Touchless computer interfaces in hospitals: A review

Seán Cronin
GLANTA Ltd, Ireland

Gavin Doherty iD
Trinity College Dublin, Ireland

Abstract

The widespread use of technology in hospitals and the difficulty of sterilising computer controls has increased opportunities for the spread of pathogens. This leads to an interest in touchless user interfaces for computer systems. We present a review of touchless interaction with computer equipment in the hospital environment, based on a systematic search of the literature. Sterility provides an implied theme and motivation for the field as a whole, but other advantages, such as hands-busy settings, are also proposed. Overcoming hardware restrictions has been a major theme, but in recent research, technical difficulties have receded. Image navigation is the most frequently considered task and the operating room the most frequently considered environment. Gestures have been implemented for input, system and content control. Most of the studies found have small sample sizes and focus on feasibility, acceptability or gesture-recognition accuracy. We conclude this article with an agenda for future work.

Keywords

gestures, hands-free, infection control, sterile, touchless, user–computer interface

Background

Healthcare associated infections (HCAIs) are a major problem. In the United States, HCAIs cause 99,000 attributable deaths and cost US$6,500,000,000 every year.[1] In Europe, they result in 16,000,000 extra days spent in hospital, 37,000 attributable deaths and €7,000,000,000 in cost every year.[1]

Modern technology can contribute to patient care by allowing healthcare staff rich and immediate access to patient information and imaging. However, computers and their peripherals are difficult to sterilise effectively, and keyboards are natural breeding grounds for various pathogens.[2]

Corresponding author:
Gavin Doherty, Trinity College Dublin, Dublin 2, Ireland.
Email: gavin.doherty@tcd.ie

In order to reduce the spread of HCAIs, hospitals must implement multimodal strategies, including such measures as provision of anti-bacterial gels, ongoing hand hygiene training and education, and proper equipment sterilisation and cleaning. However, when a healthcare worker (HCW) washes their hands, it is unlikely that 100 per cent of pathogens are eliminated.[3] As such, a simple and effective solution for preventing contamination of surfaces by people's hands, and of people's hands by surfaces, is to remove the need to touch those surfaces at all during use. Given the widespread use of technology in hospitals, interacting with computer-based systems using touchless interaction may be helpful for reducing opportunities for contamination.

Touchless control of computers has become more common in recent years. The most important factor facilitating this progress has been the introduction of more reliable and affordable consumer grade hardware, particularly time-of-flight (ToF) cameras such as the Microsoft Kinect. It is thus appropriate at this point to examine the literature to assess the current state of the art and to identify where further efforts should be directed.

Methods

Search strategy

The aim of the search was to find papers concerning user interactions with medical devices or other information technology in a medical context that do not involve touching them with the hands.

Selection process

The literature review was performed across three databases:

ACM: To cover the field of computer science research.

PubMed: To cover the field of biomedical research.

Web of Science: To cover more general scientific research.

Each database was searched using the same methodology, covering a period from January 2000 to January 2016. Three distinct groups of search terms were combined. In Level 1, search terms were 'gesture recognition', 'voice recognition', 'speech recognition', 'gaze tracking', *touchless*, *contactless*, *hands-free* and *touch-free* and were used in order to restrict the corpus to papers that pertained to touchless control of any system/interface. In Level 2, search terms were used to identify the environment (to filter out those papers concerned with touchless interaction in other environments), *hospital*, *medical*, *hygiene* and *sterile*. In Level 3, search terms were used in order to restrict the corpus to those papers in the medical devices/technology field, *interaction*, *interface*, *device* and *control*. For each search, all the returned paper titles were read, and all papers with titles considered relevant were extracted for further investigation. Titles were deemed relevant based on the presence of key terms in the title, and reference to appropriate contexts, such as hospital or operating room (OR).

In the next step, the abstracts of all papers returned were read and analysed for relevance. Papers that were deemed relevant were accepted for acquisition and full review. The criteria used to determine relevance included whether a paper had key search terms included in the abstract and abstracts that made reference to interaction with a medical device or computer. The final step was to read the full paper contents to determine whether or not touchless interaction with computer interfaces in hospitals was a core theme of the paper. In keeping with recent practice, a further search was

Table 1. High-level search results by database (simplified).

Database	Total results	Rejected at title	Rejected at year	Rejected at abstract	Rejected for no access	Rejected at paper	Final
ACM Digital Library	1229	1208	2	5	0	5	9
PubMed	811	700	23	50	6	5	27
Web of Science	376	371	2	0	0	1	2 (both duplicates)
Overall	2416	2279	27	55	6	11	36 (38 with duplicates)

performed using Google Scholar to identify any papers of significance that may have not have been returned when searching the other databases.

Data synthesis and analysis

A simplified overview of the breakdown of search results by database is given in Table 1. For each paper, the motivations for using touchless technology were noted, along with the nature of the technology used, and the user group, tasks and outcomes of any evaluation presented. User group is defined as the group of individuals who made use of the system and whose performance and feedback was collected and presented by the authors. Outcomes include quantitative and qualitative outcomes such as gesture-recognition rates and subjective user-reports on ease of use. Many of the papers discuss difficulties encountered, and so, these are qualitatively analysed for common themes, as well as any non-functional requirements (such as usability or reliability) discussed.

Results

The main search resulted in the identification of 36 unique articles. An additional 5 papers were identified through Google Scholar (noted separately in Figure 1), resulting in a final corpus of 41 unique articles. Before presenting the results, we should note that the corpus is very varied, ranging from brief papers on technical feasibility through to human–computer interaction (HCI) papers with rich discussion of user interaction. A breakdown of geographical location of the research is given in Table 2. There were no studies describing clinical outcomes, and so, no meta-analysis is presented.

