Investigating Clutching Interactions for Touchless Medical Imaging Systems

Sean Cronin
Trinity College Dublin
Dublin, Ireland
secronin@tcd.ie

Euan Freeman
University of Glasgow
Glasgow, Scotland
euan.freeman@glasgow.ac.uk

Gavin Doherty
Trinity College Dublin
Dublin, Ireland
gavin.doherty@tcd.ie

ABSTRACT

Touchless input could transform clinical activity by allowing health professionals direct control over medical imaging systems in a sterile manner. Currently, users face the issues of being unable to directly manipulate imaging in aseptic environments, as well as needing to touch shared surfaces in other hospital areas. Unintended input is a key challenge for touchless interaction and could be especially disruptive in medical contexts. We evaluated four clutching techniques with 34 health professionals, measuring interaction performance and interviewing them to obtain insight into their views on clutching, and touchless control of medical imaging. As well as exploring the performance of the different clutching techniques, our analysis revealed an appetite for reliable touchless interfaces, a strong desire to reduce shared surface contact, and suggested potential improvements such as combined authentication and touchless control. Our findings can inform the development of novel touchless medical systems and identify challenges for future research.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**; *Haptic devices*; User studies.

KEYWORDS

Clutching, Gestures, Medical Imaging Systems, Midas Touch, PACS, Touchless Interaction

ACM Reference Format:
Sean Cronin, Euan Freeman, and Gavin Doherty. 2022. Investigating Clutching Interactions for Touchless Medical Imaging Systems. In *CHI Conference on Human Factors in Computing Systems (CHI '22), April 29-May 5, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3491102.3517512

1 INTRODUCTION

Interactive medical imaging systems have been of significant benefit to clinician work. PACS (Picture Archiving and Communication Systems) enable all-digital workflows, which dramatically reduce the time taken to access and review patient images, and make it easier for clinicians to view patient reports and scans during medical procedures. These have had an overall positive impact on clinical work and patient outcomes.

Sterility is a major concern when interacting with computing systems in medical settings, motivated by the need to manage healthcare associated infections. There is growing evidence of the role of contact with contaminated surfaces in the transmission of various nosocomial pathogens, such as *Clostridium difficile* and Norovirus; healthcare workers' hands can acquire various pathogens through contact with surfaces in the absence of direct patient contact [32]. Combining effective hand hygiene, PPE (personal protective equipment), appropriate surface cleaning, and reduction of surface contact can help reduce the transmission of such nosocomial pathogens.

In some clinical areas, such as the operating theatre, sterility is vital; however, the need for clinicians to maintain sterility has presented challenges when interacting with imaging systems using conventional input devices, e.g., mouse and keyboard. Users have found themselves unable to use PACS and other software easily in an aseptic fashion, resulting in an interaction gap between preoperative planning/image review and intraoperative control of key imaging. Despite the value of timely and comprehensive access to pertinent imaging in the sterile field, allowing operators to rapidly modify their therapeutic plan through an analytic decision-making process, existing solutions for controlling key imaging in the operating room are often time-consuming, inefficient, error-prone, and disrupt the workflow [15]. Often, to interact with key imaging, clinicians either have to break asepsis (introducing delays due to the need to rescrub), rely on colleagues to interact with systems on their behalf (potentially leading to undesirable cognitive and workflow disruption), or simply forgo interacting with imaging entirely (limiting options for additional planning opportunities during surgical procedures) [5, 6, 19, 23, 28, 41]. Such workarounds are inefficient and could impact patient outcomes.

Touchless interaction is the key to giving clinicians direct, sterile control of medical imaging systems. Touchless input modalities like speech and mid-air gesture provide a sterile alternative to traditional devices for interacting with PACS (since no surface contact is necessary), and recent advances in touchless technology make this the ideal time to consider practical aspects of their deployment in hospitals. A growing body of research across HCI (human-computer interaction) and medicine shows the potential of touchless interaction in medical contexts, e.g., [14, 16, 21–23, 27–31, 37].

Touchless interfaces have practical issues that limit their adoption. An important usability concern for a touchless user interface is the Midas Touch problem [24]. Sensor-based interactions are inferred from continuous streams of sensor data—in the case of touchless gesture input, these sensors typically respond to hand or body movements. Sensing systems often have a large but ambiguous

CHI '22, April 29-May 5, 2022, New Orleans, LA, USA

ACM ISBN 978-1-4503-9157-3/22/04.
https://doi.org/10.1145/3491102.3517512

input range [8, 10] and any body movements within that range may be recognised by the sensors. As a consequence, incidental actions (e.g., hand movements as part of another activity, like pointing to an object or waving to a colleague) may be detected by input sensors, incorrectly classified as intentional input, and mistakenly acted upon to cause an accidental or unintended effect in an interactive system. Such false-positive gesture recognition happens because the system incorrectly inferred a person's intent to interact [34]. The resulting unintentional input can cause frustration and may require additional actions to be taken to undo its effects [17]. For medical imaging systems, such unintentional input could be disruptive and may even go unnoticed, since health professionals are focused on their patients, with primary tasks that demand more care and attention. Hand gestures are often used by health professionals to communicate during medical procedures [23, 28, 31] and so being able to distinguish between interactive and non-interactive gestures is crucial for mitigating the Midas Touch problem and reducing the likelihood of costly unintentional input in critical scenarios.

In this work, we investigate clutching techniques for a touchless PACS interface with speech and gesture input. Clutching refers to the functionality of activating and deactivating control over a system, effectively giving users a way to *address their input* [3, 8] to a touchless gesture system and mitigate the Midas Touch problem. We implemented a touchless PACS interface with four clutching techniques (voice, gesture, gaze, and active zone). These were chosen from the literature as suitable candidates for medical usage contexts and are intended to help touchless systems determine when a PACS user intends to provide input to the system, rather than talking or gesturing to their colleagues. We performed a study exploring these with an ecologically valid sample of hospital-based healthcare professionals (n=34). Quantitative interaction measures and qualitative analysis of interviews with the clinicians provides rich insight into the usability of these clutching methods and their suitability for clinical use. Our findings also identify benefits and challenges of integrating touchless interaction into medical settings. We build on previous work by presenting a detailed investigation of clutching in this context and evaluate clutching as a key component of touchless interaction, rather than as a minor component of a larger interaction technique. In doing so, we take a key step towards the deployment of usable and effective touchless PACS interfaces, which have the potential to transform clinical activity.

Our contributions include: (1) a novel touchless PACS interface prototype that integrates four clutching methods adapted from the literature; (2) a mixed-methods experiment with 34 hospital-based health professionals, carried out in situ; (3) detailed findings about the efficacy and usability of these clutching methods, extending prior findings with added insight about their suitability for PACS usage and identifying compelling challenges for future research.

2 BACKGROUND

2.1 Touchless Interaction in Medical Settings

Touchless interaction modalities like speech and mid-air gesture are compelling for use in medical settings because they can allow health professionals to provide input without concerns of sterility. Sterility is crucial in medical settings, both for protecting patients who are often vulnerable to HCAI (health-care associated infections) and for protecting the well-being of health professionals (most recently highlighted by the COVID-19 pandemic). Touchless modalities allow users to interact directly and immediately without having to break asepsis, delegate commands to another person, or utilise many workarounds that have been developed for interacting with computers in clinical contexts [5, 6, 23, 28]. Many prototypes have shown the potential benefits of touchless input in medical settings [29], e.g., using mid-air gestures [14, 16, 21, 22, 30, 31, 33, 37], foot input [14, 22], voice [14, 27, 30], and proxemics [28]. A common finding in this body of work is that health professionals recognise the potential benefits of touchless input (especially with sterility and direct control over imaging systems). Good usability is key for these to translate to real working environments and such works have identified challenges of deploying touchless clinical systems in a usable and effective way.

Unintentional input is a key usability concern with touchless interaction because touchless sensing systems are 'always on' and attempting to recognise intentional input within a large room-scale space [8]. Unintentional input can be especially problematic in medical usage contexts, because teams of health professionals work in close physical proximity and frequently gesture and talk to each other during clinical work [6, 23, 28, 31]. Being able to differentiate between interactive and non-interactive gestures and sentences is therefore key for reducing unintentional input. This is important because unintentional input could have disruptive or unnoticed effects, e.g., when a health professional is communicating with a colleague and a system interprets this as a command. In this work, we investigate methods that address concerns of unintentional input in medical settings.

2.2 Clutching Interaction Techniques

One way of reducing the likelihood of unintentional input to a sensing system is to employ a *clutching mechanism* [17], interactions that allow users to signal their intention to *address* the interface [3]. Clutching mechanisms may require purposeful actions that act as a 'mode switch' [36] or show a person's intent. Alternatively, systems may infer intent from a set of constraints and contextual cues (e.g., position in the room or eye-contact with the display).

Speech interfaces often use wake words as a clutch; these are words unlikely to occur in everyday conversation, e.g., Alexa, OK Google, Hey Siri. These tell the system to 'wake up' and treat the following words as input, reducing the likelihood of false-positive recognition. Voice can also be used by other interaction modalities for mode switching, since it does not interfere with manual actions, like mid-air gestures. More complex voice clutches can be integrated into multimodal interactions and used alongside other input methods; for example, Put-That-There [4] used speech and gesture to identify operations and their parameters. Speech acted as a clutch, since a gesture without an accompanying utterance, or an utterance without a gesture, would not be treated as input. Voice clutches were first considered for novel medical imaging systems in the early 1990s. Hinckley et al. [18] investigated the use of tangible props for interacting with neurosurgical visualisations. A voice clutch was considered and discarded due to the performance limitations of speech recognition technology at the time;

they thought speaking the clutch phrases (*"move «prop»"* and *"stop «prop»"*) would introduce a frustrating delay and might distract from other tasks. However, these issues have not prevented the uptake of speech user interfaces in hospitals [5] and contemporary speech recognition is less affected by such latency.

