duplicate of the live fluoroscopy image. These face out towards other members of the clinical team. A control room occupies one end of the theatre (RHS of Figure 1b) – a ¾ height protective glass divide is used to separate the control room from the radiation but being open at the top remains open to the acoustics and communication from the main part of the theatre. Various monitors and PACS (Picture Archiving and Communications System) machines are situated in this control room and provide control points for the various imaging systems such as the fluoroscopy and live overlay images.

(a)

(b)

Figure 1. The endovascular suite: (a) view of the suite from the control area; (b) plan view of the suite showing approximate layout

The specific area of vascular surgery that was the focus of our work was that of complex aneurysm repair in the aorta. These procedures are based around the insertion of custom-made stent grafts into the aorta. A schematic of the repair with the deployed stent in the aorta can be seen in Figure 2(a). Of significance with these stents is that they have holes (fenestrations – see Figure 2(b)) and/or branches (limbs off the main stent). A core feature of the surgery is the alignment of these fenestrations and branches with the corresponding openings to the arteries that branch off in particular directions from the main aorta. In some complex cases there may be up to four fenestrations/branches that all need to be correctly aligned in 3D space with the anatomy of the patient's aorta.
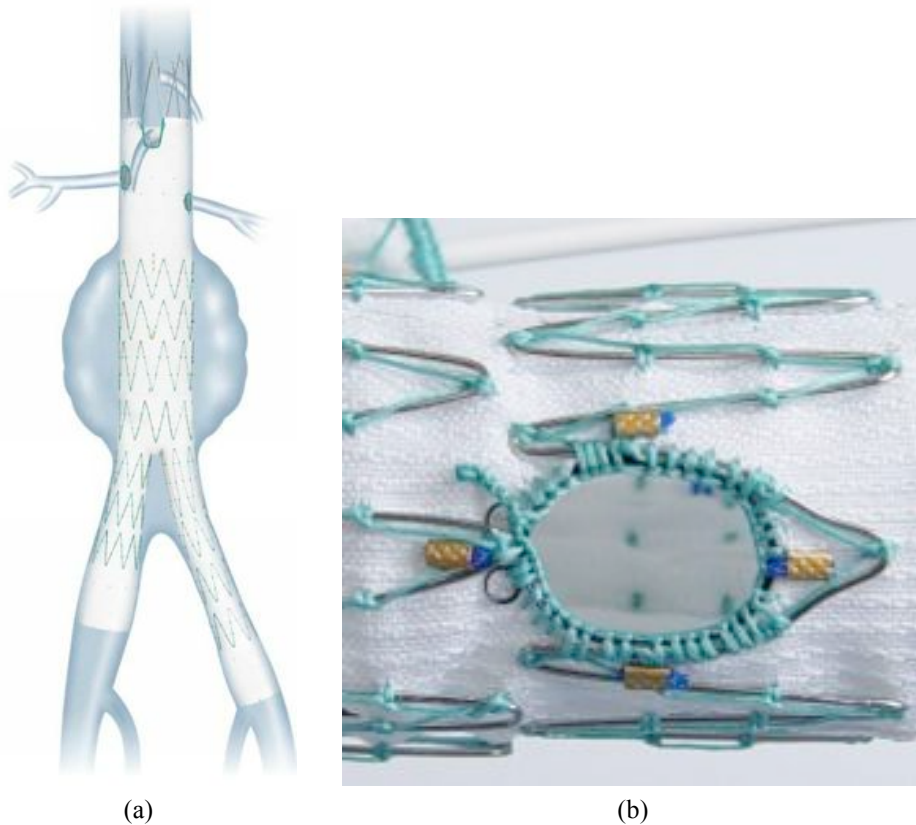

(a)   (b)

Figure 2. (a) A schematic of the fenestrated repair; (b) fenestration (opening in the stent)

In order to position these stents the surgeon introduces a sheath into the arteries. The sheath allows guide wires and catheters and finally the stent devices to be introduced into the arteries. To direct them in these activities, the clinicians use fluoroscopic imaging techniques in which x-rays provide real-time moving images of the aorta, guide wires and stent devices as they are manipulated into position. This kind of imaging shows only radio opaque materials, such as the bones, the guide wires and specially designed unique markers on the stents that help identify each fenestration/branch (e.g. see four radio opaque markers around the fenestration in Figure 2(b)). In Figure 3(a), we can see an example of the fluoroscopy image showing the catheter wire and revealed groupings of

11

radio opaque dots. Each configuration of dots uniquely identifies particular locations on the stent such as the entry points for particular fenestrations. The shape of the marker configurations (and corresponding perspective distortions) and their position depicted in the 2D image provide an indication of the position and orientation of the stent within the aorta and consequently with respect to the branching blood vessels to which they need to be aligned. The live fluoroscopy images are typically shown on Monitor A of the monitor bank (see Figure 1b). As can be seen in Figure 3(a), there are no blood vessels visible in the live fluoroscopy image. The reason for this is that the blood vessels are not radio opaque and hence do not show up on the x-ray. To reveal them a radio opaque contrast dye has to be injected into the arteries. As the dye flows through the blood vessels, a "run" of x-ray images is taken to reveal the a 2D depiction of the anatomy of the arteries. Images from this run are then selected by the surgeon to use as reference maps/images that are cross-referenced with the real fluoroscopy images to help interpret the position of the wires and the stent devices. These static reference images are typically displayed on Monitor B of the monitor bank in Figure 1b.
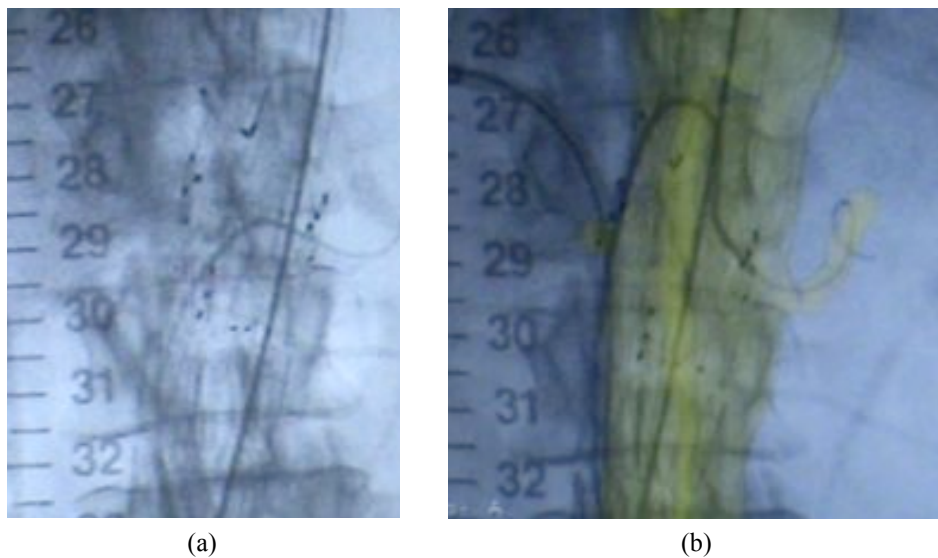

(a)   (b)

Figure 3. Fluoroscopy image: (a) showing catheter wires and radio opaque markings on the stent; (b) with 3D overlay of the aorta

As well as the live fluoroscopy and x-ray runs, the clinical team in this particular hospital utilises an additional imaging resource of a 3D volumetric rendering of the aorta built up from pre-operative CT scans (see Figure 3(b)). This 3D image is overlaid on top of a screen capture of the live fluoroscopy image and presented on Monitor C of the monitor bank (see Figure 1b). This particular imaging system offers a number of potential values to the clinicians during these procedures. First of all the 3D rendering of the aorta can provide additional information about the position and direction of branching arteries and catheters, which may be hidden or ambiguous in the 2D fluoroscopy and x-ray reference images. Second, there are times where the contrast dye does not reach all of the blood vessels leaving them unrevealed in certain x-ray views. The 3D overlay offers a potential resource for helping resolve particular features of these clinically constrained x-ray reference images.