Motivations for using touchless control

Sterility. Sterility is the most commonly cited motivation for touchless control (27 papers). Complications and infections caused by non-sterile interactions can be very costly,[4] with both financial and human costs. It is noted by Wachs et al.[5] that the current most prevalent means of HCI in hospitals remains the mouse and keyboard. Keyboards and mice are a potential source of contamination (up to 95% of keyboards have been shown to be contaminated[6]). However, computers and their peripherals are difficult to sterilise.[7]

While the increased use of technology in highly sterile settings such as the OR, and particularly the use of imaging, is noted, the need for access to non-sterile technology is a problem:

> Unfortunately, the necessary divide between the sterile operative area and the non-sterile surrounding room means that, despite physical proximity to powerful information tools, those scrubbed in the OR are unable to take advantage of those resources via traditional human–computer interfaces.[8]
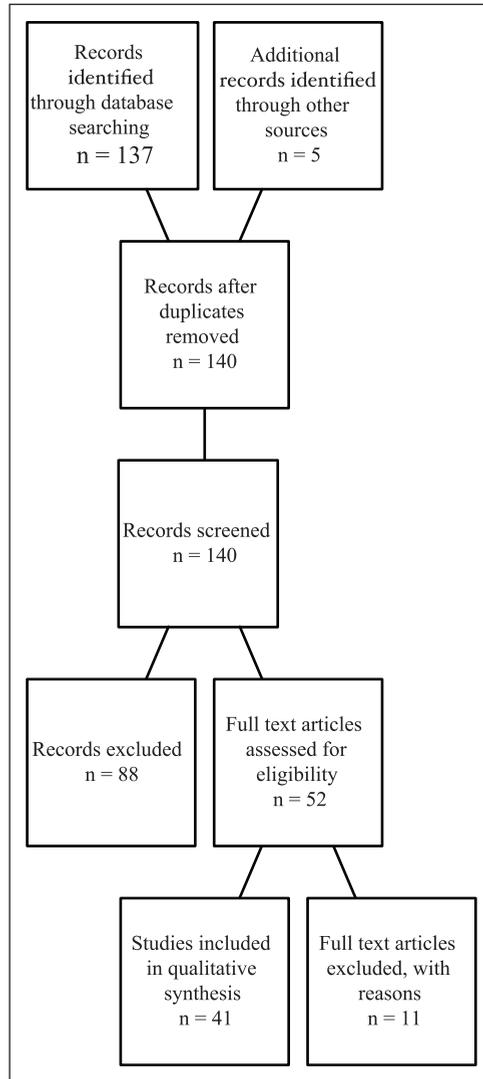
Figure 1. Study selection flow diagram.

Surgeons must remember details from prior review of each case, either asking assistants to control devices for them or using ad hoc barriers[4,8] – one strategy used by some surgeons is to pull their surgical gown over their hands, manipulating the mouse through the gown.[9] As a result, surgeons are less likely to use computer resources if they have to step out of the sterile surgical field due to the time and effort of scrubbing back.[8] Dela Cruz et al.[10] stated that breaks and interruptions to workflow leads to increased chance of medical error and poorer patient outcomes, and so, they should be avoided where possible. Cleaning to prevent bacterial contamination during surgery after checking a computer can take up to 20 min, sometimes adding a full hour to surgery.[11] However, surgeons may need direct control to mentally 'get to grips' with what is going on in a procedure.[12]

Table 2. Distribution of papers by location of origin.

United States	14	Argentina	1
Germany	5	Australia	1
United Kingdom	4	Brazil	1
Italy	3	Colombia	1
Denmark	3	Czech Republic	1
Canada	3	Finland	1
Israel	1	Japan	1
Switzerland	1		

Three-dimensional applications. In total, 21 of the reviewed papers referred to three-dimensional (3D) applications, such as manipulating 3D imagery and data sets. A major advantage of 3D hands-free interaction is the ability to navigate 3D data in 3D space[13] when compared to a conventional mouse, which only operates in two dimensions. In interpreting the thousands of images that modern scanners can produce, conventional analysis of two-dimensional (2D) images is potentially cumbersome.[11] This 3D interaction potentially enables more efficient navigation of 3D data.[12]

Hands busy. Five of the reviewed papers referred to hands-busy interaction with the system, for example, a surgeon holding surgical tools needing to manipulate a medical image.[14] Johnson et al.[9] observe that at times, radiologists have their hands full holding and manipulating wires and catheters.

Removing barriers. Touchless interaction enables not only a potential speed-up for specific tasks, for example, image manipulation in an OR, but also enables interaction modes not previously available, as well as supporting hospital sterility. O'Hara et al.[14] observe that

> 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.

Currently, significant barriers exist to the usage of technology in sterile environments: 'In current practice, therefore, the use of modern technology in the OR is at best awkward and fails to realise its full potential for contributing to the best possible surgical outcomes'.[8]

Context of use

OR. As it stands the single most frequently considered use case for such touchless control has been in the OR (mentioned in 32 papers) during surgical procedures, notably for the purpose of providing control of medical image systems such as Picture Archiving and Communications (PAC) systems based on the DICOM standard. In total, 10 papers discuss PAC /DICOM in the OR context. Kim et al.[15] found that it is feasible to manipulate surgical tools and execute simple surgical tasks, such as controlling OR lights, imaging data or positioning operating booms, using current commercially available contactless tracking technology, such as the Microsoft Kinect.

Interventional radiology. Interventional radiology was discussed as a context of use in 11 papers, making it the second most considered use case for touchless control. Image interaction in

interventional radiology takes place within a surgical context rather than a purely diagnostic context, which makes sterility a key requirement.[9]

Tasks. Table 3 lists the tasks and user groups for evaluations where this has been specified. As can be seen, the most common system application for touchless, gesture-driven interfaces has been various forms of image navigation, closely aligned with the OR and interventional radiology contexts listed above. Medical image navigation is the goal in seventeen papers,[4–9,11,14,16–24] with the more specific subset of magnetic resonance imaging (MRI) image navigation being the goal in six papers.[5,7,8,9,14,16] Task context was also significant in identifying tasks for testing the systems implemented, such as measuring a lesion on MRI images.[17]

Context-sensitive systems. A context-sensitive system is a system that is affected by or reacts to the context it appears in. While the term 'context-sensitive systems' was not found in the corpus, the concept of context was found in five of the papers. Several papers investigated the relationship between context and system functionality. Contextual information, such as focus of attention, can be employed to improve recognition performance.[7] Another method for improving recognition accuracy is switching between vocabularies of gestures based on context as suggested by Wachs et al.[36] This approach could improve performance in gesture-recognition systems, as separate gesture-recognition algorithms are used for smaller gesture subsets. For touchless interaction systems, there are a number of contextual cues that should be considered, both on a situational and an individual level. On a situational level, for example, in the OR, there are specific activities that a surgeon can be expected to be engaged in, such as focusing on the patient and the key image, as well as staying close to the patient.[5]