Mid-air gesture systems can likewise use clutch mechanisms to avoid unintentional input [36]. Gestural equivalents of wake words can be used to indicate the beginning of a gesture command sequence. Like wake words, these gestures should be unlikely to occur incidentally. Clutch gestures can be discrete movements or poses that act as a mode switch prior to performing other gestures; e.g., finger snapping [7] or making a fist [36]. Alternatively, a clutch pose can be held continuously as part of another action; e.g., the hand-on-hip 'teapot' pose while performing gestures with the free hand [38], a finger-to-thumb pinch gesture while moving the hand [39], or an extended thumb while pointing with the index finger [11]. Confidence in user intention can be further increased through the use of a dwell period, where a clutch gesture is held for a brief period. These clutches are typically evaluated indirectly as part of a complete touchless user interaction sequence and so little is known about the actual user experience of clutching itself [36].

For clinical contexts, it is important to choose clutch gestures that will not occur during other activities or interactions with other health professionals, who often gesticulate to each other [23, 28, 31]. A further constraint is introduced by the need for sterility, which restricts where hands can safely move or be placed during interaction. For example, Strickland et al. [35] used a hands-above-head gesture as a clutch; this was otherwise unlikely to occur when viewing medical images and kept hands away from non-sterile surfaces. O'Hara et al. [31] discuss several other examples of clutch gestures that were found to be unsuitable in surgical contexts. For example, dwell was deemed unsuitable because health professionals would often pause for reflection while viewing images.

Finally, the intent to interact can also be inferred by *how* or *where* users perform actions. Baudel et al. [2] described the use of an "active zone", an area of space where sensed movements are treated as intentional input. This can reduce the available input space significantly by excluding input in other regions and, if clearly signified, can help users understand where input will be sensed [25]. Alternatively, information about body posture and gaze can be used to infer an intention to interact, as in existing surgical human-human turn taking, where a nurse delivers surgical instruments to a surgeon based on explicit requests (e.g., uttering "scalpel") and implicit requests expressed as body language [42]. Schwarz et al. [34] computed an 'intent to interact' score, finding that it improved false-positive recognition rate by inferring intention from body language and similar cues; however, users experienced uncertainty about its behaviour and preferred to combine it with gestures (e.g., raising a hand) as that improved the user experience and gave a sense of agency over the interaction. Zhou and Wachs present an early turn-taking prediction algorithm using Long Short-Term Memory, a type of recurrent neural network. Through experimentation, they found their algorithm could *"predict the incoming turn-taking intention much earlier than humans, thus providing early prediction capability instead of just classification"* [42]. Jacob et al. [20] used information about head and body orientation to determine when a user was intending to interact with an MRI (magnetic resonance imaging) system. While this reduced the false-positive gesture rate in their study, O'Hara et al. [31] suggested those contextual posture cues may be misleading in other usage scenarios (e.g., when looking at images with other health professionals). We consider these to be examples of *implicit* clutch mechanisms because the intention to interact is inferred automatically from sensor information, unlike *explicit* methods where users perform a deliberate action to mode switch (e.g., a gesture or speech command).

2.3 Clutch Interactions in Medical Settings

We identified a variety of clutch mechanisms for touchless interaction through a literature search and informed by our own prior work on this topic [8, 9]. In this work, we perform a direct comparison of four of these (illustrated in Figure 1): an unlock gesture, a spoken unlock phrase, active zone and gaze. These were chosen as representing explicit and implicit clutching mechanisms appropriate for medical contexts. In the case of explicit mechanisms, these comprised the most common touchless interaction modalities (hand gesture and speech), and for implicit mechanisms both physically constrained (active zone) and unconstrained (gaze) mechanisms, which may have particular implications within these contexts. Whilst some have been adopted in previous touchless medical imaging systems (e.g., unlock gestures [31, 35], speech [18], gaze [20]), there is only limited insight into their usability and efficacy in this setting.

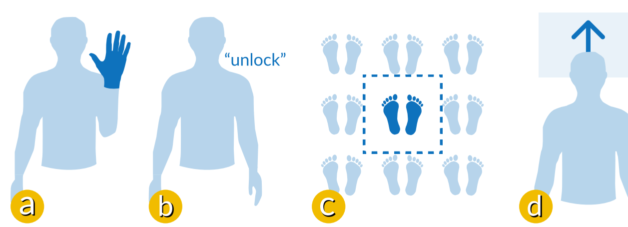

Figure 1: Our study looks at four clutch methods: (a) unlock gesture, (b) unlock phrase, (c) active zone, and (d) gaze.

Gesture clutches have been used in a variety of touchless user interfaces across many application domains, including medical imaging systems. For example, O'Hara et al. [31] described a touchless medical visualisation system that used an unlock gesture where both hands were held close to the body. Their work outlined many considerations for choosing a good clutch gesture (and other interaction design decisions), but did not evaluate it or reflect on how well it worked in practice. We include an unlock gesture in our study because it fits with our other gesture actions and is an example of a widely-used clutch method in other application areas. Our goal is not to find the 'best' unlock gesture, but to evaluate the idea of using a deliberate gesture to unlock a PACS interface.

Voice clutches have also been considered in medical settings. For example, Hinckley et al. [18] considered the use of key phrases for clutching in a tangible user interface for medical visualisation. We include an unlock phrase in our study because this is a ubiquitous mode switching method and complements the speech commands one might find in a speech-based PACS interface. Speech-based interfaces have been evaluated in medical contexts before, but our work focuses on its use for clutching, rather than evaluating the merits of speech input in this setting.

We also include the **gaze** and **active zone** methods, which infer intention to interact and unlock the touchless PACS interface accordingly. These have not been evaluated in this usage context before, but could be advantageous because explicit actions are not required by the user, potentially reducing the cognitive load in an already demanding environment. An intentional limitation of the active zone method is that it restricts the space where interactions can be sensed, which offer users less flexibility for interaction. However, in clinical contexts, users have restricted freedom for movement and will typically interact from key locations in well-defined workflows [31]. We include gaze as an implicit mechanism which does not have the same physical constraints, and is potentially low-effort (just gaze at the display). We thus evaluate both clutch methods to investigate their efficacy in this usage context and to see how their implicit clutch behaviour compares to the explicit clutch actions.

This paper extends prior work on clutch interaction techniques with a focused investigation of clutching for touchless medical systems. Whilst others have highlighted the importance of clutching to reduce unintended input and devised novel input techniques to address this problem, we directly compare four of these, provide detailed insight into how clutching affects the user experience of a touchless PACS interface, and consider the benefits of implicit and explicit clutch methods. This work addresses the limited usability evaluation of clutch interactions found in the literature and considers the extent to which the limitations of our chosen clutch methods apply in usage environments where users face many challenges that may affect interaction. Finally, we contribute a detailed qualitative exploration of the challenges and opportunities for deploying touchless UIs to other interactive systems in medical contexts.

3 USER STUDY

3.1 Study aims

We conducted a mixed methods user study to evaluate clutching methods for a touchless PACS interface, with an ecologically valid sample comprising a large cohort of practising hospital clinicians. We asked health professionals with PACS experience to complete a series of tasks with a generic touchless PACS interface, using four clutching techniques (outlined in the previous section) to (un)lock the touchless user interface. Our aim was to investigate how clutching affects user experience and to better understand the challenges of integrating touchless input into PACS workflows. This study addressed the following research questions:

- **RQ1:** How does clutching affect the user experience of using a touchless PACS interface?
- **RQ2:** How can clutching be effectively integrated into a touchless interface in a clinical setting?
- **RQ3:** What are the differences between several commonly used clutching methods for a touchless PACS interface?
- **RQ4:** What are the opportunities and challenges associated with touchless PACS interaction?

PACS are used in challenging usage contexts where users have many competing attention demands. We need to better understand how clutching affects user experience (**RQ1**), so we can minimise cognitive demand and avoid disrupting the important tasks that happen alongside PACS usage. By comparing a variety of clutching methods, we will be able to make informed recommendations about how to integrate these into a touchless PACS interface and similar clinical applications (**RQ2**). Our chosen clutching methods include explicit and implicit interactions, where there is a potential trade-off between user control (favouring explicit) and interaction demands (favouring implicit); we explore this trade-off, to see which methods are most appropriate for this context (**RQ3**). Our findings will give insights for researchers and designers creating touchless interfaces for health professionals, including PACS (**RQ4**).

3.2 Study design and procedure

We used a within-subjects design with four conditions, corresponding to the clutch techniques: (1) gesture, (2) speech, (3) active zone, and (4) gaze. These clutch techniques include both explicit (e.g., gesture and speech where the user performs an affirmative action) and implicit interactions (e.g., inferring intent from body position in the active zone or from eye contact with the PACS interface).

Participants were required to complete tasks using a custom PACS implementation (described later). We created our own system so that we could integrate all clutch interaction modalities and minimise potential bias as a result of user familiarity with existing PACS software. We chose frequently used PACS tasks (e.g., opening patient files, image navigation and manipulation, viewing patient reports) that required sustained attention and interaction. The system was placed on a table with a 1 m^2 active zone marked on the floor, 1.5 m in front of the display.

We asked participants to complete off-screen secondary tasks while carrying out the PACS tasks. The inclusion of these secondary tasks was informed by previous research with healthcare professionals who described a need to lock the system to perform tasks such as interact with a patient [6]. This dual task design was also similar to the study design presented by Schwarz et al. [34] in their evaluation of implicit clutching techniques. These were performed in front of the input sensors, necessitating use of the clutch interactions to avoid false-positive input recognition. Divided attention and multitasking scenarios like this are representative of situations where PACS are used; e.g., discussing images and reports with colleagues, referencing other paperwork and materials, attending to patients, etc. For the secondary task, participants had to step away from the system, read a question about an on-screen report, then write down the answer before returning to the system. Users were allowed to consult the report when answering the questions, so often had to look at and/or use the system to navigate. This was intended to emulate a PACS user switching attention between a display and attending to a patient or discussing images with colleagues.