4. Touchless Control of the 3D Overlay with Kinect

Prior to the development of our system, a dedicated controller (operating a PC located in the control room area of the Theatre) manipulated the 3D overlay. The clinicians would issue particular verbal instructions to the controller who would then implement the requested commands. In light of the arguments outlined in the introduction, we developed a touchless interface to the system to allow the clinicians to control the 3D overlay directly while at the patient table. The system utilised a Kinect sensor positioned just below Monitor C on which the 3D images were displayed (see figure 1b). Interaction used a combination of hand gestures and voice commands, with voice for discrete commands/mode changes and gestures being used for the control of continuous image parameters.
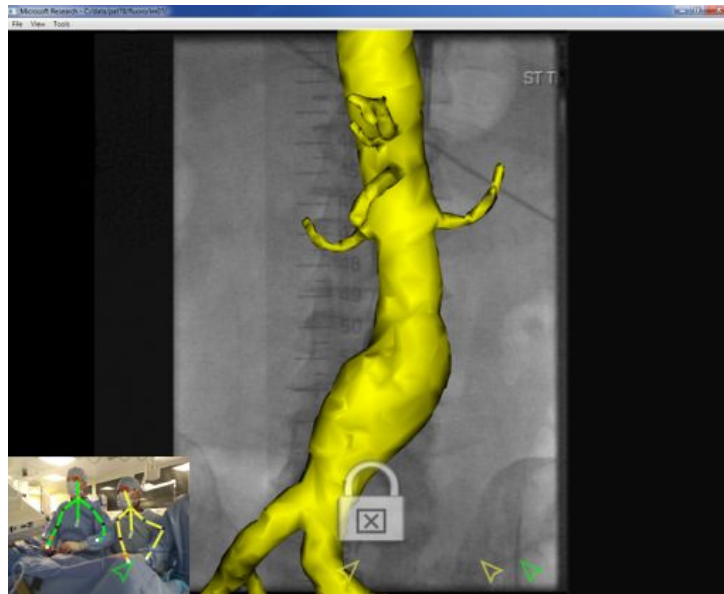

Figure 4. The touchless interface to the 3D overlay system

The particular details of the interface and vocabulary of the gestures and voice commands were informed by our fieldwork observations and developed iteratively with participation from the clinicians throughout in the process. We moved from functional specs through to increasingly higher fidelity prototypes involving review by the clinicians at each of the key points. The clinicians were able to try early prototypes out both in the lab and subsequently in the endovascular suite (outside of any surgical procedure). Feedback from these sessions was used to refine the system design for the purposes of deployment. The final system was organised around two modes, *locked* and *unlocked* mode.

Locked mode. This mode was the default state for the system in which much of the image manipulation functionality is turned off to avoid inadvertent triggering of system commands caused by body movements and gestures not intended for the control of the system. In this mode, the movements of the clinicians hands in the x and y planes were used to control the position of pointers on screen (see Figure 4(a)). The aim behind this was to allow the clinicians to use the onscreen pointers to point to particular parts of the

13

image in the context of their talk. Where there is additional functionality available in this mode, it is designed to be performed using a single hand or with voice. Such functionalities were allocated on the basis of the clinical need to be operated while potentially holding surgical instruments with the other hand. First of all, the system allows a mark to be placed on the surface of the rendered aorta. Typically, this is used to mark the entrance to a branch off the aorta. With the pointer positioned in the desired location for the mark, the verbal command "*Kiko – Place Mark*" is used to place a small dot on the surface of the aorta and a line traced from the mark to the corresponding point on the underlying fluoroscopy image (see Figure 5(a)). A mark can subsequently be deleted by issuing the command "*Kiko – Delete Mark*". Second was the functionality to control the *opacity* of the 3D overlay to reveal the underlying fluoroscopy image and understand correspondence between the two. To do this, the right hand is used to raise the pointer to the top of the screen, at which point a slider bar is revealed. By moving the hand in the x plane, the opacity slider is moved to the desired level.
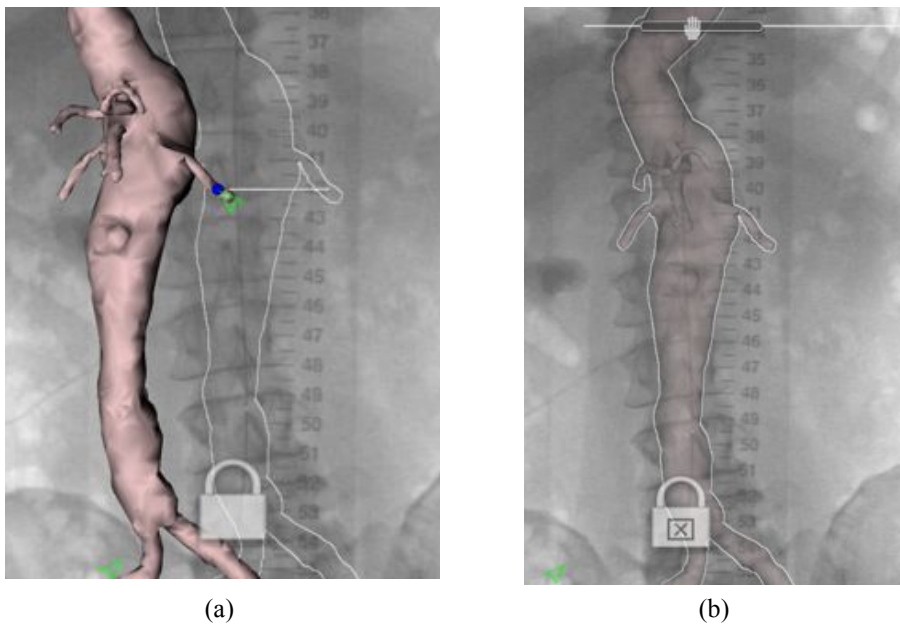

(a)   (b)

Figure 5. Screen shots showing (a) placed mark and ray trace; (b) opacity slider

Finally, a verabal command was used to return the image to its origin position. The command was "*Kiko – X Position*". The x-position was an understood term among the clinicians in which the image would be returned to its origin position (where there is no rotation, zoom or pan applied). For reasons of perspective, this is the only position where there is a fully accurate correspondence with the underlying fluoroscopy image.

Unlocked Mode. In order to have the full range of gestures recognised, the controller must enter the unlocked mode. This is achieved by raising the right hand (to signify the controller to the Kinect) and saying the command "*Kiko - Control*". When the mode is activated, a narrow green/yellow border appears around the edge of the screen corresponding to the colour of the tracked skeleton in control. Once in this mode, the following continuous parameter functionalities are available:

- *rotate* - movement of the right hand (palm forward) in the x-y plane controls rotation of the overlay and underlying fluoroscopy around the x and y axes.
- *zoom* - two hands (palms forward) spreading apart in the x-y plane controls increases zoom, while bringing the hands together decreases zoom.
- *pan* - two adjacent hands (palms forward) moving together in the x-y plane controls panning along the x and y axes.

The use of one or two hands for these functionalities was judged clinically appropriate by the clinicians. Two hands were used when particular functionality was deemed more likely to be used at points in the procedure where clinicians were able to put down tools.

To leave this mode and return to the locked mode there were two alternatives. First, the verbal command "*Kiko - Lock Image*" could be used. This command would retain the image in its current state. Second, by issuing the *"Kiko – X position"* command in this mode, the image would return to the x-position and additionally revert to the locked mode.

Another particular characteristic of the gestures is that they are designed to constrain movement within the area projecting in front of the clinician's torso. Again there are clinical considerations at play with such gesture design. Perhaps most importantly is an adherence to ritual practices relating to sterility. The area above and below the torso are designated as non-sterile in strict sterile practice and so surgeons constrain their arm movements to within the sterile torso area. The gestures were designed to adhere to these particular constraints of movement arising from sterile rituals. A second clinical reason for such decisions is that the tableside is a crowded area (cf. Johnson et al, 2011). Restricting the gestural movements was intended to enable gestures to be performed when in close proximity to other clinicians. To help achieve this the system employed a *clutching* mechanism in which hands are withdrawn from the imaginary x-y plane back towards the body. When the hands are close to the body the system ignores movement and allows repositioning of the hands before moving back in to the active recognition plane – for example to allow a continuation of the zoom gesture without the need to stretch the hands beyond the projected plane of the torso. The user would move their hand forward in front of their chest, then move their hands apart to begin the zoom, then move their hands back towards their chest to reposition them together to start the gesture again and continue the zoom manipulation.