Technologies

Available technologies have improved markedly over time. At the time of publication of the study by Gallo et al.,[16] there existed no reliable and mature technology for effective gesture control. Over time, researchers have investigated using eye gaze technology (EGT),[37] capacitive floor sensors[22] and inertial orientation sensors;[21,22] colour cameras such as the Canon VC-C4,[5] the Loop Pointer[17] and MESA SR-31000 ToF cameras;[13] Siemens integrated OR system;[38] wireless hands-free surgical pointer;[29] to the Apple iPad;[17] leap motion controllers;[24–26] and the Microsoft Kinect ToF camera G1.[4,11,15–17,19,27,28] Data frequency has ranged from devices with 15-fps output, up to the leap motion controller with greater than 100-fps output. Regarding camera configurations, Chao et al.[17] suggest that moving to a stereo camera set-up might improve accuracy for touchless interaction. Along with advances in the hardware, there have also been complementary developments in available software for implementation of touchless systems. Many of the papers that utilised the Microsoft Kinect used additional specialised software, such as Primesense drivers,[17] OpenNI software[7,17] and skeletal tracking NITE.[17] Across the papers, there was a mix of purely gestural commands, and combined gesture and voice commands (a form of multimodal interaction).

The primary modes of human communication are speech, hand and body gestures; facial expressions; and eye gaze,[5] and much of the HCI literature on these forms of interaction draws motivation from this 'natural' nature rather than the touchless properties of these interaction modalities: 'Gestures are useful for computer interaction since they are the most primary and expressive form of human communication'.[36]

Only one paper by Chao et al. directly compared the efficacy of multiple devices. In their paper, the authors compared the Microsoft Kinect, Hillcrest Labs Loop Pointer, and the Apple iPad. Theirs was the only paper to use either the Hillcrest Labs Loop Pointer or the Apple iPad. Their

Table 3. Tasks, users and outcomes.

Task	User group	Sample size	Main reported outcome
Medical image interaction[7]	Surgeons	10 users	Mean recognition accuracy of 97%. The system was deemed to be easy to learn and to remember and moderately easy to perform.
Medical image navigation[4]	Radiologists	29 radiologists	69% believed the system could be useful in an interventional radiology practice.
Medical image navigation[8]	Surgeons	6 procedures	Users reported that they accessed more medical images than normal using the system.
Moving, zooming, windowing medical images[20]	Medical professionals	10 users	Mouse control was found to be faster than Kinect control, both with and without previous experience of the medical image viewing software.
Measuring a lesion on MRI images[17]	Radiologists	29 users	The iPad was found to be the most usable, and the Kinect the least usable, with tasks taking nearly 4 times as long with the Kinect.
MRI navigation[5]	Surgeons	1 operating procedure	The system was found to be easy to use, responsive and fast to train with. Gesture-recognition accuracy of 96%.
Locating an aortic stent and navigating to bifurcation[21]	Biomedical engineers	10 users	The inertial sensors were shown not to inhibit user's movement. System was deemed to be responsive and precise, albeit slower than a mouse and keyboard.
Medical data browsing[6]	Hospital employees	20 users	Calibration took less than 10 s, and task completion rates were 95% or higher.
3D medical data navigation[11]	Unspecified	18 users	Users completed tasks faster when using interfaces with appropriate numbers of degrees of freedom.
Presentation control[25]	A professor	99 gestures	Approximately 25% of gestures were not fully recognised because they were not performed in the working space or the gesture was not performed properly.
Classifier performance[13]	Unspecified	15 users	The system operated at 13.8 fps without classification and at 10.69 fps with classification.
Varied[13]	Primarily computer science students	7 users	On a Likert-type scale of 1–5 (5 being the best) the system scored: response time of the system: 4.14; adaptation to the system: 3.57; comfort of gesture set: 4.0; intuitiveness of the gesture set: 4.0.
Arbitrary parameter adjustment[26]	Unspecified	12 users	Users significantly improved their performance with practice over a small number of repeated uses of the system.
A 10-step predefined scenario[22]	Computer science researchers	5 users	Gesture-recognition accuracy was 74% for foot gestures and 79% for hand gestures. User feedback was generally positive, especially noting simpleness of use.
Controlling an OR table[27]	Unspecified	10 users	The system was graded as an above average interface.

(Continued)

Table 3. (Continued)

Task	User group	Sample size	Main reported outcome
Controlling an interactive hospital room[28]	Unspecified	18 users	97% voice and 93% gesture accuracy. Users prefer to perform whole tasks using only one type of interaction.
Peg transfer and pointing tasks[15]	Experienced surgeons	5 users	The da Vinci system had the lowest latency, the lowest tremor radius, and the fastest time to task completion when compared to the 3Gear and Mantis systems.
Laparoscopic cholecystectomy procedures[29]	Instructing surgeons (6), operating surgeons (29) and camera assistants (48)	83 users	No negative effects on surgery completion time when using wireless hands-free surgical pointer (WHaSP). The WHaSP was found to be comfortable, easy to use and easy to control. Furthermore, the WHaSP improved communication effectiveness in the OR.
Voice input to EHR[30]	Nurses(7) and others (3)	10 users	Voice input took an average of 304.5 s per record versus 1459 s for keyboard input.
Creating and editing medical reports using voice control[31]	Pathology assistants (48), residents (12) and attending physicians (20)	80 users	Use of voice recognition has led to a marked reduction in report turnaround time (554–102 min).
Creating and editing radiology reports using voice control[32]	Radiologists	7 users	Reports generated using voice recognition were approximately 24% shorter in length and took 50% longer to dictate than those transcribed conventionally. The reports generated using voice dictation also contained more errors (5.1 errors/report vs 0.12 errors/report).
Creating and editing radiology reports using voice control[33]	Radiologists	2 users	Productivity for one radiologist was calculated at 8.6 MRI reports per hour using voice recognition and 13.3 MRI reports per hour using a transcriptionist.
Voice recognition of sentences[34]	Users with no medical background	8 users	The system performed at a higher rate of classification in command mode than free speech mode (81.6% vs 77.1%). Training the system without background noise improves recognition rate (85.5% vs 77.8%).
Updating anaesthesia record using speech recognition[35]	Doctors and nurses	12 users	Users found it slightly more difficult to update the anaesthesia record by voice, although voice input required significantly less time than traditional input and almost two times as many medication registrations were made with voice input as compared to without voice input.