At the start of the session, participants were trained in using the touchless PACS interface. Video tutorials demonstrated interaction and participants could practice using the system for as long as they liked. The lead researcher was present in the room to answer questions and assist during the training phase. Each condition (i.e., clutch technique) had one block of tasks and condition order was balanced using a Latin square design. Each block had 14 on-screen PACS tasks and off-screen tasks, as described before; tasks were presented in a randomised order.

3.3 Measurements

Each study session was video-recorded and interaction data was logged by our touchless PACS system. Interaction logs were manually annotated by the lead researcher with help from the videos, so that we had complete logs with all user actions, even if the system did not detect them. This enabled us to capture more complete information about interaction, e.g., identifying when users performed an incorrect action, 'correct' input action not being recognised, etc.

We measured the total interaction time for each task, ending when the final PACS action was completed. Within this, we measured the cumulative time the system was unlocked during each block, as this could suggest if a clutch method was at risk of false-positive recognition (e.g., if the system is actively sensing input more often). We logged all clutch interaction events and timestamps, so that we could count the number of transitions between clutch states (i.e., locked/unlocked) and the total time spent in each state. We measured clutch success rate as the ratio of successful clutch actions to total clutch actions (i.e., including unsuccessful attempts). After each block of tasks, participants completed the NASA-TLX survey as a measure of task workload [13].

We used repeated-measures ANOVA (with post hoc t-test comparisons) to analyse the effect of condition on timing data, as this data met the assumptions for this approach (i.e., continuous data, from a normal distribution, within-subject design) and is commonly used for interaction time measures. We used Friedman's test (with post hoc Nemenyi tests) to analyse the effect of condition on the number of clutch transitions, success rate, and TLX survey ratings, as these data are not appropriate for parametric tests (i.e., ANOVA).

We conducted a semi-structured interview after the final task block, to complement the quantitative data with qualitative feedback. Interviews were loosely structured around findings from previous research that investigated the needs and experiences of clinicians when accessing medical imaging [6]. These interviews aimed to assess user satisfaction when using touchless interaction, assess their thoughts on the four clutching techniques, explore their views on touchless control as an alternative to traditional mouse and keyboard interaction, and discuss attitudes towards touchless interaction with PACS. Interview prompts are available in the supplementary material.

Interviews were audio-recorded and transcribed, then responses were thematically analysed by two of the authors. We coded transcripts using the Framework Method described by Gale et al. [12], using the constant comparative method to identify and develop themes in the transcripts. After an initial round of coding, these were structured into the higher level themes discussed later. These qualitative findings provide additional insight, complementing the quantitative results, and drawing on the experiences of a large sample of practising hospital clinicians and PACS users.

3.4 Design and implementation

3.4.1 PACS interface. We developed a generic PACS interface for this study. This used a simple DICOM (Digital Imaging and Communications in Medicine) loader for loading authentic medical imaging files, and supported PACS tasks like image browsing, image manipulation (e.g., zoom, rotate, invert, flip), and viewing reports. The final set of PACS tasks implemented combined several typical PACS

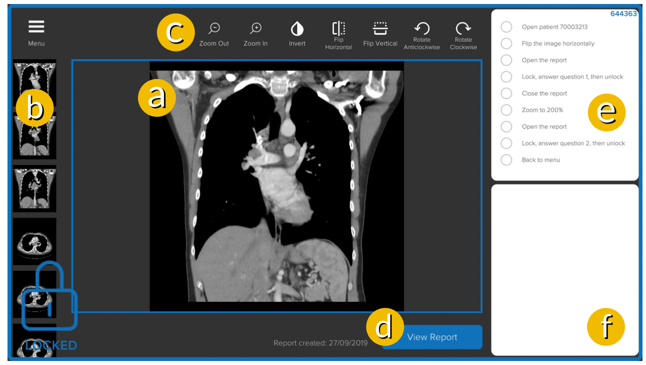

Figure 2: Screenshot of the PACS user interface (a–d), experiment tasks (e) and touchless interaction feedback panel (f).

functions presented by Madapana et al. [26] with input from clinicians. Users provided input using speech commands for direct manipulation (e.g., "rotate image clockwise", "zoom to 200%"), with corresponding mid-air gesture commands.

Figure 2 shows our PACS interface. This has a similar layout to existing PACS software: selected images are shown in the central area (a); a sidebar shows the set of images in the patient file (b); available image manipulation actions are shown in a toolbar (c); and patient reports can be viewed in a new window using the 'View Report' button (d). Task instructions and interaction feedback were also shown in the user interface: a list of PACS task instructions were shown in the top right sidebar (e) and interaction feedback was given in the bottom right sidebar (f).

For the experimental tasks, we used a dataset built from open-source DICOM images (.dcm files), acquired from online repositories like DICOM Library. Images were chosen to represent common medical imaging. Patient metadata was modified to standardise patient IDs across all experiments. In order to remove the need for specific medical knowledge (since our participants had different areas of expertise), a standardised report was used for all experiments, allowing participants to answer secondary task questions, e.g., "what is the patient's name", directly from the reports without interpretation requiring specialised medical knowledge.

3.4.2 Touchless PACS commands. Several common PACS functions were supported, including: open patient, open/close report, set image zoom %, rotate image (anti)clockwise, invert image, flip image horizontal/vertical, return to menu. Voice and mid-air gesture commands were mapped to PACS functions, e.g., "zoom to 150%" or raising left hand to shoulder height to close a patient report. Interactions were introduced during the tutorial at the start of the study session and participants were able to practice using them.

3.4.3 Clutch interactions. Our PACS prototype supported all four clutching interaction modalities. These were used to transition between locked and unlocked states in the touchless PACS user interface. When the system was unlocked, it would respond to the speech and gesture commands.

The **mid-air gesture clutch** was a single raised hand, held above shoulder height for 1000ms (Figure 1–a). This toggled between the locked and unlocked states. We chose this as a mode switching gesture because it can be robustly detected by a simple depth sensor, even when users are at a distance. More importantly, it requires no motion and allows the users' hands and arms to be kept close to the torso; this is crucial for maintaining sterility (e.g., when in surgery) and would be suitable when standing beside other clinicians. From a technical perspective, this action can also be recognised from rudimentary skeleton tracking, ideal when users' hands may be occluded by PPE. It is also an action unlikely to occur in situ, as clinicians are not likely to stand with a single arm raised.

The **voice clutch** used "lock" and "unlock" to switch between lock states (Figure 1–b). We chose these because they are not typical operations found in a PACS interface so would not overlap with commands. However, we note our focus was on the use of a speech clutch in general, rather than the specific phrases used to (un)lock the touchless interface.

The **active zone clutch** used body position to implicitly (un)lock the touchless interface (Figure 1–c). We defined an active zone relative to the position of the screen showing the PACS interface: a $1\,m^2$ square, starting from a distance of 1.5m from the centre of the display. This was marked on the floor with tape. We used the Kinect's torso position to determine when the user was in the active zone (as their legs would not be visible if they were standing behind an operating table). This region gave users some flexibility in where they stood, but was appropriately sized so that, along with gaze oriented towards the screen, we could reliably infer an intention to interact. When the user was standing in the active zone and looking at the screen, the system was automatically unlocked and responsive to input actions.

The **gaze clutch** used estimated gaze direction to implicitly (un)lock the touchless interface (Figure 1–d). We used head posture to estimate gaze direction as this could detect users over a greater distance and would be less affected by PPE visors. If the user was facing the screen showing the PACS interface (head direction within $\pm 15^{\circ}$ of the Kinect sensor), the system was automatically unlocked and responsive to input actions; conversely, if they were facing away from the screen, the system was locked. We could have alternatively included body posture for a more robust estimation of intention to interact (e.g., as in [34]). However, PACS usage will often require users to divide attention between multiple tasks, so users may glance to and from the screen without necessarily turning their body towards it. Therefore, we use head direction alone to infer engagement with the interface.

3.4.4 Apparatus. Our software was implemented on the Microsoft Azure platform. We used the Language Understanding service from Azure Cognitive Services to map users' speech commands to user interface operations. We used the Face service from Azure Cognitive Services for facial detection, which determined when the user was engaged and looking at the screen. We used the Body Tracking SDK (software development kit) from the Azure Kinect DK for user detection; this was used to track position in the room (active zone) and body posture (gestures).

For input sensing, we used the Microsoft Azure Kinect device. This device has a 1MP depth sensor with room-scale range for user tracking, and an omni-directional microphone array for robust speech tracking. We used this because it can be used to recognise all four of our clutching methods, supports the touchless interactions

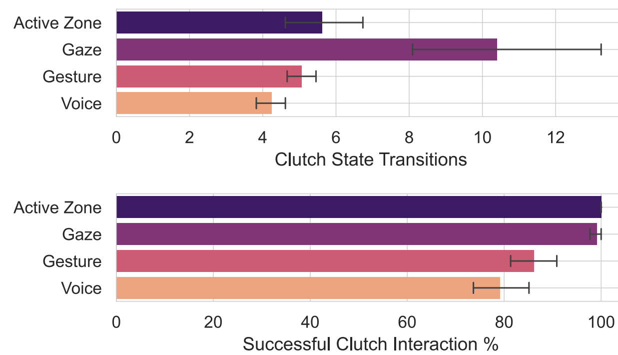

Figure 3: Number of transitions and clutch action success rate. Error bars show 95% CIs.

with our interface, and integrates well with the aforementioned cloud services for accurate and reliable interaction sensing. For visual output, we used a 30" monitor raised on a table surface.

3.5 Participants

We recruited 34 healthcare professionals (32 practising clinicians and 2 dedicated PACS staff) through existing contacts and through a university-affiliated teaching hospital. All were health professionals with PACS experience, with varying grades of seniority in the hospital. Table 1 gives an outline of participants, their speciality and years of PACS experience. We asked participants if they had experience with touchless computer interfaces; 21 had used voice interfaces, 10 had used gesture or motion controls, and 7 had used other devices (e.g., foot pedals). 18 participants wore face masks during the study due to COVID–19 restrictions; facemasks and PPE are often worn in situ (especially in surgical contexts). Participants were not compensated for the approximate 45 minute study time. Research ethics approval was obtained from the lead author's institutional ethics review committee.