A final point in the design concerns the provision of feedback to the controller – again an issue highlighted in (Johnson et al, 2011). Of particular note here, as shown in Figure 4(a) is the picture-in-picture display of the Kinect camera view and tracked skeletons. The picture highlights how the skeleton, as perceived by the Kinect, maps on to particular components in the live image. Additional icons are used to show various other features such as when the image is locked (see Figure 5(a)) or in the X position (see Figure 5(b)).

Having developed and refined the system with the clinicians, we then installed it in the Endovascular Suite for the clinicians to use during real surgical procedures. Some of the surgeons who had been involved during the development phase had some familiarity with the basics of the system having had chance to use earlier iterations prior to the installation. With the system installed, the lead surgeon took the opportunity to further familiarise himself with the system immediately before the first procedure while the theatre was being prepared for surgery. For the other surgeons (not all of whom were

present in the first procedure), they took the approach of learning on the job, trying it out just prior to a procedure or during periods of downtime in a procedure. These efforts were typically supported by prompts from the lead surgeon who acted as a champion for the system. In this sense, familiarization with the system was something that was integral to the ongoing work of the surgical team rather than based on a dedicated training session. With the system installed, we observed a further 5 complex aneurysm procedures. As with the fieldwork prior to the system installation, we augmented the observations with in situ interactions with the surgical team (surgeons, radiologists, radiographers and operating assistants) to help interpret the procedures and actions being undertaken. Video recordings were again made of the procedures from the same vantage points outlined earlier.

The analytic orientation we adopted was with how these images and representations, with these new touchless interaction capabilities, became embedded in surgical practice and production of meaningful action within these settings. Our analytic attention then was not simply with the touchless system interactions in and of themselves but rather in how they were situated within the broader spatial and material arrangement of team members, objects and artefacts constitutive of the collaborative practices of these procedures.

5. Findings

5.1 Dynamic control of images

In order to provide us with a way into the findings let us consider the following remark from one of the vascular surgeons who had used the system for various procedures:

> *"Until recently I was relying on calling across the room to a controller – left a bit right a bit, yes that's where I want it that's quite difficult thing to do when you are in a long complicated operation and concentrating hard on the technical aspects and trying to communicate with a third part often with quite subtle anatomical information so this Kinect system allows me to get control of the system without interfering with the flow of the operation and without me having to engage my brain to communicate to a third party.... I can interact with that my self with the Kinect system I can pick those points precisely – rather than issuing commands to someone with advanced anatomical knowledge operating the system – with the Kinect system I can go straight to the anatomical feature without having to translate into their language. It's much easier to engage and disengage. It's there on standby all the time. I've noticed I use it several times more in the operation than I would if you are having to get the attention of the third party and going through the slightly laborious procedure of trying to convey what projection or image you want. Some of it is gaining further information. Sometimes its just reassurance and sometimes it's a positive steer for where you want to be."*
> *[Vascular Surgeon]*

While such remarks articulate the values of the system to the clinician in high level terms that are broadly in line with our initial motivations for developing the system, a closer inspection starts to hint at the deeper significance of the remarks in thinking about the value of the system. Of importance here are the issues of communication between

clinician and operator and what this means for the ways that the images come to be manipulated and used. Not only is the communication between clinician and operator laborious and cognitively demanding in the context of other features of surgical practice, it is also constrained by what can reasonably be expressed *to*, and understood *by*, the operator in control. Much of what could be gleaned from the manipulation of images and much of what the clinician can see in the images in relation to the other imaging resources and the body, cannot always be neatly packaged up into easily expressible forms that could be understandable to the controller.

As such, in the operator controlled scenarios, the use of the overlay can be characterised as the production of image "states" – end points of a manipulation process rather than an ongoing interleaving of image manipulation into the interpretative, discursive and explanatory practices of the clinicians. In this respect the surgeon requests a manipulation and inspects the image once the requested state is achieved. Of significance in the quote is the sense the clinician has that he has been using the overlay imaging system more often during the procedures as a consequence of the direct control over the system enabled by the device. What this points to, then, is something more than just having a more efficient mechanism for achieving the same set of static image states. Rather by giving the surgeon direct control over the interactions and removing the imaging resources from the laborious bounds of an instructed communication, the imaging resources can be mobilised in new ways in the context of practice. As we shall see, such concerns relate on the one hand to the individual clinical interpretation, while on the other, to the ways the imaging resources can be mobilised in collaborative context.

In order to illustrate these issues let us consider the following sequence illustrated in Figure 6. The three clinicians[1] are lined up at the patient table and engaged in discussion around the fluoroscopy image. In the midst of the discussion, the chief surgeon, who is standing rightmost in the images, decides to refer to the 3D overlay (Figure 6(a)). As he is looking at the image, he sweeps his raised hand across to the left to fade out the overlay to full transparency (Figure 6(b)). His hand remains raised as he momentarily inspects the image before sweeping his hand in the opposite direction to fade overlay to partial opacity, then back out, then back in to full opacity (Figure 6(c)). He repeats this several times, back in and back out (Figure 6(d-e)), inspecting the image and all the while keeping his hand in position ready to continue the gestures. All the while he is doing this, his attention is on the overlay screen. As he repeatedly manipulates the opacity he is also engaged in discussion, commenting as he talks and listening to the response of the other clinicians. Having performed these repeated actions several times he begins to fade out again (Figure 6(f)). As he starts the gesture, his gaze shifts to the fluoroscopy screen (Figure 6(g)). As he turns his head, his hand pauses the motion of the gesture and remains in position in mid air in preparation to complete it. When his gaze returns to the overlay screen (Figure 6(h)), he continues the gesture to fade out once again (Figure 6(i)) before finally fading back in to full opacity.

---

[1] The other two clinicians are standing close to the left of the chief surgeon - the most salient figure in the images - and as such are partially obscured by him.

Figure 6. Dynamic control of the image

What we see in this vignette, then, is not so much a concern with the simple production of an image end state for inspection but rather the continual and dynamic manipulation of the image in the context of medical interpretation. Previously, when they had been instructing a controller, they would typically work with the production of one or two end states ("can you fade in please" – inspect – "fade out"). With the Kinect controller, we see how the surgeon fades in and out many times in succession combined with simultaneous inspection. By repeatedly fading the overlay in and out here, the surgeon is able to actively construct a visual correspondence between features of the overlay and features of the underlying fluoroscopy image in order to interpret what he is seeing in the fluoroscopy. This is not simply a question of looking at one image state and then the other but rather about creating an animated juxtaposition between the different layers of the image. What is of particular significance here are the nuances of such dynamic manipulations in terms of the particulars of parameter adjustments and the details of their timings as they constitute the interpretive and collaborative work being performed. Inspection and attentional shifts of the surgeon within image layers, between image layers and across the other imaging displays can be intimately bound to the imaging manipulations being performed as the interpretation evolves. These timings are illustrated nicely, for example, as the surgeon pauses a manipulation to look across the screen but maintaining his hand in position in preparedness to continue the manipulation when he shifts his attention back to the overlay screen. What is being enabled here, then, is not simply a more efficient way of performing the same actions they would otherwise have done through instruction of a controller. Rather, in enabling the direct control over the imaging they are performing a richer set of interpretive practices and constructing *new* ways of seeing.