MRI: magnetic resonance imaging; OR: operating room; EHR: electronic health record.

results showed the Apple iPad to have had the greatest number of participants with prior experience of the device and the Hillcrest Labs Loop Pointer to have had the least prior experience.[17] The authors' results also showed the Apple iPad to have the greatest usability score, as well as the

lowest completion time for a sequence of measurement tasks (mean usability score: 13.5 out of 15; mean completion time: 41.1 s), followed by the Hillcrest Labs Loop Pointer (mean usability score: 12.9 out of 15; mean completion time: 51.5 s), with the Microsoft Kinect scoring the lowest (mean usability score: 9.9 out of 15; mean completion time: 156.7 s).[17]

Gesture recognition. Gesture interaction is claimed to be intuitive because users are familiar with communicating with other people by means of gestures.[27] In total, 20 of the papers discuss the design and implementation of gesture-recognition systems. Wachs et al.[36] discuss a number of forms of analysis for hand-gesture recognition:

Motion: effective and computationally efficient;

Depth: deemed to be potentially useful;

Colour: heads and hands can be found with reasonable accuracy using only their colour;

Shape: available if the object is clearly segmented from the background, can reveal object orientation;

Appearance: more robust but higher computational cost;

Multi-cue: a combination of the previous approaches

Conventional interfaces via gesture or native gestural interfaces. Several papers have tried to apply gestures directly to conventional means of interacting with a computer,[13] for example, Rosa and Elizondo[24] implemented a virtual touch-pad in mid-air. However, adhering to existing control design paradigms such as mouse and keyboard control design has been found to be a drawback.[4] Others have created complete gesture sets with purely gestural control in mind.[14,16,17] Designing the system with gestures in mind from the very start would also help to reduce the issues associated with adhering to existing mouse and keyboard control design according to Tan et al.[4] Other systems have used gesture modalities to provide more direct control within medical procedures, such as FAce MOUSe where a surgeon can control the direction of a laparoscope simply by making the appropriate facial gesture.[5]

Gesture set. Choosing the appropriate gesture set is key in system design, and hardware characteristics of input device and the application domain must be considered.[16] There have been a large number of gestures described in the literature, some more common than others. Table 4 classifies the gestures encountered into system control, content control and input, with content control being the largest category. Despite the extent of this list of gestures, O'Hara et al.[12] state that limiting the number of gestures benefits ease of use, as well as learnability. Furthermore, limiting the gesture set can enhance reliability and avoid 'gesture bleed' (where gestures containing similar movements are mistaken for each other by the system).

O'Hara et al.[14] talk about expressive richness as 'how to map an increasingly larger set of functional possibilities coherently onto a reliably distinctive gesture vocabulary', as well as how to approach transitioning between gestures.

Rossol et al.[26] designed an interface with the purpose of using gestures designed to be equally efficient to use with bare hands or hand-held tools. In order to minimise cognitive load, their system used a vocabulary of three gestures for finger – or tool – tips, and one gesture for hands. Furthermore, they implemented a means of performing highly precise adjustments by means of tapping gestures on one hand while using the closer hand for parameter adjustment.[26] In terms of fine detail

Table 4. Complete list of gestures found in the literature.

System control	Content control	Input
Click[13,24–27]	Translation[11,13,16,27]	Measuring[17,24]
Unlock[8,14]	Rotation[13,14,16,24]	Cursor[8,13,16,24,25]
Reset[13]	Scrolling[8,17]	Entering a value [25,26]
Windowing[16,17]	Zooming[14,16,17,22,24,26]	
Set idle[16]	Panning[14,17]	
Lock[8,14,27]	ROI extraction[16]	
	ROI erasing[16]	
	Animating[16]	
	Navigation[8,22,24,25,27]	
	Complex 3D[11,24,25]	

ROI: region of interest; 3D: three-dimensional.

performance, depth cameras can now track fine motions of hands and fingers.[15] Tan et al.[4] flagged issues with hand tracking and inconsistent responsiveness as issues, along with their stylistic choice of requiring two hands for gestures. Wachs et al.[36] discussed the issue of lexicon size and multi-hand systems where the challenge is to detect and recognise as many hands as possible.

Wachs et al.[36] flagged system intuitiveness as an issue, that is, there is no consensus among users as to what command a gesture is associated with. Dealing with differences in gestures between individuals is a considerable challenge.[16] Regarding system reconfigurability, Wachs et al.[36] stated that there are many different types of users; location, anthropometric characteristics and types and numbers of gestures vary.

Adjustment of continuous parameters may also be a possibility in gestural input – O'Hara et al.[14] use a combination of voice commands for discrete commands and mode changes and gestures for the control of continuous image parameters.

Soutschek et al.[13] found that a majority of processing time was spent on acquiring images and preprocessing, for example, resizing. However, as computers have become more powerful, such processing is now easily accomplished allowing for more sophisticated real-time solutions. System performance and user familiarity impact directly on the user experience. Soutschek et al.[13] assert that users dislike systems when there is a perceptible delay during use.

Unintended gestures and clutching. Not all gestures are intended as commands, for example, pointing out a feature of interest to a colleague,[25] and the misinterpretation of such gestures can adversely affect the user experience. The inclusion of a lock and unlock gesture (an example of a clutching mechanism) is essential according to Mauser and Burgert;[25] O'Hara et al.[12] used a voice command to lock and unlock the system. The aim is to ignore inadvertent commands: the system should be inactive until hailed by a distinctive action and should be locked using another distinctive action.[8] Deliberate gestures and unintentional movements need to be distinguished from each other. Unintentional movement usually occurs when the user interacts with other people or is resting their hands.[36] Rossol et al.[26] found unintentional finger – or tool – tip movement being interpreted as input to be a drawback with their system. In order to combat overlapping gestures, they suppressed any recognised gestures that overlapped the previous gesture's time window.[26] Tan et al.[4] state that fine movements were the most difficult. How to define starting and ending of a gesture[16,25] is also a difficult issue. One advantage of depth cameras is the ability to take account of motion in the Z plane to reduce unwanted gestures when returning to idle.[16]

Physical issues. At a practical level, gesture recognition works best at particular distances.[18] When designing gestures, O'Hara et al.[14] looked for ways to facilitate the work of clinicians, while maintaining sterile practices, by restricting movements to the spatial area in front of the torso. Tracking movements relative to the body may be the most appropriate,[9] specifically information from the operator's upper limbs and torso to implement the functionality of a mouse-like device.[8] Regarding the 3D gesture zone, 'this zone extends roughly from the waist inferiorly to the shoulders superiorly and from the chest to the limit of the outstretched arms anteriorly and to about 20 cm outside each shoulder laterally'.[8] Using environmental cues for intent, Jacob et al.[7] allowed users perform gestures anywhere in the field of view. However, depth segmentation has required an upper and lower threshold,[13] meaning that to use a system the user cannot be too close or too far away.