4 RESULTS

We now present analysis of interaction measurements (Section 4.1) and NASA-TLX survey results (Section 4.2). This is followed by discussion of the key themes from the qualitative analysis of the interviews (Section 4.3).

4.1 Interaction

Figure 3 shows the number of clutch transitions for each condition and the ratio of successful clutch actions. Figure 4 shows the mean task time for each condition and the mean cumulative time in the 'unlocked' state for each block.

There were a mean of 6.31 (SD 4.86) transitions per interaction task. Friedman's test found a significant effect of method on the number of transitions: $\chi^2 = 27.8, p < .001$. Post hoc Nemenyi tests found that Gaze had more transitions than Active Zone ($p = .007$), Gesture ($p = .02$) and Voice ($p = .001$).

Mean clutch transition success rate was 91%. Friedman's test found a significant effect of method on success rate: $\chi^2 = 44.1, p < .001$. Post hoc Nemenyi tests found higher success rate for Active

Role / Speciality	N	Experience	Role / Speciality	N	Experience
Senior House Officer (equiv. Resident)	13		Student	3	0–4 yrs: 3
→ *General*	2		→ *Radiography*	3	
→ *Orthopaedics*	1	0–4 yrs: 11			
→ *Paediatrics*	1	5–9 yrs: 2	Specialist Registrar (equiv. Fellow)	13	
→ *Obstetrics & Gynaecology*	3		→ *General*	4	0–4 yrs: 2
→ *Anaesthetics*	1		→ *Obstetrics & Gynaecology*	2	5–9 yrs: 10
→ *Surgical*	5		→ *Paediatrics*	1	10–14 yrs: 1
			→ *Orthopaedics*	1	
Consultant (equiv. Attending)	3		→ *Surgical*	5	
→ *Radiology*	2	10–14 yrs: 2			
→ *Surgical*	1	20+ yrs: 1	Other	2	5–9 yrs: 1
			→ *PACS Manager*	1	10–14 yrs: 1
			→ *PACS Clerical Officer*	1	

Table 1: Participant role, speciality and years of PACS usage experience.

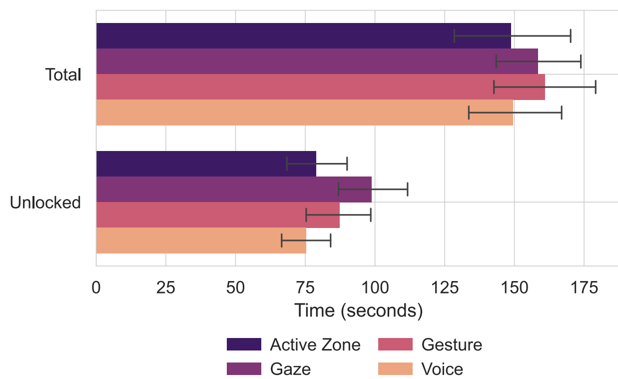

Figure 4: Total task time and total time in the unlocked state. Error bars show 95% CIs.

Zone than Gesture (p = .008) and Voice (p = .001), and higher success rate for Gaze than Gesture (p = .02) and Voice (p = .001).

Mean time-on-task for each block was 154.3 seconds (SD 53.5 seconds). A repeated measures ANOVA did not find a significant effect of method on time: F(3, 96) = 0.61, p = .61.

Mean time unlocked was 85.0 seconds (SD 34.3 seconds), which was 55% of the overall task time. A repeated measures ANOVA found a significant effect of method on total time unlocked: F(3, 96) = 3.42, p = .02. Post hoc t-tests found time unlocked was higher for Gaze than Active Zone (p = .03) and Voice (p = .001).

4.2 Task-Load Index

Overall task-load index (TLX) was calculated as the mean of the six components (on a scale of 0–100) [13]. The mean TLX was 23.1 (SD 13.9). Figure 5 shows mean TLX and CIs for each condition, including overall score and six components. We investigated the effect of condition on overall TLX and on the six TLX components: mental demand, physical demand, temporal demand, perceived performance, effort, and frustration. Note: lower scores are 'better'.

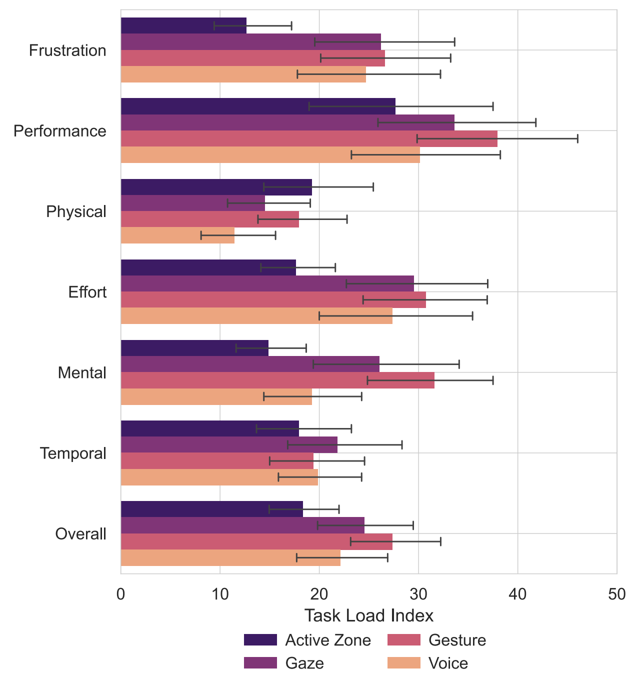

Figure 5: Mean TLX scores. Error bars show 95% CIs.

Table 2 shows Friedman's test results and significant results from post hoc comparisons. As can be seen from Overall TLX, Active Zone was less demanding than Gesture and Gaze, and Voice was less demanding than Gesture. These can be partly explained by key differences in the individual components; i.e., Active Zone was less mentally demanding and frustrating than Gesture and Gaze, and Voice was less mentally and physically demanding than Gesture.

4.3 Interview Findings

In this section we discuss key themes from our analysis of the interview transcripts. Participants are identified anonymously by a

Rating	Mean	χ^2	p-value	Sig. Comparisons
Overall	23.1	21.9	<.001	A < GE (p = .001), A < GZ (p = .006), V < GE (p = .03)
Mental	22.9	26.3	<.001	A < GE (p = .001), A < GZ (p =.02), V < GE (p = .007)
Physical	15.8	12.6	<.001	V < GE (p = .02)
Temporal	19.7	6.08	.11	
Performance	32.3	5.83	.12	
Effort	26.3	14.9	.002	A < GE (p = .005)
Frustration	22.5	17.3	<.001	A < GE (p = .008), A < GZ (p = .007)

Table 2: Mean task-load index scores (overall and each of the six components), Friedman's test results and, if appropriate, significant Nemenyi test comparisons (A: Active Zone, GE: Gesture, GZ: Gaze, V: Voice).

unique number and their professional role: PACS → Admin staff (e.g., PACS manager); STDT → Student; SHO → Senior House Officer; SREG → Specialist Registrar; CS → Consultant Surgeon; CR → Consultant Radiologist. We first discuss comments about the clutching interactions (in Section 4.3.1), followed by a discussion of touchless interaction in general (in Section 4.3.2).

4.3.1 Clutching Interactions.

Voice. For many of the health professionals, voice was the *"most straightforward"* [SHO 13] and *"easiest, most intuitive"* [SREG 3] clutch method to use. Most (21) had prior experience using voice control (e.g., digital assistants), so familiarity was perceived as an advantage. Another advantage of voice in this context was that the unlock and lock commands could be issued from anywhere: *"it removes the element of where you're standing"* [SHO 13] and did not disrupt tasks elsewhere in the room, unlike the other methods that required standing within line of sight of the Azure Kinect sensor.

Voice was not without issues, however. Some users reported it was *"frustrating when the voice wasn't listening to you to unlock and lock"* [SHO 10]. This was often the result of misrecognising speech commands: *"not recognising certain phrases from myself"* [SREG 8]. Misrecognition often occurred due to factors out of the user's control. An important observation was the potential impact of PPE: *"voice recognition when you're wearing an FFP3 mask might be a problem"* [SREG 12] because it muffles speech.

Many users pointed out the negative effect of ambient noise in a busy environment with other people, e.g., SREG 1 and SREG 6 both said the operating theatre can be very noisy and this could cause issues when detecting the clutch commands, whilst SREG 6 and PACS 1 highlighted the potential for false-positive input from others talking nearby who were not part of the interaction. Being around others could also cause other difficulties when using speech to unlock the system, as the user may need to interrupt ongoing conversations: *"it would be great if you didn't have to stop talking, issue a command, and start talking again"* [SREG 10, SHO 10].

We used "lock" and "unlock" as our key phrases for (un)locking the system, although some participants noted that these may not be appropriate for all usage contexts within the hospital. For example, *"locking and unlocking are a common term in orthopaedics for actually fixing things, so it might cause confusion"* [SREG 5]. In practice, the key phrases for clutching will need to be chosen carefully, to avoid false-positive input from words that have meaning and may occur often in that context with the intended health professionals as users.

Gesture. Gesture input was a novel interaction modality for most users (only 10 had prior experience, mostly from motion controls in games, e.g., Nintendo Wii or Xbox Kinect). Whilst there was only one clutch gesture (which acted as a mode switch), some users noted the mental demand associated with remembering this as part of the wider gesture set: *"it [was] a little tiring to remember the gestures"* [SREG 10, CR 1]; this was also reflected via TLX scores. Some participants noted that this was largely due to the novelty of these gestures and that they would become *"second nature"* [CR 1] with experience; indeed, health professionals *"are invariably tech-savvy and adaptable... it's not rocket science"* [CR 1] and gestures would become easier to learn and use with experience.