5.2 Dynamic control in the context of talk

Such interpretive work though is not simply the individual concern of the individual surgeon. Indeed, as we have outlined earlier, our concerns with medical images are not so much as representations to be manipulated, but rather with how such images are situated and made meaningful through the ways they are constructed, manipulated, attended to, pointed at and used by participants as constitutive of collaborative surgical practice. A particular feature of the above vignette, then, is the way that the image manipulations of the surgeon across the multiple image sites are situated in the context of ongoing discussion with the colleagues to his side. Having dynamic control over these image manipulations is a key feature of the way such collaborative practices are organised. To understand this further we can again draw on Goodwin's work and in particular the treatment of pointing as a situated collaborative practice (Goodwin, 2003). For Goodwin, pointing is constituted as a meaningful act through the *mutual contextualization* of a range of semiotic resources. These resources include the visible act of pointing, the talk that elaborates and is elaborated by this pointing, the properties of the space targeted by the point, the orientation of relevant participants toward both each other and the locus of the point and finally the larger activity within which the act of pointing occurs. In the vignette, the dynamic manipulations of the images in the context of talk are being used as a similar act of mutual contextualisation. Having this control over the timings of the manipulations enables the surgeon to synchronise the image manipulations and the talk such that the image manipulations could both elaborate and be elaborated by the talk. In addition, control over the timings of these image manipulations is key in how they are used to help orchestrate the orientation and attention of the collaborators to particular features of the various images of significance to the ongoing work. We can see this illustrated in the following episode.

Figure 7. Dynamic manipulations of the images in the context of talk as acts of mutual contextualisation.

In Figure 7, S is a radiologist who is currently attempting to manipulate the catheter into the correct position. The chief surgeon T is to the left of S and is overseeing the procedure, while surgeon C, behind them both is assisting in the procedure. All three are initially focused on the live fluoroscopy screen, with S using it to guide the

movement of the catheter while the surgeons monitor what is going on. With S being less experienced, he is struggling with the insertion. T ascertains this struggle and why it may be happening, from viewing the movements of the catheter on the screen. He offers his interpretation:

> T: *Not really enough angle on it – I think you are still in the iliac with your catheter. I don't think the catheter is above the bifurcation, which is the way we planned it. I think you might need a C2 to get round it see the way the limb has sprung open. Its not attached to the other side so its gone right the way across. Can we have a C2 please?*
> C: *yeah because we thought the other was just in the origin of the iliac*

At this point, T realises that some further explanation would be useful to help S interpret what is being seen in the fluoroscopy and understand the planned course of action. He decides to invoke the overlay to facilitate this.

> T: *[raises his right hand]*
> T: *[to system] Kiko - Control [engages the system]*
> T: *[to system] Kiko - Lock Image [locks the image]*
> T: *So the…*
> *[T gestures to fade in the image then gestures to move the cursor to the point of interest]*
> T: *[to system] Kiko – Place Mark.*
> T: *So that is where the aorta bifurcation is…*
> S: *[leans in over the patient to look more closely at the mark just made]*
> T: *level with….*
> S: *Oh there.*
> T: *Yeah its level with - so that's the old one – about where that dot is. See the dot on the contra lateral limb – so it's not far below.*

So here, we see how T initiates his explanation with "*So the…*", and in doing this begins the process of mutual contextualisation with S. The beginning of the utterance is then accompanied by the fading gesture and the placement of the mark. As the mark is placed, T continues with the utterance. Control over the timing of image manipulation and utterance allows T to juxtapose the utterance to the image manipulation (placement of the mark). In doing so, the utterance and manipulation become bound together in the production of meaning, and drawing the attention of S to key features of the image in the construction of the explanation. What is important here is that the talk and the image manipulations are mutually constituted– they are designed in concert with each other. It is this interdependency in meaning production that is enabled by the clinicians dynamic control over the image manipulation in ways not feasible when dependent upon the 3rd party controller. As well as elaborating and being elaborated by the controllers talk, the timing of these image manipulations is also of importance in aligning with the talk and gaze of other clinical colleagues. Control over timing of the manipulations is also evident in relation to the talk and gaze of other clinical colleagues who are not controlling the image. We can see this illustrated in the episode shown in Figure 8. In Figure 8(a), clinician C has his hand raised to gesturally "grab" the opacity control. Having faded it to an appropriate position, clinician T is leaning in over C's shoulder to examine the image.
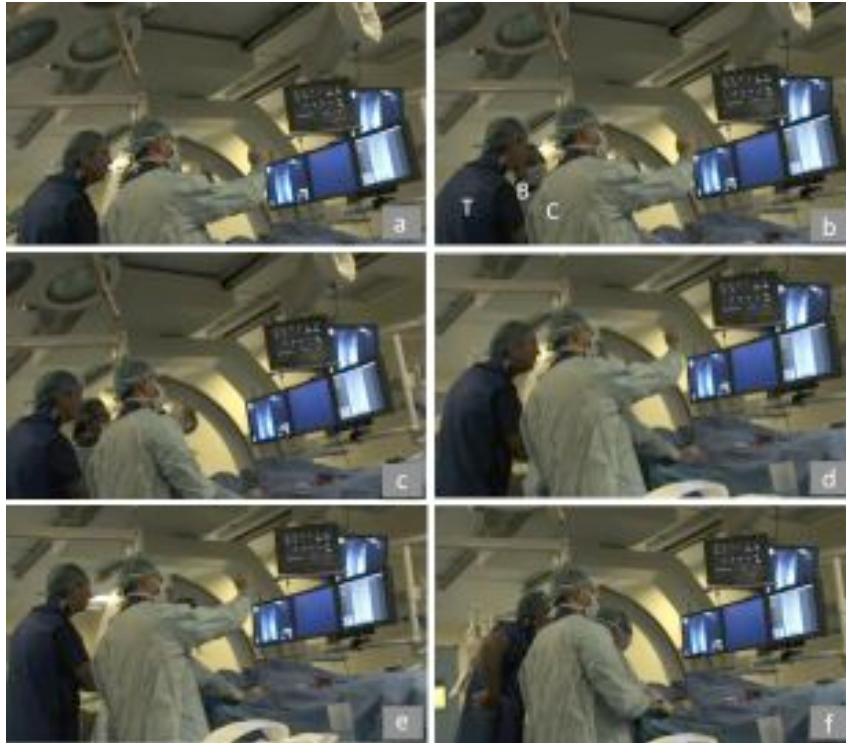
Figure 8. Timing manipulations in relation to the gaze of clinical colleagues

T: [to C] Ah good, the right…
B: [to T] it's a good position on the right.
*T: [turns to B agreeing – (*Figure 8*(b)) – C's hand remains in position for the fade gesture] The right renal is correct [inaudible].*
As T enters into discussion with B, C drops his hand to end his control over the fade gesture (see Figure 8*(c))*
T: … the left is the problem.
*T: [turns back and leans in over C's shoulder to look at the overlay] Can you fade again [*Figure 8*(d)]*
T: [verbalising out loud to himself and C] It's about 24… and a half.
[T continues staring at the overlay to confirm what he has verbalised; C maintains the raised hand in control of the fade]
*T: [*Figure 8*(e) T leans back away from the image and turns to addresses B; C maintains the raised hand in control of the fade] 24 and a half.*
T: [Leans to look at the fluoroscopy image; C drops his hand to end fade control]

In the vignette, then, clinician C is coordinating his gestural control with the gaze and verbalisations of the other clinicians. C understands when T is engaging with the overlay through T's bodily orientation combined with the content and direction of his verbalisation. When T's attention and talk are oriented to the overlay, C maintains the gesture and when T's attention shifts elsewhere, this too is understood and the gesture is dropped. What is notable about these behaviours lies in the broader ways that we come

21

to understand and conceive the importance of direct and dynamic control of these images by the surgical team as enabled by the technology. These extend beyond some of the more immediate cognitive concerns of image interpretation by a controller to encompass some of the more nuanced in situ ways in which these imaging resources can be mobilised in the collaborative context of surgical practice.