Comfort and fatigue. Regarding the issue of interaction space, Wachs et al.[36] ask is it right to assume that the user and device are static and the user will be within a standard interaction envelope? Tan et al.[4] also say that ample space is required to operate their touchless system. Sufficient operational space is one of the factors affecting user comfort. This relates to the question of fatigue, and the need to avoid intense muscle tension over time. Consideration of static ('the effort required to maintain a posture for a fixed amount of time') and dynamic ('the effort required to move a hand through a trajectory') stress is key in promoting user comfort.[36]

Training and calibration. Systems should be easy to integrate into existing ORs with minimal distraction, training or human resources according to Strickland et al.[8] Rosa and Elizondo[24] state that with a little training of the user, use of their gesture interface is easier and faster than changing sterile gloves or having an assistant outside the sterile environment. In O'Hara et al.,[14] the surgical team became familiar with the system through ongoing use of the system, 'learning on the job', rather than a dedicated training system, supported by prompts from the lead surgeon who acted as champion for the system. Chao et al.[17] determined that prior use of a device had a significant impact on task completion time, and found that gamers were faster on all devices.[17] However, Ebert et al.[20] observed no significant difference between gamers and non-gamers.

A gesture classifier typically needs to be trained by providing a large set of sample data. Regarding this training, Jacob et al.[7] state that it is imperative to use a greater number of users (high-variance training data). System calibration results in a time cost. For example, the total set up time (including calibration) was ca. 20 min for the system of Wachs et al.[5] Calibration is also required in some papers that use a Kinect.[16] However, for some systems, no calibration was required.[20]

Voice control. Perrakis et al.[38] believe that voice control has an important role to play in minimally invasive surgery, allowing the surgeon to take control of the entire OR without breaking sterility or interrupting the surgery; this potentially allows for single surgeon surgery, resulting in reduced costs. As mentioned in 29 papers, and discussed on a practical level in 20 papers, voice control has been found to be slower but more accurate than gesture control, and both were slower than traditional methods.[28] Two major issues for voice control are people's accents,[7] and ambient noise, with the noise levels of an OR making voice control extremely difficult.[5]

When designing a grammar for a speech recognition-based interface, care must be taken to select words that are easily recognisable for the various users of the system and sufficiently distinct from each other phonetically to avoid possible mis-recognitions.[34] Strickland et al.[8] suggest that the implemented gesture vocabulary does not need to allow full functionality of the PAC system, but rather a subset of the most common functions.

Voice control for text input. In total, 13 papers discussed voice control as a means of inputting text. Given a sufficiently high speech recognition engine confidence score, use of a keyword to activate the system and of another keyword to switch between command based and free text mode may allow for a completely hands-free approach.[35]

Voice recognition has been described as having accuracy as high as 99 per cent; however, some studies have shown slightly lower accuracy than human transcription.[31] According to Kang et al.,[31] the largest benefit to using voice control is a decrease in turnaround time, which results in higher administration, and clinician satisfaction, whereas the biggest disadvantage is an increased editing burden on clinicians. They report natural dictation at speeds of 160 words per minute.[31] Marukami et al.[30] investigated voice recognition input to an electronic nursing record system (ENRS) and compared the time taken to input records to an electronic health record (EHR) using voice recognition as compared to by means of keyboard input, finding that users were able to input records roughly 5 times faster using voice recognition (304.5 vs 1459 s). In contrast, Pezzullo et al.[32] found that reports generated by means of the voice recognition system were 24 per cent shorter in length and took 50 per cent longer to dictate than those transcribed conventionally. In terms of cost per report, Pezzullo et al.[32] note that use of a voice recognition system may result in a 100 per cent increase in dictation costs, caused by the significant difference in cost per hour of radiologists compared to transcriptionists.[33]

Voice control for discrete commands. Voice control is generally deemed good for discrete commands, though it is not appropriate for continuous parameter adjustment, for which gestures are better suited.[14] Dictation systems require discrete commands, with Argoty and Figueroa[28] proposing two-word commands as more meaningful to the user than single-word commands. Nagy et al.[39] say that increased length of commands plays a significant role in improving recognition hit rates. Alapetite[34] found that voice recognition displayed higher accuracy when issuing discrete commands (81.6%) as compared to free speech mode (77.1%). Use of a voice recognition interface resulted in a significantly higher quality of anaesthesia record as compared to the traditional interface (99% of medications recorded vs 56%), as well as a reduced error rate.[35] In contrast, Pezzullo et al.[32] found that their voice recognition system resulted in more errors per report than conventional transcription (5.1 errors/report compared to 0.12 errors/report) and go on to suggest that 'radiologists are not good transcriptionists'.

Training for voice control. Hoyt and Yoshihashi[40] state that the success or failure of voice recognition technology in a hospital is dependent on personal experience, training, technological or logistical reasons. To this end, voice recognition vendors may provide 'train the trainer' sessions to users with high levels of aptitude ('superusers').[40] Rossol et al.[26] found that users can significantly improve their performance with practice over a small number of repeated uses. In Kang et al.,[31] new users took a 1-h training session, and setting up a new voice profile took approximately 10 min. Alapetite[35] found that setting up and training a new voice profile took roughly 30 min.