Similar to the voice clutch modality, participants raised concerns about the negative effects of PPE on the clutch gestures, both in terms of recognition difficulties and their own inability to perform certain gestures: *"[they] probably wouldn't be captured, just because you're wearing a gown in theatre, if you were scrubbed"* [SREG 1], *"could never use that one if I was wearing any type of scrubs or gown"* [SHO 5]. This poses an interaction challenge, as those with the greatest need to interact with PACS in theatre are likely to be the health professionals who are scrubbed.

Gaze. For a small number of health professionals, the gaze clutch method was *"definitely the best"* [SREG 10], *"my preferred method"* [SHO 12], as it streamlined interaction and required no additional effort to begin issuing speech or gesture commands for the PACS interface: *"you do not have to raise your hand or say anything, that was good"* [STDT 1]. Participants also commented positively on the lack of clutch errors, something supported by the data (99.1% recognition). However, it was the least preferred and the majority found it challenging—it was *"very unnatural feeling"* [SHO 13].

From the discussions, it seems most participants experienced a clash between their mental models of the gaze clutch and its actual behaviour. Gaze was an implicit clutch which automatically unlocked the touchless PACS interface whilst a user was looking at the system. For many users, this was undesirable—they wanted to be able to look at the system without it clutching for input (e.g., to refer to the patient files for the secondary task): *"I was looking at the screen to get the report, I automatically kept engaging it"* [SREG 3]. This loss of control over the clutch state—the inability to prevent the system unlocking when glancing at the screen—was frustrating to them. As a result, some described strategies to avoid unlocking while looking at the screen, e.g., sideways glances so the system did not think were not looking.

Despite frustrations with the gaze unlocking, some participants found it to be a particularly compelling modality for initiating interaction; they described existing frustrations with the need to repeatedly unlock computers each time they wish to use the system, due to aggressive data protection and security policies: *"locking full-stop is what bothers me the most about PACS... if there was an option to never lock throughout the whole operation that would be useful"* [SREG 3]. It was suggested that an improved version of gaze, used alongside facial recognition, could address such concerns—unlocking the system from a security perspective *and* from an input recognition perspective.

Active zone. The active zone clutch was notable for the lack of frustrations expressed by participants during the interviews. As might be expected from its 100% recognition rate, *"I don't think it missed a beat once, that was good"* [STDT 1]. Unlike gaze, which was also an implicit clutch method, participants seemed positive about the automatic (un)locking with the active zone: e.g., *"I think that's good for safety, it means you don't have to unlock it ... [then] you just walk away and it automatically locks"* [SHO 1]. Participants also suggested this could simplify data protection practices: e.g., *"I can see the active zone being useful, like when you walk out of your office, it will lock it automatically"* [STDT 1].

A key advantage of the active zone method was that it was a good fit for existing PACS usage habits and could integrate with existing workflows and practices. Participants explained that in real PACS usage contexts, they will often move to a computing terminal, issue commands via mouse and keyboard, then step away again. Active zone required similar behaviours, e.g., *"stepping away is something you'd be doing anyway"* [SREG 7].

General views on clutching. Participants generally appreciated the value of clutching as a tool for avoiding accidental input—*"the most important part of this"* [SHO 10]—especially in situations where having the right information on screen is vital. As mentioned with gaze previously, security and data protection also came up in discussion about clutching. The need to authenticate before PACS usage adds friction and, whilst out of the scope of this research, many saw the potential for touchless interaction to aid authentication. Common suggestions included using facial or speech recognition to identify the active user of the system, such that clutching and authentication could happen at once.

4.3.2 Touchless Interaction. After discussing the clutch methods, we asked participants to discuss their thoughts on touchless PACS interaction in general. Five main themes emerged during analysis: (1) Workflow: about how touchless interaction could impact existing clinical practice; (2) Sterility: about benefits of touchless input for sterility, especially during the COVID-19 pandemic; (3) Environment: about challenges of the usage context; (4) Adoption: about integrating touchless input into the health profession; and (5) Applications: about promising uses for touchless input.

Workflow. Many saw the potential for touchless interaction to improve their workflow and were excited about its potential: e.g., *"it's the future of what we're doing in PACS... you just look at the screen, it will come on, you just say open that patient, it will just come up... it's gonna replace keyboard and mouse I think, you know, it's just a matter of time"* [CR 2]. For most, the key benefits were the potential for more efficient interaction: e.g., *"much easier, quicker, much less frustrating"* [CS 1] and *"it's innovative and potentially the future, and allows one to potentially be more efficient"* [CR 1]

Also important was the ability to interact with PACS directly, which is not always possible in situ. To maintain sterility and avoid cross-contamination, health professionals cannot always directly use computing devices. Instead, they need to give commands to another person who controls the system on their behalf, although this is not without friction: e.g., *"what we usually do was ask one of the scrubbed nurses to open up, but they don't necessarily [know] what image sequence you wanted to open up"* [SHO 2].

Touchless interaction avoids this indirect input and lets the health professional interact directly: e.g., *"you have to ask for someone to stay with you for several hours to look at the scan, [doing it yourself] is really very helpful"* [SREG 13], *I think it would be really useful in theatre if we wanted to scroll through images while we're scrubbed and couldn't use a mouse and keyboard* [SREG 5], and *"you're trying to avoid having to ask other people in the room, such as the circulating nurse who may not be familiar with the technology... certainly if you can talk to it, or gesture to it to pull up the exact image that you need, I think it would be very useful"* [CS 1].

Whilst these benefits are compelling, it was clear that reliability is a key factor in willingness to adopt a touchless PACS interface: e.g., *"nothing's 100%, if you can rely on it 95%, people will engage... [but if not] people have something that works, even if it's slower, they know it works"* [CR 2]. Some noted it would take time for touchless interaction to be integrated into their workflow, due to familiarity with existing input modalities: e.g., *"I'd prefer a touchless system in the long run, but all systems require you to kind of engage with it, you know, and get used to it... at this moment in time, just more used to mouse and keyboard.* [CR 2], *"I'm used to using the mouse and keyboard... it just takes time, because you're thinking about what you're doing before you do it... with a bit of practice, it probably would be as easy"* [SREG 6], and *"when you're doing a repetitive task, such as we do in radiology, it would become second nature"* [CR 1].

Participants thought different interaction modalities would be better suited to different tasks within the PACS workflow. Speech and gestures would be ideal as shortcuts for simple and repetitive actions like zooming, panning, moving through image sequences, but less suited to more complicated tasks: e.g., *"once it is kept simple, it is better than mouse and keyboard... but if there was a command for zooming in, raising your hand or something, yeah... for anything more complicated, I would prefer mouse and keyboard"* [STDT 1].

Fatigue. Among those health professionals that commented on the effort involved, they generally did not find using the touchless system to be fatiguing, though it was noted that that extended use may result in it becoming more so. Several participants noted that recalling gestures was demanding; however, users felt that over time, their familiarity with using the system would increase and mental fatigue would decrease. *"I found it a little tiring to remember the gestures. But if you learn them, they will become far more intuitive than the current system."* [SREG 10].

Sterility. One of the main perceived benefits of touchless interaction was that it would allow health professionals to interact with PACS in sterile conditions and reduce cross-contamination between surfaces and people: *"it decreases kinds of cross-contamination in* *a hospital setting"* [SREG 11]. Touchless input can overcome this challenge: e.g., *"we shouldn't be touching things, I don't think we should be touching things... I should be able to sit at a computer and be able to sift through, use a hand gesture, to sift through a scan"* [SHO 10], *"I think in theatre I can imagine it being used a lot, say if like somebody is scrubbed and wanted to see a scan and there was a big monitor on the wall, I think it'd be really useful... not even in theatre but in other scenarios where people are scrubbed or in sterile environments"* [SREG 6], and *"I envision touchless interaction being particularly useful in high-risk areas where there is a high risk of contact being made, such as a busy ED ward in a COVID setting or theatre, in what is meant to be a sterile area"* [SREG 8].

The significance of sterility was amplified by the COVID-19 pandemic, during which this study took place. Social distancing and increased focus on sterility had a disruptive effect on workflow and interaction with computing systems, and is likely to have a lasting effect on shared hardware in the health profession. This increased the perceived benefit of touchless input for PACS and other computing tasks: *"because of this pandemic and COVID, everything is being moved towards minimal touch or touchless, so yes, why not [PACS displays]"* [SHO 11], *"I was very pleased with it, especially in times of COVID where we're trying to minimise contact with things, so I think it's a great idea... it would be a perfect time to introduce this into the wards"* [SHO 9] , and *"I think it's exceptionally useful, particularly at the moment in the middle of a COVID pandemic when we're all sharing the same keyboard, and maybe aren't paying attention to sanitising between them, so certainly from that point of view I'd feel a whole lot happier with [touchless]"* [CS 1].

Environment. Whilst this study took place in several hospital settings, the locations were quiet and free from disruption for the study sessions. In practice, the usage context is busy, noisy, perhaps even chaotic, and many participants identified challenges that might affect reliability and usability. As mentioned above, PPE could cause interaction difficulties, e.g., facemasks muffling speech and protective gowns causing gesture detection issues. No mention could be found in the literature regarding the possibility of bodily substances, e.g., blood, having an effect on infrared reflectivity that would impact on the gesture recognition process, but this issue would deserve further investigation.

Health professionals often work with others and the presence of bystanders was seen as a potential issue for gesture recognition: *"it may not pick up your [gestures] or may not see you exactly doing them"* [SHO 12]. Having others nearby could also cause issues of occlusion and ambiguity over who the active user is, especially for the active zone clutching method: e.g., *"one barrier I envisage is space, particularly for the active zone feature"* [SREG 8], since that might create an area of the room that other health professionals need to avoid standing in.

Background noise was seen as a key problem for voice recognition: *"hospitals are noisy, noisy places"* [SREG 6], including conversations between others that might affect the PACS interface: e.g., *"if you were talking and then I was talking, the system wouldn't understand"* [PACS 1] and *"[it would be good if] it isn't distracted by other voices speaking around it... it might pick up their voice"* [SREG 3]. Recognition failure was one concern with noise, but false-positive actions could also cause disruption, e.g., *"more importantly, if [I am] spending an hour on a report and someone walks in and says a word like 'cancel' for whatever reason, and it deletes everything, no one's going to be happy"* [SREG 12].