5.3 Control while using other surgical artefacts

As we highlighted in our earlier design discussions our understanding of these gestural systems is not simply a concern with how the controller can successfully deploy gestures to interact with the system. Rather, we must consider more broadly how the gestural system comes to be used in the context of other aspects of surgical practice and the theatre setting. A particular concern for us from the outset was with the engagement of clinicians' hands with other surgical instruments and the ways that this would affect the organisation of action around the imaging resources. In orienting to this concern, we highlighted the potential clinical significance of considering how our gestural vocabulary should be distributed across one or both hands. From our observations, we saw ways in which these design decisions played out but also some of the ways in which the surgical team organised their imaging interactions to overcome the constraints of the setting and artefact use.

Let us consider how practices were oriented to the one-handed control possibilities as illustrated in the following episode. As can be seen in Figure 9, the 3 clinicians are lined up at the table in the standard positions 1, 2, and 3. The consultant clinician overseeing the procedure is standing behind the 3 clinicians in an advisory capacity. A scrub nurse is to the right of clinician D. Clinician A (position 1) is carefully manipulating the catheter wire into position, being guided by the live fluoroscopy image. The catheter wire is long and extends across the length of the patient's body. Because of the tensile qualities of the catheter it can move and curl up and resist further down the shaft as the clinician is making fine-grained movements to get it in to position. When the wire becomes unbalanced or resistive it can affect the fine-grained control required by the surgeon to position things exactly. One of the key reasons for having the supporting clinicians along the table is to steady the wire and facilitate the control of the primary clinician. As Clinician D is interacting with the overlay image using both hands, the scrub nurse (SN) steps in to steady the wire that he would otherwise be holding (see Figure 9(a)).

Figure 9. Holding a catheter while controlling the system

At this point the scrub nurse needs to access something from the operating trolley and has to let go of the wire to turn towards the table. As a consequence of this, the wire

becomes unsteady, and in noticing this the consultant clinician A says: *"D, sorry, can you just steady that wire a sec"*. In this request, the consultant clinician is acknowledging that D is in the middle of something but highlighting that he needs to attend to the wire as soon as feasible but leaving it to D's judgement to time his actions accordingly. While still gesturing to the system with his right hand, D reaches down with his left hand to steady the wire (see Figure 9(b)). Still holding the wire, D then issues as verbal command *"X-position"* which returns the image to its correct registered view. Here then is an example of voice control in the context of the other hand steadying the wire – so its one handed and hands free control.

5.4 Proxemic Constraints and Distributed Control

There is a further feature of this vignette that is significant in the team's organisation of their imaging practices and in particular how this relates to other ways in which the setting and ecology of artefacts influence opportunities for interacting with the system. The point of note in the vignette is that the surgeon leading the procedure is not the person controlling the images in this particular instance of use. Rather, we see here how control is taken up by one of the supporting clinicians. Unpacking this further, we begin to get a richer sense of complexities involved in the deployment and use of these touchless systems and the additional factors that constrain and shape use in particular circumstances during the procedure. Of particular concern here are the spatial demands of the Kinect sensor in terms where one needs to be positioned to effect any system behaviour – and how these demands may vie with other features of the work and environment that affect the spatial configuration of the clinicians at certain points in the procedure. In this illustration from the vignette, the lead surgeon has assumed responsibility for manipulating the catheter and deploying the stents. For this, he needs to be positioned at the head end of the patient table to facilitate access and delicate control of the catheter wire and stent. The primary imaging resource for the surgeon in this activity is the real time fluoroscopy and as such the visibility of this display at this point is the main determinant of the positioning of the monitor bank. In consequence, the physically connected overlay display and Kinect sensor end up being positioned at the other end of the patient table. With the lead surgeon is engaged in catheter and stent manipulation, not only is this cognitively demanding for him, but his body posture, orientation and hand positions are constrained in particular ways that prevent the body and hand positions necessary for engaging with the Kinect sensor as it is positioned. In light of this, it is the supporting clinician here who ends up being in the best position and orientation with respect to the Kinect sensor. As such, he adopts temporary responsibility for controlling the image manipulations. The point here is that there remain elements of the imaging interactions with the system that are a distributed and collaborative concern albeit within the clinical team. Overcoming the constraints of sterility in itself is not sufficient to render the distributed organisation of control an unnecessary feature of the work. Situated features of the work and spatial demands of the sensor continue to see control as a collaboratively organised concern.

On the face of it, such spatial concerns might point to a simple matter of adjusting the positioning of the sensor in appropriate ways and indeed this is something that was considered closely with the clinical team. Important here though, was that the sensor be positioned with respect to the overlay screen such that the controller could be in an appropriate orientation to both view and control the images. In this respect then, the sensor needed to be positioned directly below the overlay monitor. In our observations,

we did see some attempts to compensate for this positioning when lead surgeon T asked for the monitor bank (with sensor on) to be moved further up the head end on the patient table and angled slightly to give a better orientation to the overlay screen and sensor. Such positioning though could not be optimised fully around the demands of the overlay and sensor and around the spatial demands of only the lead surgeon. The positioning of the overlay screen and sensor are not independent of the positioning of the live fluoroscopy screen that is the primary focus. As such, adjustments to the sensor and overlay screen could only be made within the bounds of what is an acceptable view for the live fluoroscopy screen. Likewise, the screens and sensor needed also to be available to the other supporting clinicians at the patient bedside and, as such, any adjustments needed also to accommodate their viewing and potential control role during the procedure. Balancing these issues then was not a simple task for the clinicians but was nevertheless something of which they were conscious in the organisation of their practice. In light of this, they suggested that things might be made easier if the order of the screens on the monitor bank were changed such that the live fluoroscopy screen would be in the middle with the overlay and other reference screen either side. As a purely technical exercise, this should, in theory have be fairly trivial adjustment to make but in practice turned out to be organisationally more complicated with the requirement for dedicated engineers from the manufacturers to be involved.

The point here is not so much the particular circumstances highlighted in the vignette, but rather in their illustration of what it means to situate these touchless imaging technologies in the context of other artefact use during the procedures. On the one hand this raises a host factors that can be considered in the context of design. But on the other hand these features of the work and setting are things that simply need to be attended to and managed by the clinicians. From this perspective, what is of significance is how such contingencies and dependencies come to bear on the organisation of the imaging practices and the specific ways that the touchless control of the images can be deployed in the construction of a "professional vision".

5.5 Working to be sensed

In thinking about these gestural control systems, it is very easy to idealise how the gestures, as defined in an abstract gesture vocabulary, are enacted in practice. While we defined our gestures relative to the body of the controller, an important distinction needs to be considered in terms of the body as understood by the actor and the body as understood by the system. For the system, the understood body is in the form of an abstracted skeleton and it is in the way that this abstracted skeleton enacts movements and gestures through which the control of the system is determined. Ideally, there is a one-to-one correspondence between the abstracted body of the skeleton as understood by the system and the body as experienced and understood by the controller. The reality, though, is that such a correspondence is contingent upon a range of factors that compromise in various ways the fidelity of this correspondence in practice. This has significance in how we come to understand the bodily performance of the clinicians as they interact with the system. More specifically, what was apparent from our observations is that the surgeons are not simply performing defined actions with respect to their own understood body but rather engage in particular forms of work that make themselves understood and sensed by the system. We can see a simple illustration of this in the sequence depicted in Figure 10. In the sequence, the clinician slowly moves his

arms forwards then pulls backwards slightly until his hands are being detected (as indicated by the revealing of the cursors) at which point he can enact a zoom gesture. In making himself sensed, he is reliant on the visual feedback provided by the screen. What is important is the interplay between his body movements and the feedback of the system.

Figure 10. Clinician making himself sensed

A key issue here is that the particular setting presents a number of challenges for the system in its abstraction of the controller's skeleton. As we can see in Figure 11, showing the view from the system camera, the close proximity of colleagues to the clinician controlling the system means that the system does not always unambiguously isolate the body of one clinician from that of another. In this instance the abstracted skeleton extends from the controlling clinician across the arm of a colleague. What the systems sees as the controller's body in this situation, then is different from the controller's actual body. As a consequence, the enactment of gesture as understood by the clinician does not correspond with the enactment of gesture as understood by the system.