Fatigue in voice control. Marukami et al.[30] considered the issue of user fatigue when implementing voice recognition as compared to keyboard input and gathered user feedback regarding both input tools by means of a questionnaire. Their results indicated that users found that voice input caused less fatigue and was easier compared with keyboard operation, despite being inexperienced with voice input.[30]

Time of day also seems to impact on performance. Luetmer et al.[41] identified an increased error rate in laterality in radiology reports during the evening and overnight shifts (0.154% during the evenings, 0.124% overnight, compared to 0.0372% during the day). They also found that reports

generated using voice control had similar major laterality error rates to those generated without using voice control.[41]

Performance of voice control. Alapetite[35] found that the voice recognition interface led to shorter action queues than the traditional interface, but users found that it required slightly more concentration and was slightly more difficult to update the anaesthesia record using voice recognition input. Pezzullo et al.[32] declare that with diminished speech recognition accuracy comes an increase in time spent editing reports, resulting in a decrease of user satisfaction.

Alapetite[35] observed a speech command must be said in one go, distinctively and without any dysfluency. Perrakis et al.[38] found that speaker's accent did not have an impact on system accuracy, with functional errors in using the system being approximately the same for non-native and native German speakers. Alapetite[34] found that when a user profile was trained with background noise, there was a slight increase in free speech recognition performance (78.2% vs 75.6%), in contrast to the results for command mode, and stated that 'background noises have a strong impact on recognition rates'.

Other technologies. Nine papers discussed the use of EGT as an interaction modality. Modern eye trackers have the advantage of being very easy to install and to use.[37]

Two papers discussed inertial-type sensors attached to the users' bodies to capture gesture input. Jalaliniya et al.[22] stated that advantages of such a system were that the system did not require a direct line of sight for the user and that it would allow only a designated person (the wearer of the sensor) to interact with the system. Bigdelou et al.[21] discussed the hardware issues of inertial orientation sensor-based systems, highlighting issues such as noise and drift.

Non-functional requirements

The most commonly referenced non-functional requirements in the corpus are usability and reliability of gesture and voice control. System design should consider real-time interaction, sterility, fatigue, intuitiveness/naturalness, robustness, ease of learning, unencumbered use and the scope for unintentional commands.[5] Another requirement is low cognitive load, with short, simple and natural gestures.[36] Natural interaction is taken to include the use of voice and gesture commands.[36] Furthermore, it is suggested that systems should focus on being stable and providing basic access rather than trying to be more powerful and versatile.[8] The system needs to support both coarse and fine-grained system control through careful design of the gesture vocabulary,[9] and how gestures are mapped to control elements in the interface.[9] Reliability is identified as a key non-functional requirement,[8] which is impacted by the issue of unintentional gestures, and gesture control interfaces may need to sense the human body position, configuration and movement in order to achieve this.[16]

Evaluations

A variety of outcomes are studied in the literature (Table 3) with accuracy of gesture recognition being the most frequently reported outcome (seven papers). There are a number of factors that should be considered when evaluating a system.[36] Validation of sensitivity and recall of gestures, precision and positive predictive value, f-measure, likelihood ratio and recognition accuracy should all be rigorously evaluated using standard, public data sets.[36] Furthermore, usability criteria such as task completion time and subjective workload assessment and user independence should be evaluated. The usability of interfaces is described by different

standards, which focus on efficiency, effectiveness and user satisfaction.[27] Soutschek et al.[13] deem aspects such as classification rate, real-time applicability, usability, intuitiveness and training time as relevant aspects for evaluation.

Both quantitative and qualitative assessments of a system should also be carried out comparing gesture interaction to other technologies such as voice control and mouse and keyboard.[36] Subjective evaluation by experienced physicians is important and is likely more insightful than technical factor comparisons.[17] This obviously introduces an obstacle in terms of access to potentially large numbers of qualified personnel, as well as ethical and safety concerns for eventual real-world evaluations. However, a number of the papers did perform evaluations with representative end users, for example, Tan et al.[4] asked 29 radiologists to evaluate their system for efficacy as well as possible advantages and disadvantages.

Technical evaluations. In Jacob et al.,[7] development and validation involved three steps, lexicon generation, development of gesture-recognition software and validation of the technology. Papers have evaluated their systems on both technical and subjective (user experience) levels.[13] When performing a technical evaluation, data regarding technical accuracy need to be collected and analysed.[13] Technical evaluation might focus purely on accuracy of gesture recognition at the early stages of development, or on task performance at later stages, in which performance time can also be measured. The choice of realistic tasks is important. High accuracy is a requirement for medical implementation, with an accuracy of 95 per cent upwards suggested for use in a medical context.[6]

Feasibility. With the advances in technology using touchless technology, the focus is no longer on technical feasibility. Rather, it is important that we understand how systems and their design impact on the patterns of behaviour of hospital staff.[14] However, it is claimed that much existing work from a medical background lacks consideration of practical elements and implementations, remaining experimental and work originating from a technology background often suffers from over-simplification of medical complexity.[9]

Acceptability and satisfaction. A range of methods have been used for qualitative evaluations including contextual interviews, individual questionnaires and subjective satisfaction questionnaires. Wachs et al.[5] and Ebert et al.[20] made use of subjective questionnaires to determine user satisfaction with the system. Questionnaires may be used to gauge issues such as previous task experience, perceived ease of task performance, task completion time and overall task satisfaction.[5] Robustness is key to the acceptability of a system; Wachs et al.[36] specifically mention robustness for camera sensor and lens characteristics, scene and background details, lighting conditions and user differences.

Overall, there is a consensus that systems should be subject to both technical and qualitative evaluations using public data sets and demonstrate a very high level of gesture-recognition success in order to be appropriate for medical use.[6] There is also a recognised need to minimise unwanted side effects such as accidental gesture recognition.

Discussion

One issue across papers containing experimental results was a persistently small sample size during user testing of the systems, and no solution was tested in more than one hospital. In total, 10 of the studies were executed using non-medical personnel. Those papers that performed experimental work in hospital environments may have been constrained by access to hospital staff, whose time may be difficult to obtain. While these sample sizes are sufficient for early-stage prototyping and

investigating feasibility and acceptability, they are too small to determine whether the solutions proposed are appropriate for wide-scale deployment. No studies reported outcomes relating to contamination. Ultimately, outcome focussed evaluations showing reduced levels of contamination will be required to promote widespread adoption of these technologies.