Some noted the importance of room layout, so that displays would be clearly visible and input sensors would be able to capture input: e.g., *"sometimes you operate from the top of the patient or the bottom of the patient, so looking back at the screen might be an issue"* [SREG 12] and *"I'm short-sighted, so I have to stand particularly close to be leaning in to see, I guess you could just always reposition the camera if that was the case"* [SHO 12].

Adoption. There were many comments about factors that may affect the adoption of touchless interaction in hospitals. Two users suggested people may feel uncomfortable using touchless input actions around other people: *"you may think you look a bit foolish doing the arm movements, but overall, if you have sort of a quiet area, it's definitely better"* [SHO 12].

Many of the health professionals identified the need for training and discussed the learning curve they experienced during the study: *"it did take me a little bit of time to realise how to use it properly"* [CS 1] and *"you might need just to train them well"* [SREG 13]. There was a positive outlook on this by some: *"there has to be good training and adaptation...I wouldn't envisage that as a problem"* [CR 1].

One participant suggested that training could also encourage adoption, by helping novice users see the time-saving benefits of touchless input: *"find the 10 or 20 things that they actually use a computer for and say this is how you do it, and it's actually going to be much quicker for you"* [CR 2]. Demonstrating the benefits could help encourage those who are reluctant to change working practice, which is *"a pretty systemic issue"* [SREG 2] and so *"showing people that it is actually going to make their life easier is probably the most important thing"* [CR 2].

Applications. Our focus in this study was generic touchless PACS usage, but many of the health professionals discussed other areas in the medical profession where touchless interaction could be useful. The operating theatre was commonly suggested, due to the need for complex imaging but inability to directly interact with it: e.g., *"in the operating theatre, if we're doing a complex case that relies on complex imaging such as cross-section CT scanning angiography where you're trying to match the image to the operative site... you're trying to avoid having to ask other people in the room, such as the circulating nurse who may not be familiar with the technology... if you can talk to it, or gesture to it to pull up the exact image that you need, I think it would be very useful"* [CS 1].

Radiology was also suggested as a compelling use case, where voice control is desirable as it does not require body movement: *"I would see it as being useful in both diagnostic and interventional radiology, whereby you don't have to be employing other parts of your body to do stuff that you're trying to concentrate on... voice control would be handy in intervention, to take an image to mag up, rather than having to instruct a radiographer or to do it yourself* [CR 1].

Finally, the catheterization lab was another environment where touchless interaction could provide workflow improvements. Like the other suggestions, touchless input would remove the need to work through a technician to control the system: *"in the cath lab, if you want to look at an old image there is a technician who sits outside who will load the images, like a radiographer, for the person* *who is performing the procedure to have a look at... usually, you'd look at it before the procedure, but if you needed to remind yourself, maybe it would be useful to be able to ask for it to come up—if you could say 'load it up on screen two'... if you had control of loading and scrolling through the image at a certain speed... you're not trying to communicate probably quite a sophisticated interpretation through someone else* [SREG 3].

5 DISCUSSION

Touchless user interfaces are compelling for hospital environments because they can reduce the spread of pathogens. This is more important than ever after the COVID-19 pandemic heightened awareness of the risks with shared input devices—especially in hospitals. Touchless input also has the potential to improve PACS usage through faster interaction and by allowing direct input without the need for a proxy user. For touchless technology to be deployed successfully in this context, it needs to be reliable, help rather than hinder clinical practice, and be easily usable in mentally demanding situations. We looked at four methods for mitigating false-positive input, a crucial aspect of reliable touchless interaction. Our quantitative and qualitative findings provide insight into how touchless interaction can be integrated into clinical contexts, and can inform the design of clutching methods for better input in this context.

5.1 Touchless Interaction in Clinical Contexts

Our study evaluated a touchless PACS interface with a large diverse sample of health professionals in hospital environments. One of the key strengths of our contribution is that we evaluated with real PACS users, gaining valuable insight into their needs and wants for a touchless user interface. We learned about how touchless interaction could benefit PACS usage and saw desire for adding touchless capabilities to other hospital systems.

Immediate benefits could be found in any aseptic environment where imagery access is vital, e.g., the operating room or during intraventional radiology **[RQ4]**. The *Workflow* and *Sterility* themes show that existing workflows are built around maintaining asepsis: i.e., avoiding contact with input devices and shared surfaces, issuing commands to a nurse or intermediate user, and anticipating imagery needs in advance. Other work notes that users may ignore imagery entirely when input is not possible due to sterility concerns [6], which is not ideal for patient outcomes.

Participants recognised that touchless technology can address these barriers to interaction: it gives them direct control and allows interaction whilst maintaining asepsis **[RQ1]**. Participants expressed a strong desire for this level of control, citing how much it could improve their workflow. Touchless input can also remove a source of frustration and inefficiency: e.g., the challenges of explaining precise imagery needs to a non-expert. In the worst case, breaking asepsis for interaction then scrubbing back in can add 15 minutes per event. Another potential benefit of touchless input was the faster interaction it affords, e.g., by allowing users to issue a speech or gesture command from anywhere in the room, even without taking their hands away from a patient. These findings are consistent with prior work, which has also highlighted the benefits of touchless input in this context (as discussed in Section 2).

There are many challenging aspects of the clinical usage context, however, as revealed by the *Environment* theme **[RQ4]**. Gesture recognition was sometimes impeded by protective gowns or visors, and participants noted that crowded environments could confuse input sensing. Similarly, speech recognition could be affected by facemasks and visors. These ecological challenges have implications for interaction design: e.g., which gestures can be robustly sensed through PPE, which gestures require the least physical space, which commands are less affected by muffled speech? Similarly, there are implications for deployment: e.g., what is the most suitable sensing technology, where is the optimal sensor placement? Such questions provide an agenda for future research, with relevance beyond clinical application areas (e.g., industrial settings where PPE, ambient noise, and crowded environments are also common).

We found that background noise and ambient sounds could also be challenging for voice control in certain hospital environments; e.g., the operating theatre can be particularly noisy. A voice controlled touchless system would need to work reliably in such noisy spaces, as participants suggested repeating commands would negate any efficiency benefits. Another auditory challenge comes from other people; clinicians often work alongside many people. Nearby conversations could disrupt speech recognition and there was a sense of frustration that a touchless PACS user would need to interrupt other ongoing conversation so that speech commands could be issued in silence. Deep learning solutions have been demonstrated for differentiating speakers. While such approaches are currently limited by decreasing accuracy with increasing numbers of speakers, further development could help address this [1].

We used a Microsoft Azure Kinect in our prototype system, a commodity device with powerful SDK support that simplified development of our touchless PACS interface. Low cost devices and emerging sensing frameworks (e.g., Microsoft Azure, Google MediaPipe, NVIDIA DeepStream) will make it easier and cheaper for robust touchless user interfaces to be developed on the scale that, e.g., regional hospitals or a national health service will require.

5.2 Clutching Methods for Touchless Input

Our main aim was to investigate clutching methods for a touchless PACS interface. Clutching gives users a way to indicate intent to interact; this helps sensing systems overcome the Midas Touch problem [24] by mitigating false-positive gesture recognition (whereby incidental body movements are mistakenly treated as input). This is an important problem to address, because touchless input offers great potential for PACS, but the Midas Touch problem could cause disruptive unintended input and frustrate users. A variety of clutching methods have been proposed in the literature and we directly compared four, bringing clarity about how they perform in clinical usage contexts. Our findings extend prior discussions [31] of clutching methods that were focused more on technical issues (e.g., gesture segmentation) rather than on usability and user experience. We now discuss key findings about each method, including:

- *Gesture*: +Reliable performance; +Silent; -Interrupts manual tasks; -PPE may affect recognition; -Learning.
- *Voice*: +Similar to delegating commands; +Familiar; -Issues with recognition can frustrate; -PPE impacts performance;

- *Gaze*: +High input accuracy; +Silent; -Uncertainty over its behaviour and loss of agency/control frustrated users;
- *Active Zone*: +High input accuracy; +Silent; +Least demanding; +Works well with PACS usage workflows; -Unknown challenges of how to deal with multiple users, limited space, different positions in room, etc.

Voice, gesture and gaze clutches each had issues that made them less suitable for interaction in clinical settings **[RQ3]**. Voice was the most familiar interaction modality and other work has suggested a preference for speech commands in hospitals (e.g., [6]). However, environmental and recognition issues can impede speech recognition. We also found the need to choose words carefully, e.g., "lock" and "unlock" are frequently used during orthopaedic procedures.

As an implicit clutching method, we anticipated gaze to be easy for users to understand: i.e., just look at the system to unlock it. This caused more frustration than anticipated, as users often wanted to look at the system to refer to PACS images and reports, without it unlocking and becoming responsive to input. This happened often, suggested by the higher time unlocked for this method. Some users tried to work around this behaviour, e.g., glancing to avoid making direct eye contact. In this case, the clutch method was not harmonious with the way PACS images and reports were used during secondary tasks, which often required glancing at the screen. Users perceived a loss of control over the clutch behaviour, similar to findings reported by Schwarz et al. [34] in their evaluation of implicit clutching methods. We used head orientation to estimate gaze, but future work could consider use of a specific gaze target with eye-tracking, such as an on-screen button, so that users can glance at the screen without unlocking the touchless interface. Such an approach would be unaffected by screen size, permitting larger screens without increasing the risk of false-positive gaze.