Figure 11. Tracked skeleton extends across multiple bodies

There are a number of ways that we might interpret and approach such breakdowns. For example, in highlighting these tracking difficulties arising from the particular circumstances of the operating theatre, we might pose an interesting set of challenges for the computer vision community. That is we can frame these difficulties as a limitation of the current technology to be solved with more sophisticated sensing and tracking solutions. Indeed there is strong merit in this as a technical challenge, but what is also of concern from our observations are the ways that the clinicians deal with these challenges and the work they do to make their bodies unambiguously understandable by the machine. To illustrate this work let us consider the following sequence in which Clinician T is attempting to place a mark on the overlay.

25

T: Kiko place mark
[Nothing happens – T checks the Kinect system camera view]
T: E I think that's you actually [E is in the picture and being picked up as the controller]
T: Kiko place mark [nothing happens]
T: I'll just block you out [T shifts across as he is saying it and sees that the system is no picking him up clearly]
G: I think its still on green
T: There we go
T: yeah I just walked in front of her.

In the sequence, T realises that the system is not behaving in the ways expected from the gestures and commands he is enacting. To ascertain what is happening, T checks the picture-in-picture view from the Kinect camera with the model of the skeleton being interpreted by the Kinect). He sees from this that the skeleton is not corresponding cleanly to his own understood body shape but rather is latching on to his colleague's body in the background of the scene. In order to overcome this, T shifts his position such that his body obscures E's body from the view of the Kinect sensor and the skeleton correctly aligns to his body. That is he positions and shapes his body not simply to enact gesture but to do so in a way that is understandable to the system. Through this then, we begin to see the importance of providing the picture in picture view of the Kinect camera view as an explicit design concern and why it is important to reframe our understanding in terms of these two distinct perspectives on the body. By reflecting this information back out to the controller, it provides a resource through which they can come to understand which particular features of the scene and setting may be interfering.

5.6 Collaborative Work of Being Sensed

Of further significance in the above vignette is that it starts to point to a key feature of this work of making the body sensed, namely that it is not so much an individual concern but rather a collaborative affair in which the bodies of other clinicians in the theatre come to have an influence. While T in the above sequence is able to deal with the presence of another body through his own movements, perhaps more important were the ways this work of being sensed was collaboratively achieved and organised and the ways in which this impacted other practices in these settings. To illustrate this further let us consider the sequence depicted in Figure 12.

Figure 12. The collaborative work of being sensed

In this sequence, clinician T raises his hand to get ready to assume control of the system (Figure 12(a)). The system does not appear to respond to the actions of T. The clinician C, next to T, notices that the system does not respond and adjusts his position slightly in an attempt to give T a clear space between their bodies (Figure 12(b)). Here, clinician C has assumed the intent of T from his gestures and that these should have had a particular effect that is not manifest on the screen. In making sense of the situation, he can see from the Kinect's camera view that his positioning may be having an impact and makes an adjustment to see whether this will resolve the problem. T continues to try and rotate the image, enacting the rotation gesture three times but without success. In response to this he then attempts a different tack issuing the verbal command "Kiko lock image" then immediately "Kiko control". He then successfully manages to imitate some rotation but it is clearly not as responsive to his movements as he would like so is still aware there is a problem. Looking at the picture-in-picture he says:

> T: It's picking up you isn't it."
> [moves his hands across as he plays around to get a better skeleton shape]
> C: Over here it is picking it up as me and that's what I was talking about – the dominance of the yellow [skeleton].
> T: Can you step off to the side or something or other.
> [C moves a little further away – see Figure 12(c)]

It is as this point that (see Figure 12(d)) that the radiologist, M, comes into the scene behind T and C, in order to see what is going on with the system. He is unaware that his presence will compound the difficulties they are experiencing with the tracking system. T continues to try and engage with the system.

27

T: Kiko lock image. Kiko place…. Ooops [the cursor moves as he places issue the command and so it is placed in the wrong place. [T moves his hand to get the cursor in the correct position]
T: Kiko place mark [it places a mark but again the cursor moves slightly so it is slightly off position. T adjusts and repeats]
T: Kiko - Place Mark; Kiko - X-Position.
[T then performs the gesture to fade but he is unable to successfully pick up the fade control and move it.
T: Why can I not grab that?
T: [looking at the skeletal feedback from the Kinect camera view he sees that it his skeleton is extending onto the body of the radiologist, M]. Oh is it M? (with humour).
[M steps back away from C and T (Figure 12(e))]
C: No it's still me
T tries to fade again
T: It's got me but I can't move the slider across any more… I should be able to slide that across. It's tracking my hand
G [imaging assistant]: But the other person is still in the background.
T: M it's you hovering in the background. It's sabotage.

Everyone laughs and M adjusts his position further so that he is hiding himself behind T but then shifts his head round so he can still see what is going on (Figure 12(f)).

T: (humorously) We were doing perfectly alright until…
T: Kiko - Delete mark, no Kiko Place Mark – that's better.
[T then turns to the clinician C]
T: C, why don't you come in and try and place the mark.
[C moves in and tries to assume control – he waves his hand]
C: It's not tracking me at all.
[C moves from side to side in an attempt to get tracked and waves his hand and tries a few things before saying Kiko control].

Noticing that there are still problems with the tracking, T can see from the camera view that despite M's attempts to hide himself from the camera, his presence is still interfering. T turns to M and points towards a "safe" location (Figure 12(g)):

T: I think it's you M. Can you just move over to the sides there… [to all, humorously] It's a way of keeping radiologists out of the operation I think. There we go look at that – we get rid of M… Definitely sabotage – anyway lets get on with it.
T: can you just take a step to your right C (Figure 12(h)).
C: mmmhhmm.
[T completes the fade (Figure 12(i))].
T: Ok that's what I want.
[C moves back into position].

The act of being sensed here, then, is not a simple matter for the controller. Given the properties of the sensor and tracking algorithm, there emerges a rather curious *zone of interference* around the controller, to the sides and behind, in which other bodies in the theatre come to impact on the system performance. This zone is not precisely bounded or

defined but is something that the clinicians come to understand through their use of the system. And it is by presenting back the state of the system to the controller through which they and those around can come to understand and define the boundaries of influence in the moment. What we see in the vignette is how the clinicians and theatre staff need to work collaboratively in order to accomplish this act of sensing. But we see, too, how as a consequence of this zone of interference, there is introduced additional constraints on the way the clinicians can organise themselves in the context of any engagement with the system. This in turn will shape the opportunities and ways in which the image manipulations become constitutive of surgical practice and the construction of a particular professional vision.

What is more, we see how background activities in the theatre acquire a new significance of which the team members need to become aware in the organisation and timing of their action. When members are unaware of these consequences, as in the above sequence, their previously innocuous actions can become problematic. Of interest, though, in our observations, was how clinical team members who weren't using the system began to learn about their potential influence and to accommodate this in their coordinated timing of action. A good illustration of this can be seen in the sequence depicted in Figure 13.

In this sequence, the scrub nurse, N, is standing at the instrument trolley behind clinician C as he is engaging with the system (Figure 13(a)). The nurse needs to get something from the patient table but as she turns, she is aware that C is gesturing to the system. With her head still turned towards C to monitor his activity she leans on the instrument trolley waiting (Figure 13(b)). She looks away while waiting (Figure 13(c)) but then notices C move forward towards the patient table indicating that he has finished with the system (Figure 13(d)). As he reaches the patient table she moves forward (Figure 13(e)) and leans over to pick up a catheter wire from the patient table (Figure 13(f)). What we see, then, is how in being visibly aware of his use of the system, the nurse organises the timing of her actions so as not to interfere with his use of the system.