Standardising evaluations

Effective and systematic protocols for evaluation of touchless technologies should be established for a range of medical environments. In particular, producing a set of standardised tasks for a particular context and user group would allow experiments comparing time-on-task measures for different designs and interaction modalities and also studies of outcomes based on contamination (e.g. comparing a touchless to a touch-based design). Critically, it would also allow for meta-analysis across different studies. Further standardisation of a gesture set (or set of voice commands for speech) to support a particular task would allow comparison of different implementations with regard to recognition speed and accuracy, again opening up the possibility of meta-analysis. Reliable comparisons of different hardware technologies would require use of the same tasks, gesture set and user group and thus would ideally be conducted as part of the same experiment.

Beyond the OR

While several OR and interventional radiology use cases have been explored, there has been no systematic consideration of use of touchless systems in other contexts around a hospital. This is an opportunity for future work as pathogens can be spread anywhere in a hospital. However, with this opportunity comes additional considerations and restrictions. For example, regarding voice control, audio feedback to the user is an option. However, within some parts of the hospital environment (such as intensive care), it may be important to minimise ambient noise levels.

Contextual cues and clutching

A number of contextual cues were also considered in the papers, including the cue of gaze. Several papers examined EGT for controlling systems. While gaze-based control is challenging, determining if a user's gaze is directed at the system is by comparison of a relatively simple task. This provides the potential for unencumbered gesture use of a system where gaze is used as a contextual cue to process or ignore gestures (clutching). Similarly, voice control can be accurate and well suited for discrete commands for clutching. Strickland et al.[8] identified system activation in a crowded operative space as being the primary source of issues during use of their system. Ultimately, multimodal systems combining speech and gesture, gaze and gesture or gaze and speech, although technically challenging, may have the best chance of success.

Implementation

Regarding implementing gesture recognition in systems, we see in the literature a gradual movement from implementing conventional interaction methods using gestures, to designing systems with gestures in mind from the ground up. Advancements in technology provide emerging opportunities in terms of potential accuracy of both depth and colour inputs. Developments such as the second generation of Microsoft Kinect, which was not used in any of the papers investigated, are a sufficiently significant improvement when compared to the first generation of such technologies that one can reasonably anticipate a noticeable improvement in possible applications.

Best practice

As touchless control moves towards being a viable option for use in the hospital environment, it is appropriate to consider best practice in the design and evaluation of these systems. While different in scope and focus, several of the findings presented in this article are supported by the recent independently conducted review of touchless interaction in the OR and interventional radiology by Mewes et al.[42] Major themes in their analysis echo the conclusions presented in this article regarding recent improvements in the feasibility of touchless control; the need for improved evaluations; the need to improve usability, including issues surrounding accuracy and unintended gestures; and the potential of multimodal interaction to address some of the practical difficulties in making these systems appropriate for deployment. With regard to best practice, these findings support careful consideration of usability in the design of touchless systems, using multimodal input to support clutching, using realistic tasks and conducting larger studies with representative HCWs.

Conclusion

The literature shows a gradual move in concern away from technical difficulties towards more fundamental issues such as the design of gesture languages and the potential impact of touchless systems on medical practices, particularly in the OR. It is clear that while progress has been made in the field, the literature does not support any instance of the technology being mature enough to gain widespread acceptance or adoption. The dramatic improvements in, and commoditisation of, the technologies involved have allowed significant advances in performance, and the technology is constantly improving; however, these capabilities have not yet been fully exploited. While the literature supports the technical feasibility of these types of system, and a useful variety of explorations of imaging-related tasks in the OR, the lack of larger studies and ecologically valid evaluations is a serious barrier to adoption. Providing benchmark tasks for particular contexts would allow for comparative studies, particularly in the context of future studies examining contamination as an outcome. While there is an understandable focus on the OR and interventional radiology as the most frequently examined use cases, the use of touchless systems in other areas of the hospital environment should also be explored.

ORCID iD

Gavin Doherty https://orcid.org/0000-0002-9617-7008

References

1. World Health Organization. *Report on the burden of endemic healthcare-associated infections worldwide*. Geneva: WHO, 2011.
2. Schultz M, Gill J, Zubairi S, et al. Bacterial contamination of computer keyboards in a teaching hospital. *Infect Control Hosp Epidemiol* 2003; 24(4): 302–303.

3. Cooper BS, Medley GF and Scott GM. Preliminary analysis of the transmission dynamics of nosocomial infections: stochastic and management effects. *J Hosp Infect* 1999: 43131–43147.

4. Tan JH, Chao C, Zawaideh M, et al. Informatics in radiology: developing a touchless user interface for intraoperative image control during interventional radiology procedures. *Radiographics* 2013; 33(2): E61–E70.

5. Wachs JP, Stern HI, Edan Y, et al. A gesture-based tool for sterile browsing of radiology images. *J Am Med Inform Assoc* 2008; 15(3): 321–323.

6. Feied C, Gillam M, Wachs J, et al. A real-time gesture interface for hands-free control of electronic medical records. *AMIA Annu Symp Proc* 2006; 2006: 920.

7. Jacob MG, Wachs JP, Packer R, et al. Hand-gesture-based sterile interface for the operating room using contextual cues for the navigation of radiological images. *J Am Med Inform Assoc* 2013; 20(e1): e183–186.

8. Strickland M, Tremaine J, Brigley G, et al. Using a depth-sensing infrared camera system to access and manipulate medical imaging from within the sterile operating field. *Can J Surg* 2013; 56(3): E1–E6.

9. Johnson R, OHara K, Sellen A, et al. Exploring the potential for touchless interaction in image-guided interventional radiology. In: *Proceedings of the SIGCHI conference on human factors in computing systems (CHI '11)*, Vancouver, BC, Canada, 7–12 May 2011, pp. 3323–3332. New York: IEEE.

10. Dela Cruz JE, Shabosky JC, Albrecht M, et al. Typed versus voice recognition for data entry in electronic health records: emergency physician time use and interruptions. *West J Emerg Med* 2014; 15(4): 541–547.

11. Gallo L. A study on the degrees of freedom in touchless interaction. In: *Proceedings of the SIGGRAPH Asia 2013 technical briefs*, Hong Kong, China, 19–22 November 2013. New York: ACM.

12. O'Hara K, Dastur N, Carrell T, et al. Touchless interaction in surgery. *Commun ACM* 2014; 57(1): 70–77.

13. Soutschek S, Penne J, Hornegger J, et al. 3-D gesture-based scene navigation in medical imaging applications using time-of-flight cameras. In: *Proceedings of the 2008 IEEE computer society conference on computer vision and pattern recognition workshops (CVPR Workshops)*, Anchorage, AK, 23–28 June 2008, pp. 2–7. New York: IEEE.