Both our quantitative and qualitative findings show the active zone [2] as a good choice for this context. Its simple sensing made it easier for human and computer alike: users understood how it worked and were able to exert control over the system by moving in/out of the active zone, and the system was able to robustly detect when the user was intending to interact as body position could be reliably estimated regardless of posture, PPE, etc. Perhaps most importantly, active zone was a good fit for existing usage habits and clinical workflow, so came naturally to many users. Marking the active zone on the floor with tape, so that users knew where it was, was also an effective analogue solution to a key usability challenge when using sensor-based interfaces. Unlike gaze, our other implicit clutch method, the active zone did not interfere with a user's desire to look at the screen without unlocking it. Implicit clutching is promising because it can reduce cognitive load and streamline interaction, so long as it complements the user's workflow.

Active zone is compelling for use in clinical contexts, although its use raises interesting questions for future work. *Where should the active zone be located in the room, and how large should it be?* O'Hara et al. [31] and Mentis et al. [28] note that a surgeon's position may be dictated by the clinical demands of a particular procedure. An implication of this, is that the ideal position for an active zone may change; this needs to be conveyed to users and some configuration may be required when preparing for a procedure. *Could the active zone impede other health professionals in the operating theatre?* Some settings are crowded and the PACS user may be in close proximity to other members of the surgical team [31], who may enter the active zone with no intention to interact. An implication of this is that the size of an active zone must balance ease-of-use for the user with reduced detection of non-users, as asking them to avoid the active zone may not be possible. *Should any user in the active zone be given control of the system, or just a primary user, and how would user identification be facilitated?* This again arises from the size of the active zone and the potential for several people in close proximity. Individual user tracking could mitigate issues of ambiguous control by allowing one person to retain control of the system when in the active zone, although there are interesting technical challenges associated with this; e.g., how to identify and track users wearing similar PPE. *Should there be multiple active zones to facilitate multiple users or the need to move during surgery?* This relates to the relationship between clinical procedure, collaboration, and need for PACS interaction. Future research should look more at the specific demands of different PACS usage contexts. *Could the active zone impact situations where rapid activation is required?* There may be situations where the user needs to leave the active zone but requires access to the PACS interface with urgency; existing PACS workarounds could indeed be used (e.g., delegation), but could alternatives be offered through other clutching modalities – e.g., a voice clutch that overrides the active zone when needed? Such a multimodal clutch seems sensible, similar to the beneficial redundancy of supporting both speech and gesture commands for medical imaging [27]. *How should the active zone be implemented?* Depending on sensor placement, the surgical table may occlude the surgeon's body [31]. This has implications for how presence in the active zone is determined. We estimated this using torso position, but multiple sensors may offer more reliable detection without occlusion [40].

5.3 Clutching + Authenticating

Hospitals understandably have strict security policies to protect sensitive patient data, often requiring terminals to be unlocked each time they are used and automatically locking users out after very short periods of inactivity. This is a key source of frustration with existing PACS usage: authentication slows down interaction, e.g., logging in each time when periodically moving through a sequence of images. Many of the health professionals expressed a desire for clutching to work alongside authentication; indeed, this would be necessary in practice, since a touchless user interface would not be responsive if a terminal was locked/logged out.

Many participants seemed aware that touchless user interface sensors could potentially identify them; e.g., biometric approaches like facial recognition from optical gesture trackers, retina scans from gaze trackers, and voice recognition **[RQ2]**. This is a compelling topic for future work, as there are significant efficiency gains by making it easier for users to 'unlock' a touchless PACS UI, both from an authentication and touchless input perspective. We envisage a multimodal approach that streamlines access to medical imagery and reports in a touchless PACS interface, e.g., by combining active zone clutching with a form of biometric identification. There are interesting challenges associated with the use of PPE, although persistent user tracking could allow authenticated users to be tracked before they scrub into protective equipment and would keep the system unlocked so long as those users remain in view.

During our study, Ireland's Health Service was affected by a ransomware attack that brought PACS and other hospital systems offline, severely reducing work rates. Participants reflected on poor security practices during the interviews. They reported that sharing passwords was common to get around frustrating security policies; poor security practice and the increasing risk of attacks add further motivation for new authentication practices to be explored.

5.4 Limitations

There are some limitations of our approach that should be considered when interpreting our findings. Our study incorporated individual participants completing tasks in semi-controlled hospital environments. The consistent lighting conditions, absence of other clinical staff and background activity meant input sensing conditions were were good. Likewise, users were positioned to face the Kinect sensor, which meant occlusion was not an issue. In practice, one might expect diminished sensing performance due to, e.g., difficulty tracking the active user's body when there are others in close proximity, the challenges of tracking the body in different orientations, or difficulty isolating the active user's speech commands in the presence of nearby conversations. An implication of this is that our quantitative results reflected generally successful interactions and users had only limited negative experiences to consider during interviews.

Our off-screen task was intended to induce cognitive demand, necessitate context switching, and have users perform actions that meant clutching was necessary. Whilst these tasks replicated some elements of professional practice, they were not representative of the stressful and urgent activities carried out in some areas of the hospital, e.g., in the operating theatre. To help address this, we explored the interaction challenges of more stressful usage contexts through the experience and perspectives of our participants. Another implication of our task design was that users were able to focus more on the interaction and may have been more positive about the interaction techniques because they were in a low-risk situation, where usability issues had no significant consequence (e.g., on health outcomes).

6 CONCLUSION

Touchless input is compelling for clinical user interfaces because touchless input modalities like gesture and speech can allow health professionals to interact directly while maintaining asepsis. Sterility has key implications for interaction and the role of computing systems in clinical practice, e.g., as seen in our qualitative analysis and in prior work on this topic. Touchless interaction has the potential to improve user experience and, more importantly, clinical workflow through sterile interaction, and the technologies that enable this are increasingly capable and ready for deployment.

We took an important look at clutching interactions for initiating and directing interaction towards a touchless PACS interface, a key practical consideration for deployment. When interacting with touchless medical imaging systems, unintended input can lead to user frustration and delays resetting key imaging. Clutching helps to avoid such issues by reducing the likelihood of false-positive gesture and speech recognition, key for reliable touchless input.

We evaluated four existing clutching techniques in a touchless PACS interface. The qualitative interview data from our sample of 34 health professionals gave additional understanding of the user experience of these mechanisms, and attitudes towards these systems. Our findings give valuable insight into the potential benefits of touchless interaction in clinical settings and the challenges of reliably using touchless input in this context. Our findings also give insight into the user experience of different clutching techniques and how well they integrate with PACS workflows.

Active zone [2] was the most promising clutch because it was the least demanding and worked well with existing PACS usage habits. However, we uncovered compelling challenges with its use. Hospital environments are often constrained (e.g., equipment that occludes the user, nearby colleagues, many ongoing activities) and so, too, are clinical procedures (e.g., where users can be situated, what they can(not) do, who else is nearby). Solutions are needed to address these, e.g., more robust tracking, resilience to multiple persons, varied zone shape, size and position, etc.

There are several benefits to touchless PACS interaction and health professionals desire touchless input capabilities for other computing systems, as this work has shown. Whilst interesting challenges for future work have emerged during our analysis, we have taken key steps towards the practical deployment of touchless PACS technology in clinical settings. Touchless interaction has the potential to improve clinician workflow, and possibly patient outcomes. Clinicians will be able to do their jobs better, helping patients while improving sterility and reducing the spread of pathogens from shared surfaces and input devices.

ACKNOWLEDGMENTS

The work of Gavin Doherty is supported, in part, by Science Foundation Ireland grant number 13/RC/2106_P2 to the Adapt Centre and 13/RC/2094_P2 to the Lero Centre.

REFERENCES

[1] Mohamed Asni, Daniel Shapiro, Miodrag Bolic, Tony Mathew, and Leor Grebler. 2018. Speaker Differentiation Using a Convolutional Autoencoder. *2018 IEEE International Conference on the Science of Electrical Engineering in Israel (ICSEE)*, 1–5. https://doi.org/10.1109/ICSEE.2018.8646307

[2] Thomas Baudel and Michel Beaudouin-Lafon. 1993. Charade: Remote Control of Objects Using Free-Hand Gestures. *Commun. ACM* 36, 7 (1993), 28–35. https://doi.org/10.1145/159544.159562

[3] Victoria Bellotti, Maribeth Back, W. Keith Edwards, Rebecca E. Grinter, Austin Henderson, and Cristina Lopes. 2002. Making Sense of Sensing Systems: Five Questions for Designers and Researchers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '02*. ACM Press, 415–422. https://doi.org/10.1145/503376.503450

[4] Richard A. Bolt. 1980. "Put-That-There": Voice and Gesture at the Graphics Interface. *ACM SIGGRAPH Computer Graphics* 14, 3 (1980), 262–270. https://doi.org/10.1145/800250.807503

[5] Seán Cronin and Gavin Doherty. 2019. Touchless computer interfaces in hospitals: A review. *Health Informatics Journal* 25, 4 (2019), 1325–1342. https://doi.org/10.1177/1460458217748342

[6] Séan Cronin, Bridget Kane, and Gavin Doherty. 2021. A Qualitative Analysis of the Needs and Experiences of Hospital-based Clinicians when Accessing Medical Imaging. *Journal of Digital Imaging* (4 2021). https://doi.org/10.1007/s10278-021-00446-1

[7] David Fleer and Christian Leichsenring. 2012. MISO: A Context-Sensitive Multimodal Interface for Smart Objects Based on Hand Gestures and Finger Snaps. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology - UIST '12*. ACM Press, 93–94. https://doi.org/10.1145/2380296.2380338

[8] Euan Freeman, Stephen Brewster, and Vuokko Lantz. 2016. Do That, There: An Interaction Technique for Addressing In-Air Gesture Systems. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*. ACM Press, 2319–2331. https://doi.org/10.1145/2858036.2858308

[9] Euan Freeman, Gareth Griffiths, and Stephen Brewster. 2017. Rhythmic Micro-Gestures: Discreet Interaction On-the-Go. In *Proceedings of 19th ACM International Conference on Multimodal Interaction - ICMI '17*. ACM, 115–119. https://doi.org/10.1145/3136755.3136815