Figure 13. Nurse times her actions to avoid interference with the system interaction

6. DISCUSSION AND CONCLUSION

In this paper, we have presented a study of a Kinect-based system that allows clinicians to interact with and manipulate medical images in the context of a surgical procedure through the use of voice and in-air gestural commands. The driving motivation behind the use of such interaction modalities was that they enable interaction with the imaging system without the need to physically touch any interaction artefact. Within surgical settings, such a property is of particular significance because of need to maintain strict boundaries between that which is designated as sterile and that which is designated as non-sterile. With more traditional touch-based mechanisms, this orientation to asepsis is manifest in the particular ways that imaging practices in the operating theatre are organised, for example by clinicians instructing third parties to manipulate images on their behalf, by de-gloving and rescrubbing, or simply by restricting and avoiding certain types of image interactions during the surgery. A central theme of the paper, and drawing on arguments in Science and Technology Studies literature, is that, over and above the image representations in themselves, such practices constitute a particular way of seeing in surgery. That is, in the context of asepsis, the socially constructed ways of seeing or *professional vision* are bound up in what is and what isn't enabled by particular touch-based interaction mechanisms.

A key feature of this argument is that imaging practices are neither prescribed nor entirely routine in their enactment during surgery. Rather, such practices are grounded in clinical judgement as to what might be a useful resource in the moment. As a resource, any imaging manipulation is part of a host of other information resources and data points available to the clinicians at any particular time. Consequently, any judgements to manipulate an image in a particular way is bound up both in what it might enable clinically as well as what it entails to enact it, e.g. time, effort, downing tools, exposure to radiation, exposure to contrast dye and so on. Such a perspective is critical to the way we have approached this space of touchless imaging control within these surgical settings both in terms of how we go about the design of these technologies but also how we come to understand their impact in use. The development of these kinds of technologies have typically been rationalised by recourse to often simplistic efficiency narratives, such as for example, obviating the need for time consuming rescrubbing practices or time consuming and clumsy communication of instructions to 3rd party image controllers. The argument continues that if we assume a clinician engages x number of times during a procedure then we can save x(scrubbing time) minutes and so on. The important point that we want to make here is not that these systems simply allow quicker ways of doing the same activities that would otherwise be performed. Rather, by overcoming these aspects of existing imaging practices, we lower the barriers to image manipulation such that they can and will be incorporated in new ways in surgical practices. Having such a focus on creating new opportunities for imaging practices, then leads us to different kinds of design decisions and rationale. For example, from a purely technical perspective we could have achieved touchless control over the system by using voice recognition alone. However, we explicitly restricted the voice modality to the control of discrete and articulable mode changes and commands. Gestures were used for the control of continuous imaging parameters to enable the more subtly timed and less articulable manipulations by the clinician. Similarly, in our gesture design, we looked for ways that would enable the clinicians to maintain sterile practices and facilitate their working in the close confines of other colleagues by restricting movements to the spatial area in front of the torso. Again the attempt here is to reduce potential barriers that might restrict opportunities to use the system at particular points during the procedure. Likewise, in the ways that we distributed our gestures across one and two hands, our aims were to open up certain opportunities for image manipulation while engaged with other clinical artefacts. Finally, our orientation to feedback provision was not just in the facilitation of the interaction mechanics but related more broadly to facilitating of the social construction of being sensed.

While these design features are well grounded in the perspective of practice, such a rationale only takes us so far in our ways of thinking about such systems. One of the key concerns of our work in this paper has been to start thinking more seriously about what it means for these systems to be used in context and more precisely to understand the new practices that are enabled by such systems as well as the practices necessary to accommodate them. In this respect, our aim has been to move beyond the overarching and accepted narrative around sterility that motivates this work to more fully articulate what this means. What we see in our fieldwork, then, is that by overcoming the constraints of asepsis and providing the clinicians with direct and dynamic control over the manipulation, we enable some important yet nuanced shifts in practice. In the first instance, we see this in the form of much richer manipulations of the image in the context of clinical inspection. In particular what is critical here is the control over the timing of these manipulations as they relate to the hidden work of image understanding.

Importantly, our observations situate such manipulations in the context of an array of imaging and data resources that are combined together in situ in the production of medical interpretation. Of significance is that the close timing of these manipulations allows them to be aligned with shifts of attention across both different layers of the imaging system and across the other displays. But this same issue of control extends also to the ways that these image manipulations can be effectively incorporated into the broader collaborative discussion that takes place among colleagues in their collective decision-making. The dynamic control over the timing allows the image manipulations to be interleaved and combined with talk to draw attention to particular features of significance – and it is through these that the images come to acquire their meaning. Notable too, is that by reducing the barriers to image manipulation, they come to be deployed in ways beyond their previous use such as to facilitate explanations to less experienced colleagues.

The fieldwork too highlights some important issues about the use of these technologies in the context of other aspects of the surgical procedure. Indeed while we oriented to such concerns in our designs, the extent to which these played out was a little different to what we had anticipated. We did observe, for example, some one-handed use of the system while holding catheters, but such instances of artefact interaction might be regarded as fairly minimal and typically done by one of the assisting clinicians rather than the lead surgeon driving the procedure. But what is of more general significance and applicability here is that the fieldwork revealed a number of additional complexities that relate to the opportunities for system use in the context of other artefacts. Some of this has to do with the constraints on position and bodily orientation, in light of, for example, actions being performed on the patient or the need to attend more closely to certain displays during the procedure. That is, positions and actions are organised with respect to multiple things in these procedures and not always ideally with respect to the sensor and imaging system that are the focus of the paper. A notable feature of the practices arising from this was that the control of the system became something that was not just performed by the lead surgeon as initially intended but rather was distributed across different clinicians according to their level of involvement in the procedure and spatial relationship to the display.

Related to these concerns, our observations have also drawn out attention to the work performed by the clinicians in order to make themselves sensed by the system. In this respect, clinicians do not simply perform abstract gestures but do so in a way that these gestures are understandable by the system. The distinction here is an important one and has pointed to the importance of making what the system sees visible to the clinicians interacting with the system. At play here is not a simple usability concern but rather arises from situating such systems within the operating theatre and the challenges to tracking systems this poses. Critical here is presence of other clinical colleagues working both adjacent to and behind the clinician interacting with the system. In the presence of these systems, the actions of these colleagues performing their work acquire new consequences because of their potential to interfere with the accurate body tracking of the controller. This highlights a broader set of considerations for the understanding of such systems, since in managing these issues, the clinical team have to reconfigure themselves in new ways. While they are able to manage this, the system does introduce additional constraints on how they can organise themselves in the context of their collaborative activities. Again, revealing what the system "sees" and understanding the impact of their bodies on the system is an important way of enabling these reconfigurations to take place.

Of additional general significance, though, in our evaluation of these systems is developing a further understanding of these reconfiguration consequences for surgical practice.

In presenting this work, we have focused on the domain of vascular surgery as it is practiced in a particular hospital setting. While there are specific features of this surgery in terms of the details of environment, artefacts, imaging systems and the collaborative practices undertaken, our aim has been to use these specifics as illustration of a set of more general considerations for the design and evaluation of these systems in other domains of surgery. In particular, our intentions have been to extend the narrative around these touchless imaging systems beyond the basic concerns with sterility and the introduction of work efficiencies to think more broadly about how and why they become constitutive of the social practices of surgery. This applies both to how we approach the design of these systems but also the ways that we need to understand them. In doing this, we have explicitly sought to articulate key properties of the technical components of our system, from the particular sensing and tracking capabilities through to gestural vocabulary and on-screen feedback and relate these to key properties of surgical practices. The field of touchless interaction in surgery, with several demonstrations of technical feasibility in existence, is now at a point where this kind of articulation is becoming increasingly important. Accompanying technical developments in the field then, we need to begin a much richer explication of system design rationale for particular implementation choices of touchless systems in these settings that is grounded in clinical practices. Further, what we hope to have demonstrated in the current paper is the need to build a much richer corpus of studies of these different systems in the context of real surgical practice.

REFERENCES

Alač, M. (2008) Working with Brain Scans: Digital Images and Gestural Interaction in fMRI Laboratory. S*ocial Studies of Science,* vol. 38, no.4, August 2008, pp. 483-508.

Cassell, J. (1987) On Control, Certitude, and the "Paranoia" of Surgeons. *Culture, Medicine and Psychiatry,* vol. 11, no.2, June 1987, pp. 229-49.