14. O'Hara K, Gonzalez G, Penney G, et al. Interactional order and constructed ways of seeing with touchless imaging systems in surgery. *Comp Support Comp W* 2014; 23(3): 299–337.

15. Kim Y, Leonard S, Shademan A, et al. Kinect technology for hand tracking control of surgical robots: technical and surgical skill comparison to current robotic masters. *Surg Endos* 2014; 28(6): 1993–2000.

16. Gallo L, Placitelli AP, Ciampi M, et al. Controller-free exploration of medical image data: experiencing the kinect. In: *Proceedings of the 24th international symposium on computer-based medical systems (CBMS)*, Bristol, 27–30 June 2011, pp. 1–6. New York: IEEE.

17. Chao C, Tan J, Castillo EM, et al. Comparative efficacy of new interfaces for intra-procedural imaging review: the Microsoft Kinect, Hillcrest Labs Loop Pointer, and the Apple iPad. *J Digit Imaging* 2014; 27: 463–469.

18. Mentis HM, OHara K, Sellen A, et al. Interaction proxemics and image use in neurosurgery. In: *Proceedings of the 2012 ACM annual conference on human factors in computing systems (CHI)*, Austin, TX, 5–10 May 2012, pp. 927–936.New York: ACM.

19. Ruppert GCS, Reis LO, Amorim PHJ, et al. Touchless gesture user interface for interactive image visualization in urological surgery. *World J Urol* 2012; 30(5): 687–691.

20. Ebert LC, Hatch G, Ampanozi G, et al. You can't touch this: touch-free navigation through radiological images. *Surg Innov* 2012; 19(3): 301–307.

21. Bigdelou A, Schwarz LA and Navab N. An adaptive solution for intra-operative gesture-based human-machine interaction. In: *Proceedings of the 2012 ACM international conference on intelligent user interfaces (IUI '12)*, Lisbon, 14–17 February 2012, pp. 75–84. New York: IEEE.

22. Jalaliniya S, Büthe L, Smith J, et al. Touch-less interaction with medical images using hand foot gestures. In: *Proceedings of the 2013 ACM conference on pervasive and ubiquitous computing (Ubicomp '13 Adjunct)*, Zurich, 8–12 September 2013, pp. 1265–1274. New York: IEEE.

23. Liu D, Zucherman M, Tulloss WB, et al. Six characteristics of effective structured reporting and the inevitable integration with speech recognition. *J Digit Imaging* 2006; 19(1): 98–104.

24. Rosa GM and Elizondo ML. Use of a gesture user interface as a touchless image navigation system in dental surgery: case series report. *Imaging Sci Dent* 2014; 44(2): 155–160.
25. Mauser S and Burgert O. Touch-free, gesture-based control of medical devices and software based on the leap motion controller. *Stud Health Technol Inform* 2014; 196: 265–270.
26. Rossol N, Cheng I, Shen R, et al. Touchfree medical interfaces. *Conf Proc IEEE Eng Med Biol Soc* 2014; 2014: 6597–6600.
27. Schröder S, Loftfield N, Langmann B, et al. Contactless operating table control based on 3D image processing. *Conf Proc IEEE Eng Med Biol Soc* 2014; 2014: 388–392.
28. Argoty JA and Figueroa P. Design and development of a prototype of an interactive hospital room with kinect. In: *Proceedings of the XV international conference on human computer interaction (Interacción)*, Puerto de la Cruz, 10–12 September 2014. New York: ACM.
29. Trejos AL, Siroen K, Ward CD, et al. Randomized control trial for evaluation of a hands-free pointer for surgical instruction during laparoscopic cholecystectomy. *Surg Endosc* 2015; 29(12): 3655–3658.
30. Marukami T, Tani S, Matsuda A, et al. A basic study on application of voice recognition input to an electronic nursing record system-evaluation of the function as an input interface. *J Med Syst* 2012; 36(3): 1053–1058.
31. Kang HP, Sirintrapun SJ, Nestler RJ, et al. Experience with voice recognition in surgical pathology at a large academic multi-institutional center. *Am J Clin Pathol* 2009; 133(1): 156–159.
32. Pezzullo JA, Tung GA, Rogg JM, et al. Voice recognition dictation: radiologist as transcriptionist. *J Digit Imaging* 2008; 21(4): 384–389.
33. Strahan RH and Schneider-Kolsky ME. Voice recognition versus transcriptionist: error rates and productivity in MRI reporting. *J Med Imaging Radiat Oncol* 2010; 54(5): 411–414.
34. Alapetite A. Impact of noise and other factors on speech recognition in anaesthesia. *Int J Med Inform* 2008; 77(1): 68–77.
35. Alapetite A. Speech recognition for the anaesthesia record during crisis scenarios. *Int J Med Inform* 2008; 77(7): 448–460.
36. Wachs JP, Klsch M, Stern H, et al. Vision-based hand-gesture applications. *Commun ACM* 2011; 54(2): 60–71.
37. Faro A, Giordano D, Spampinato C, et al. An interactive interface for remote administration of clinical tests based on eye tracking. In: *Proceedings of the 2010 symposium on eye-tracking research & applications (ETRA)*, Austin, TX, 22–24 March 2010, pp. 69–72. New York: ACM.
38. Perrakis A, Hohenberger W and Horbach T. Integrated operation systems and voice recognition in minimally invasive surgery: comparison of two systems. *Surg Endosc* 2013; 27(2): 575–579.
39. Nagy M, Hanzlicek P, Zvarova J, et al. Voice-controlled data entry in dental electronic health record. *Stud Health Technol Inform* 2008; 136: 529–534.
40. Hoyt R and Yoshihashi A. Lessons learned from implementation of voice recognition for documentation in the military electronic health record system. *Perspect Health Inf Manag* 2010; 2010: 1e.
41. Luetmer MT, Hunt CH, McDonald RJ, et al. Laterality errors in radiology reports generated with and without voice recognition software: frequency and clinical significance. *J Am Coll Radiol* 2013; 10(7): 538–543.
42. Mewes A, Hensen H, Wacker F, et al. Touchless interaction with software in interventional radiology and surgery: a systematic literature review. *Int J Comput Assist Radiol Surg* 2017; 12(2): 291–305.