[10] Euan Freeman, Dong-Bach Vo, and Stephen Brewster. 2019. HaptiGlow: Helping Users Position their Hands for Better Mid-Air Gestures and Ultrasound Haptic Feedback. In *Proceedings of the IEEE World Haptics Conference (WHC '19)*. IEEE, 289–294. https://doi.org/10.1109/WHC.2019.8816092

[11] Euan Freeman, Julie Williamson, Sriram Subramanian, and Stephen Brewster. 2018. Point-and-Shake: Selecting from Levitating Object Displays. In *Proceedings of the 36th Annual ACM Conference on Human Factors in Computing Systems - CHI '18*. ACM Press, Paper 18. https://doi.org/10.1145/3173574.3173592

[12] Nicola K Gale, Gemma Heath, Elaine Cameron, Sabina Rashid, and Sabi Redwood. 2013. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Medical Research Methodology* 13 (12 2013). Issue 1. https://doi.org/10.1186/1471-2288-13-117

[13] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, Peter A. Hancock and Najmedin Meshkati (Eds.). Advances in Psychology, Vol. 52. North-Holland, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

[14] Benjamin Hatscher and Christian Hansen. 2018. Hand, Foot or Voice: Alternative Input Modalities for Touchless Interaction in the Medical Domain. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction* (Boulder, CO, USA) *(ICMI '18)*. Association for Computing Machinery, New York, NY, USA, 145–153. https://doi.org/10.1145/3242969.3242971

[15] Julian Hettig, Patrick Saalfeld, Maria Luz, Mathias Becker, Martin Skalej, and Christian Hansen. 2017. Comparison of gesture and conventional interaction techniques for interventional neuroradiology. *International Journal of Computer Assisted Radiology and Surgery* 12 (9 2017). Issue 9. https://doi.org/10.1007/s11548-017-1523-7

[16] Julian Hettig, Patrick Saalfeld, Maria Luz, Mathias Becker, Martin Skalej, and Christian Hansen. 2017. Comparison of gesture and conventional interaction techniques for interventional neuroradiology. *International journal of computer assisted radiology and surgery* 12, 9 (2017), 1643–1653.

[17] Ken Hinckley, Randy Pausch, John C. Goble, and Neal F. Kassell. 1994. A Survey of Design Issues in Spatial Input. In *Proceedings of the 7th Annual ACM Symposium on User Interface Software and Technology - UIST '94*. ACM Press, 213–222. https://doi.org/10.1145/192426.192501

[18] Ken Hinckley, Randy Pausch, John C. Goble, and Neal F. Kassell. 1994. Passive real-world interface props for neurosurgical visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '94*. ACM Press, 452–458. https://doi.org/10.1145/259963.260443

[19] Behrang Homayoon, John Chung, Ripal T. Gandhi, and David M. Liu. 2020. Early clinical experience with a touchless image navigation interface for the endovascular suite. *Minimally Invasive Therapy & Allied Technologies* 29 (6 2020). Issue 3. https://doi.org/10.1080/13645706.2019.1612440

[20] Mithun Jacob, Christopher Cange, Rebecca Packer, and Juan P Wachs. 2012. Intention, context and gesture recognition for sterile MRI navigation in the operating room. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications - CIARP '12*. Springer, 220–227. https://doi.org/10.1007/978-3-642-33275-3_27

[21] Mithun George Jacob, Juan Pablo Wachs, and Rebecca A Packer. 2012. Hand-gesture-based sterile interface for the operating room using contextual cues for the navigation of radiological images. *Journal of the American Medical Informatics Association* 20, e1 (12 2012), e183–e186. https://doi.org/10.1136/amiajnl-2012-001212

[22] Shahram Jalaliniya, Jeremiah Smith, Miguel Sousa, Lars Büthe, and Thomas Pederson. 2013. Touch-Less Interaction with Medical Images Using Hand & Foot Gestures. In *Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication* (Zurich, Switzerland) *(UbiComp '13 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 1265–1274. https://doi.org/10.1145/2494091.2497332

[23] Rose Johnson, Kenton O'Hara, Abigail Sellen, Claire Cousins, and Antonio Criminisi. 2011. *Exploring the Potential for Touchless Interaction in Image-Guided Interventional Radiology.* Association for Computing Machinery, New York, NY, USA, 3323–3332. https://doi.org/10.1145/1978942.1979436

[24] Rick Kjeldsen and Jacob Hartman. 2001. Design issues for vision-based computer interaction systems. In *Proceedings of the 2001 workshop on Perceptive User Interfaces - PUI '01*. ACM Press, 1–8. https://doi.org/10.1145/971478.971511

[25] Hannah Limerick. 2020. Call to Interact: Communicating Interactivity and Affordances for Contactless Gesture Controlled Public Displays. In *Proceedings of the 9TH ACM International Symposium on Pervasive Displays* (Manchester, United Kingdom) *(PerDis '20)*. Association for Computing Machinery, New York, NY, USA, 63–70. https://doi.org/10.1145/3393712.3395338

[26] Naveen Madapana, Glebys Gonzalez, Rahul Taneja, Richard Rodgers, Lingsong Zhang, and Juan Wachs. 2019. Preference elicitation: Obtaining gestural guidelines for PACS in neurosurgery. *International Journal of Medical Informatics* 130 (10 2019). https://doi.org/10.1016/j.ijmedinf.2019.07.013

[27] Helena M. Mentis, Kenton O'Hara, Gerardo Gonzalez, Abigail Sellen, Robert Corish, Antonio Criminisi, Rikin Trivedi, and Pierre Theodore. 2015. Voice or Gesture in the Operating Room. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI EA '15)*. Association for Computing Machinery, New York, NY, USA, 773–780. https://doi.org/10.1145/2702613.2702963

[28] Helena M. Mentis, Kenton O'Hara, Abigail Sellen, and Rikin Trivedi. 2012. *Interaction Proxemics and Image Use in Neurosurgery.* Association for Computing Machinery, New York, NY, USA, 927–936. https://doi.org/10.1145/2207676.2208536

[29] Andre Mewes, Bennet Hensen, Frank Wacker, and Christian Hansen. 2017. Touchless interaction with software in interventional radiology and surgery: a systematic literature review. *International journal of computer assisted radiology and surgery* 12, 2 (2017), 291–305.

[30] Kenton O'Hara, Gerardo Gonzalez, Graeme Penney, Abigail Sellen, Robert Corish, Helena Mentis, Andreas Varnavas, Antonio Criminisi, Mark Rouncefield, Neville Dastur, et al. 2014. Interactional order and constructed ways of seeing with touchless imaging systems in surgery. *Computer Supported Cooperative Work (CSCW)* 23, 3 (2014), 299–337.

[31] Kenton O'Hara, Gerardo Gonzalez, Abigail Sellen, Graeme Penney, Andreas Varnavas, Helena Mentis, Antonio Criminisi, Robert Corish, Mark Rouncefield, Neville Dastur, and Tom Carrell. 2014. Touchless Interaction in Surgery. *CACM* 57, 1 (2014), 70–77. https://doi.org/10.1145/2541883.2541899

[32] Jonathan A. Otter, Saber Yezli, and Gary L. French. 2014. *The Role of Contaminated Surfaces in the Transmission of Nosocomial Pathogens.* Springer International Publishing, Cham, 27–58. https://doi.org/10.1007/978-3-319-08057-4_3

[33] Stefanie TL Pöhlmann, Elaine F Harkness, Christopher J Taylor, and Susan M Astley. 2016. Evaluation of Kinect 3D sensor for healthcare imaging. *Journal of medical and biological engineering* 36, 6 (2016), 857–870.

[34] Julia Schwarz, Charles Claudius Marais, Tommer Leyvand, Scott E. Hudson, and Jennifer Mankoff. 2014. Combining Body Pose, Gaze, and Gesture to Determine Intention to Interact in Vision-Based Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '14*. ACM Press, 3443–3452. https://doi.org/10.1145/2556288.2556989

[35] Matt Strickland, Jamie Tremaine, Greg Brigley, and Calvin Law. 2013. Using a depth-sensing infrared camera system to access and manipulate medical imaging from within the sterile operating field. *Canadian Journal of Surgery* 56, 3 (2013), E1. https://doi.org/10.1503/cjs.035311

[36] Hemant Bhaskar Surale, Fabrice Matulic, and Daniel Vogel. 2019. *Experimental Analysis of Barehand Mid-Air Mode-Switching Techniques in Virtual Reality.* Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300426

[37] Justin H. Tan, Cherng Chao, Mazen Zawaideh, Anne C. Roberts, and Thomas B. Kinney. 2013. Informatics in Radiology: Developing a Touchless User Interface for Intraoperative Image Control during Interventional Radiology Procedures. *RadioGraphics* 33, 2 (2013), E61–E70. https://doi.org/10.1148/rg.332125101

[38] Robert Walter, Gilles Bailly, and Jörg Müller. 2013. StrikeAPose: Revealing Mid-Air Gestures on Public Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*. ACM Press, 841–850. https://doi.org/10.1145/2470654.2470774

[39] Andrew D. Wilson. 2008. Robust computer vision-based detection of pinching for one and two-handed gesture input. In *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology - UIST '06*. ACM Press, 255–258. https://doi.org/10.1145/1166253.1166292

[40] Andrew D. Wilson and Hrvoje Benko. 2010. Combining Multiple Depth Cameras and Projectors for Interactions on, above and between Surfaces. In *Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology* (New York, New York, USA) *(UIST '10)*. Association for Computing Machinery, New York, NY, USA, 273–282. https://doi.org/10.1145/1866029.1866073

[41] Ambreen Zaman, Lars Reisig, Anke Verena Reinschluessel, Huseyin Bektas, Dirk Weyhe, Marc Herrlich, Tanja Döring, and Rainer Malaka. 2018. An Interactive Shoe for Surgeons: Hand-Free Interaction with Medical 2D Data. *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3170427.3188606

[42] Tian Zhou and Juan Pablo Wachs. 2018. Early prediction for physical human robot collaboration in the operating room. *Autonomous Robots* 42 (6 2018), 977–995. Issue 5. https://doi.org/10.1007/s10514-017-9670-9