Cramer, H. S. M., Evers, V., Zudilova, E. V. and Sloot, P. M. (2004) Context analysis to support development of virtual reality applications. In *Virtual Reality,* vol. 7, nos. 3-4, June 2004, pp. 177-186.

Douglas, M. (1966) *Purity and Danger*. London: Routledge.

Ebert, L., Hatch, G., Ampanozi, G., Thali, M. and Ross, S. (2012) You Can't Touch This: Touch-free Navigation Through Radiological Images. *Surgical Innovation*, vol. 19, no. 3, September, 2012, pp. 301-307.

Ebert, L., Hatch, G., Ampanozi, G., Thali, M. and Ross, S. (2012) Invisible touch—Control of a DICOM viewer with finger gestures using the Kinect depth camera. *Journal of Forensic Radiology and Imaging*, vol. 1, no. 1, January 2013, pp. 10-14.

Fox, N. (1992) *The Social Meaning of Surgery*. Buckingham: Open University Press.

Gallo, L., Placitelli, A.P. and Ciampi, M. (2011) Controller-free exploration of medical image data: experiencing the Kinect. *CBMS 2011. Proceedings of 24th Symposium on Computer-Based Medical Systems, Bristol, UK, 27-30 June 2011*, pp. 1-6.

Goffman, E. (1961), *Encounters: Two Studies in the Sociology of Interaction*. Harmondsworth: Penguin.

Goffman, E. (1983) The Interaction Order. *American Sociological Review,* vol. 48, no. 1, February 1983, pp. 1-17.

Goodwin, C. (1994). Professional Vision. *American Anthropologist,* vol. 96, no. 3, September 1994, pp. 606-633.

Goodwin, C. (2000) Practices of Seeing: Visual Analysis: An Ethnomethodological Approach. In van Leeuwen, T. and Carey Jewit, C. (Eds) *Handbook of Visual Analysis*. London: Sage, pp. 157-182.

Goodwin, C. (2003) Pointing as Situated Practice. In Kita, S. (Ed) *Pointing: Where Language, Culture and Cognition Meet*. Mahwah, NJ: Lawrence Erlbaum, pp. 217-41.

Graetzel, C., Fong, T., Grange, S. and Baur, C. (2004) A Non-Contact Mouse for Surgeon-Computer Interaction. *Technology and Health Care,* vol. 12, no. 3, August 2004, pp. 245-257.

Hartswood, M., Procter, R., Rouncefield, M., Slack, R. Soutter, J. and Voss, A. (2003) Repairing the Machine: A Case Study of the Evaluation of Computer-Aided Detection Tools in Breast Screening. *ECSCW '03. Proceedings of 8th European Conference on Computer Supported Cooperative Work, 14-18 September 2003, Helsinki, Finland,* pp. 375-394.

Hindmarsh, J. and Pilnick, A. (2007) Knowing Bodies at Work: Embodiment and Ephemeral Teamwork in Anaesthesia. *Organization Studies*, vol. 28, no. 9, September 2007, pp. 1395-1416.

Hindmarsh, J. and Pilnick, A. (2002) The Tacit Order of Teamwork: Collaboration and Embodied Conduct in Anaesthesia. *The Sociological Quarterly*, vol. 43, no. 2, March 2002, pp. 139-164.

Hirschauer, S. (1991) The Manufacture of Bodies in Surgery. *Social Studies of Science,* vol. 21, no. 2, May 1991, pp. 279-319.

Jacob, M., Li, Y., Akingba, G. and Wachs, J. P. (2012) Gestonurse: a robotic surgical nurse for handling surgical instruments in the operating room. *Journal of Robotic Surgery*, vol. 6, no. 1, March 2012, pp. 53-63.

Johnson, R., O'Hara, K., Sellen, A., Cousins, C. and Criminisi, A. (2011). Exploring the potential for touchless interaction in image-guided interventional radiology. *CHI 2011. Proceedings of Conference on Human Factors in Computing Systems, 7-12 April 2011, Vancouver, Canada*, pp. 3323-3332.

Kang, H., Woo Lee, C. and Jung, K. (2004) Recognition-based gesture spotting in video games. *Pattern Recognition Letters*, vol. 25, no. 15, November 2004, pp. 1701–1714.

Katz, P. (1981) Ritual in the Operating Room. *Ethnology*, vol. 20, no. 4, October 1981, pp. 335-350.

Kipshagen, T., Graw, M., Tronnier, V., Bonsanto, M. and Hofmann, U. (2009) Touch- and marker-free interaction with medical software. *World Congress on Medical Physics and Biomedical Engineering, September 7 - 12, 2009, Munich, Germany*, pp. 75-78.

Lammer, C. (2002) Horizontal Cuts and Vertical Penetration: The 'Flesh and Blood' of Image Fabrication in the Operating Theatres of Interventional Radiology. *Cultural Studies,* vol. 16, no. 6*,* pp. 833-847.

Leganchuk, A., Zhai, S. and Buxton, W. (1998) Manual and Cognitive Benefits of Two-Handed Input: An Experimental Study. *ACM Transactions on Computer-Human Interaction*, vol. 5, no. 4, December 1998, pp. 326-359.

Lynch, M. (1990) The Externalized Retina: Selection and Mathematization in the Visual Documentation of Objects in the Life Sciences. In Lynch, M. and Woolgar, S. (Eds): *Representation in Scientific Practice*. Cambridge MA: MIT Press, pp. 153-186.

Mentis, H., O'Hara, K., Sellen, A. and Trivedi, R. (2012) Interaction Proxemics and Image Use in Neurosurgery. *CHI 2012. Proceedings of Conference on Human Factors in Computing Systems, 5-10 May 2012, Austin, Texas*, pp. 927-936.

O'Hara, K., Harper, R., Mentis, H. Sellen, A. and Taylor, A. (2013). On the naturalness of touchless: putting the "Interaction" back into NUI. In *ACM Transactions on Computer-Human Interaction*, vol. 20, no. 1, March 2013, pp. 1-25.

Ruppert, G., Amorim, P., Moares, T. and Silva, J. (2012) Touchless Gesture User Interface for 3D Visualization using Kinect Platform and Open-Source Frameworks. *Innovative Developments in Virtual and Physical Prototyping. Proceedings of the 5th International Conference on Advanced Research in Virtual and Rapid Prototyping, 28 September – 1 October, Leiria, Portugal,* pp. 215–219.

Stern, H. I., Wachs, J. P. and Edan, Y. (2008) Optimal Consensus Intuitive Hand Gesture Vocabulary Design. *ICSC 2008. Proceedings of International Conference on Semantic Computing, 4-7 August, Santa Clara, CA,* pp. 96-103.

Strickland M., Tremaine J., Brigley G. and Law C. (2013) Using a depth-sensing infrared camera system to access and manipulate medical imaging from within the sterile operating field. *Canadian Journal of Surgery*, vol. 56, no. 3, June 2013, E1-6.

Svenssen, M., Heath, C. and Luff, P. (2007) Instrumental action: the timely exchange of implements during surgical operations. *ECSCW '07. Proceedings of ECSCW '07, 24-28 September 2007, Limerick, Ireland*, pp. 41-60.

Tan, J., Unal, J., Tucker, T. and Link, K. (2011) Kinect Sensor for Touch-Free Use in Virtual Medicine. http://www.youtube.com/watch?v=id7OZAbFaVI&feature=player_embedded. Uploaded 11 February 2011.

Wachs, J., Kolsch, M., Stern, H. and Edan, Y (2011) Vision-Based Hand Gesture Applications. *Communications of the ACM,* vol. 54, no.2*,* February 2011, pp. 60-71.

Wachs, J., Stern, H., Edan, Y., Gillam, M., Feied, C., Smith, M., and Handler, J. (2006) A Real-Time Hand Gesture Interface for Medical Visualization Applications. *Applications of Soft Computing: Advances in Intelligent and Soft Computing vol. 36,* pp.153-162.

Wilson, R. N. (1954) Team Work in the Operating Room. *Human Organization*, vol. 12, no. 4, winter 1954, pp. 9-14.